Cognitive and Visual Speech Contributions to Speech Perception in Challenging Listening Conditions

By:

Rachel V. Wayne

A thesis submitted to the Department of Psychology
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
November, 2016

© Rachel V. Wayne, 2016
This work is licensed under a Creative Commons Attribution 4.0 International License
Abstract

Speech perception routinely takes place in noisy or degraded listening environments, leading to ambiguity in the identity of the speech token. Here, I present one review paper and two experimental papers that highlight cognitive and visual speech contributions to the listening process, particularly in challenging listening environments. First, I survey the literature linking audiometric age-related hearing loss and cognitive decline and review the four proposed causal mechanisms underlying this link. I argue that future research in this area requires greater consideration of the functional overlap between hearing and cognition. I also present an alternative framework for understanding causal relationships between age-related declines in hearing and cognition, with emphasis on the interconnected nature of hearing and cognition and likely contributions from multiple causal mechanisms. I also provide a number of testable hypotheses to examine how impairments in one domain may affect the other. In my first experimental study, I examine the direct contribution of working memory (through a cognitive training manipulation) on speech in noise comprehension in older adults. My results challenge the efficacy of cognitive training more generally, and also provide support for the contribution of sentence context in reducing working memory load. My findings also challenge the ubiquitous use of the Reading Span test as a pure test of working memory. In a second experimental (fMRI) study, I examine the role of attention in audiovisual speech integration, particularly when the acoustic signal is degraded. I demonstrate that attentional processes support audiovisual speech integration in the middle and superior temporal gyri, as well as the fusiform gyrus. My results also suggest that the superior temporal sulcus is sensitive to intelligibility enhancement, regardless of how this benefit is obtained (i.e., whether it is obtained through visual speech information or speech clarity). In addition, I also demonstrate that both the cingulo-opercular
network and motor speech areas are recruited in difficult listening conditions. Taken together, these findings augment our understanding of cognitive contributions to the listening process and demonstrate that memory, working memory, and executive control networks may flexibly be recruited in order to meet listening demands in challenging environments.
Statement of Co-authorship

In all cases, I (Rachel V. Wayne) fully participated in study design, data collection, data analysis, and manuscript preparation. I certify that I have obtained permission from all co-authors to include the below published materials in my thesis. I certify that I have obtained written permission from copyright owners to include the below published materials in my thesis (Appendix A).

**Chapter 2** has been published in its entirety and can be cited as:


Ingrid S. Johnsrude contributed to manuscript preparation.

**Chapter 3** has been published in its entirety and can be cited as:


Cheryl Hamilton and Julia Jones Huyck contributed to study design, stimulus creation, data collection, and data analysis. Ingrid S. Johnsrude contributed to study design and manuscript preparation.

**Chapter 4** will be submitted for publication.

Tristan Mohammed contributed to data collection. Agnes Alsius Rance contributed to study design, stimulus creation, and data collection. Kevin Munhall contributed to study design. Conor J. Wild and Sam Evans contributed to data analysis. Ingrid S. Johnsrude contributed to study design and manuscript preparation.
Acknowledgements

First and foremost, I am deeply indebted to my supervisor, Ingrid Johnsrude. I feel so lucky to have had the honour to work with you over the past 7 years. It has been a joy to witness your insatiable curiosity and love for science. I am immensely grateful for your continuous wisdom, mentorship, and support, and especially for your willingness to go above and beyond in supporting me from London. You have always demanded my very best (despite my frustrations at times!), and for that I am very thankful.

I am also indebted to Kevin Munhall, my “other” supervisor, who always kept his door open to me and welcomed me into his lab during my final two years of study. You have always been incredibly encouraging and supportive, both academically and personally, and for that I am so grateful.

I would like to thank to all the participants who made this research possible, particularly those who endured an hour of the claustrophobia-inducing MRI machine, as well as the older adults who participated in the Cogmed study and gave us well over 50 hours of their time, all in the name of science.

The projects of my thesis would not have been possible without the generous contributions of my colleagues Kris Marble, Cheryl Hamilton, Julia Jones Huyck, and Conor Wild. I would also like to thank my collaborators and colleagues at University College London, Mairead MacSweeney and Eva Gutierrez for enriching my PhD experience, with special thanks to Sam Evans for his valuable input and contribution to the fMRI study. I am especially grateful to Agnes Alsius for her endless encouragement and moral support (and commiseration). It was such a pleasure to work alongside you and to watch a spontaneous research project unfold.
I thank my labmates over the years, Dora, Fabienne, Julia, Zane, Graham, Harrison, Spencer, Nida, Mashal, as well as Heather T. and L. (in addition to my newer, inherited labmates at UWO). Thank you for your unwavering support and encouragement, and for making the lab such an enjoyable place to be. To all of my friends and colleagues, thank you for making the graduate experience a most memorable 7 years of my life.

To my family, especially, my parents, Martyn and Andrea, my sister Brooke, and my grandmother Barbara, as well as my extended family, none of this would have been possible without your steadfast support and dedication. And finally to Michael, who not only made sure that I was well fed during the final frantic days of writing, but who always had faith in me, thank you for your love, support, and inspiration.
# Table of Contents

Abstract.............................................................................................................................................. i

Chapter 1 - General Introduction ........................................................................................................ 1
  1.1 Overview ..................................................................................................................................... 2
  1.2 Sources of Perceptual Ambiguity Relating to the Speech Signal .............................................. 3
    1.2.1 Energetic vs. Non-energetic Masking .................................................................................. 3
    1.2.2 Common Sources of Perceptual Ambiguity in Speech Comprehension .............................. 3
    1.2.3 Hearing Impairment as a Source of Perceptual Ambiguity .................................................. 5
  1.3 Sources of Non-auditory Contributions to Speech Comprehension ........................................... 8
    1.3.1 Cognition in Hearing ............................................................................................................ 8
    1.3.2 Cognition: Listening Effort .................................................................................................. 12
    1.3.3 Visual Speech Contributions to Speech Perception ............................................................. 14
  1.4 fMRI Techniques ......................................................................................................................... 16
  1.5 Outline of Thesis ........................................................................................................................ 18

Chapter 2 - A Review of Causal Mechanisms Underlying the Link Between Age-related Hearing Loss and Cognitive Decline ........................................................................................................ 20
  2.1 Introduction .................................................................................................................................. 21
  2.2 Operational Definitions of Hearing Loss and Cognitive Function, and Their Limitations. 22
  2.3 Is There a Link Between Hearing Loss and Cognitive Decline? .................................................. 28
  2.4 Hypothesized Relationships between Age-Related Hearing Loss and Cognitive Decline. 33
    2.4.1 Cognitive Load on Perception Hypothesis .......................................................................... 36
    2.4.2 Information-Degradation Hypothesis .................................................................................. 36
    2.4.3 Sensory-Deprivation Hypothesis ......................................................................................... 38
    2.4.4 Common-Cause Hypothesis ................................................................................................ 40
  2.5 A Working Framework ................................................................................................................ 45
  2.6 Directions for Future Research .................................................................................................... 47
  2.7 Conclusion ................................................................................................................................... 50

Chapter 3- Working Memory Training and Speech in Noise Comprehension in Older Adults... 52
  3.1 Introduction .................................................................................................................................. 53
  3.2 Methods ......................................................................................................................................... 61
    3.2.1 Subjects .................................................................................................................................. 61
    3.2.2 Screening Procedure ............................................................................................................. 61
    3.2.3 Cognitive Training Procedure ................................................................................................ 62
    3.2.4 Testing Procedure .................................................................................................................. 64
    3.2.5 Cognitive Test Battery .......................................................................................................... 64
    3.2.6 Speech Tests .......................................................................................................................... 67
  3.3 Results ......................................................................................................................................... 70
    3.3.1 Improvement on Cognitive Training (Cogmed). ................................................................. 73
    3.3.2 Transfer to Cognitive Tests ................................................................................................... 74
    3.3.3 Transfer to Speech Tests ....................................................................................................... 76
    3.3.4 Correlations between Speech and Cognitive Tests .............................................................. 77
    3.3.5 Correlations between Reading Span and Cognitive Measures ............................................. 80
  3.4 Discussion ..................................................................................................................................... 81
    3.4.1 Generalizability of Cognitive Training ............................................................................... 81
    3.4.2 Relationship between Working Memory and Speech in Noise .......................................... 85
### Chapter 4 - The Role of Attention in Audiovisual Speech Integration

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
</tr>
<tr>
<td>4.2 Methods</td>
</tr>
<tr>
<td>4.2.1 Participants</td>
</tr>
<tr>
<td>4.2.2 Experimental Design</td>
</tr>
<tr>
<td>4.2.3 Materials</td>
</tr>
<tr>
<td>4.2.4 Pilot Procedure</td>
</tr>
<tr>
<td>4.2.5 Procedure</td>
</tr>
<tr>
<td>4.2.6 Behavioral Post-test</td>
</tr>
<tr>
<td>4.2.7 fMRI Protocol and Data Acquisition</td>
</tr>
<tr>
<td>4.2.8 fMRI Data Preprocessing</td>
</tr>
<tr>
<td>4.2.8 fMRI Analysis</td>
</tr>
<tr>
<td>4.2.9 ALE Map for Audiovisual Speech Integration</td>
</tr>
<tr>
<td>4.3 Behavioural Results</td>
</tr>
<tr>
<td>4.3.1 Auditory Distractor Task</td>
</tr>
<tr>
<td>4.3.2 Speech Task</td>
</tr>
<tr>
<td>4.3.4 Post-test</td>
</tr>
<tr>
<td>4.4 Imaging Results</td>
</tr>
<tr>
<td>4.4.1 Main effect of Speech Clarity</td>
</tr>
<tr>
<td>4.4.2 Main effect of Attention</td>
</tr>
<tr>
<td>4.4.3 Main effect of AV Type</td>
</tr>
<tr>
<td>4.4.4 Attention x Speech Clarity</td>
</tr>
<tr>
<td>4.4.5 Speech Clarity x AV type</td>
</tr>
<tr>
<td>4.4.6 Attention x AV Speech Type</td>
</tr>
<tr>
<td>4.4.7 AV Type x Speech Type x Attention Interaction</td>
</tr>
<tr>
<td>4.4.8 Does AV Speech Integration Require Attention? Additional Analyses</td>
</tr>
<tr>
<td>4.5 ICA Analysis</td>
</tr>
<tr>
<td>4.5.1 Effects Related to Main Effects of Attention, Speech Clarity, and AV Speech Type:</td>
</tr>
<tr>
<td>4.5.2 Effects Related to Attention x Speech Clarity Interaction:</td>
</tr>
<tr>
<td>4.5.3 Effects Related to Attention x AV Type Interaction:</td>
</tr>
<tr>
<td>4.5.4 Effects Related to Listening Effort:</td>
</tr>
<tr>
<td>4.6 Discussion</td>
</tr>
<tr>
<td>4.6.1 Attentional Modulation of BOLD Activity in FFA</td>
</tr>
<tr>
<td>4.6.2 The Role of Attention in AV Speech Integration</td>
</tr>
<tr>
<td>4.6.3 Cognitive Control Networks in Speech Processing</td>
</tr>
<tr>
<td>4.7 Summary</td>
</tr>
</tbody>
</table>

### Chapter 5 - General Discussion

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Summary and Contributions</td>
</tr>
<tr>
<td>5.1.1 Summary of Results from Chapter 2 – Age-related Hearing Loss and Cognitive Decline</td>
</tr>
<tr>
<td>5.1.2 Summary of Results from Chapter 3 – Working Memory Training and Speech-in-Noise Performance in Older Adults</td>
</tr>
<tr>
<td>5.1.3 Summary of Results from Chapter 4 – Attention in Audiovisual Speech Integration</td>
</tr>
</tbody>
</table>
5.2 Further Discussion .................................................................................................................. 153
5.3 Directions for Future Research .......................................................................................... 156
References .................................................................................................................................... 161
Appendix A: Copyright Permissions ....................................................................................... 210
Appendix B: Pilot Study Results from Chapter 4 ...................................................................... 211
  Introduction .............................................................................................................................. 211
  Participants .............................................................................................................................. 211
  Methods and Materials .......................................................................................................... 211
  Results ...................................................................................................................................... 211
    Auditory Distractor Task ........................................................................................................ 212
    Intelligibility Data .................................................................................................................. 212
  Post-test Recognition Data ...................................................................................................... 214
    6-band ................................................................................................................................... 214
    4-band .................................................................................................................................. 214
  Discussion .................................................................................................................................. 215
Appendix C: Ethics Approval (for Studies Detailed in Chapters 3 & 4) ................................. 217
Appendix D: Speech Spatial Qualities Questionnaire ............................................................ 220
Appendix E: List of Stimuli ........................................................................................................ 228
  Chapter 3 – Coherent/ Anomalous Sentence Sets ............................................................... 228
  Chapter 4 – Experimental Sentence Sets ............................................................................... 232
  Chapter 4 – Foil Sentences (Recognition Post-test) .............................................................. 237
List of Figures

Figure 2–1: A summary of the four hypotheses for the causal relationship between age-related hearing loss and cognitive decline. ................................................................. 34
Figure 2-2: An alternative framework for conceptualizing the relationship between hearing loss and cognitive decline. ................................................................. 46
Figure 3-1: Study design. ................................................................................................. 60
Figure 3-2: Dissociating training from practice effects. .................................................. 61
Figure 3-3: Accuracy scores for the BUG (sentence matrix) speech task across the three study time points. ......................................................................................... 70
Figure 3-4: Percentage word scores for high- and low-context sentences across the three study time points. ......................................................................................... 71
Figure 4-2. ALE activation map for fMRI studies assessing audiovisual speech integration. ... 104
Figure 4-3. Behavioural data for gist task performed on attended sentence stimuli in the MR system, with pilot word report for comparison......................................................... 107
Figure 4-4. fMRI behavioural results from old/new discrimination post-test. ................. 108
Figure 4-5. t-contrasts for the main effect of speech clarity, family-wise error rate corrected, p < .05.................................................................................................................. 110
Figure 4-6. t-contrasts for the main effect of attention. .................................................. 111
Figure 4-7. t-contrasts for the main effect of AV speech type........................................ 112
Figure 4-8. Attention x speech clarity interaction. ......................................................... 116
Figure 4-9. Contrast values for the attention x speech clarity interaction in the left supramarginal gyrus........................................................................................................ 116
Figure 4-10. Attention x AV speech type interaction...................................................... 118
Figure 4-11. Contrast values for peak voxels in the attention x AV speech type interaction.... 119
Figure 4-12. Contrast values for the 3-way interaction in the right middle orbitofrontal gyrus. 120
Figure 4-13. Thresholded map (FWE p < .05) for Attention x AV Speech Type Interaction Masked by Attention x Speech Clarity Interaction (FWE = .05 corrected)........................................ 124
Figure 4-14. Contrast values for masked contrasts between two 2-way interactions......... 125
Figure 4-15. ICA spatial maps for 9 retained components. ............................................. 128
Figure 4-16. Plot of regression values by condition for each component. ....................... 131
Figure 4-17. Derived ALE Map for AV integration overlayed on the spatial map for Component 12.................................................................................................................. 134
Figure Appendix B-1. Sentence word report scores by condition for both 6- and 4-band noise-vocoded speech. Error bars reflect standard mean error........................................ 214
Figure Appendix B-2. d’ scores for recognition memory post-test. ............................... 215
Table 2-1: Description, interpretation, and psychometric properties of cognitive screening tools used in assessment of the relationship between hearing loss and cognitive decline. ................................. 26
Table 2-2: Summary of proposed four directional hypotheses for the relationship between hearing loss and cognitive decline. .......................................................... 35
Table 3-1: Description of different versions of the Reading Span test............................................. 57
Table 3-2: Summary of cognitive tests and outcome measures...................................................... 67
Table 3-3: Means and standard deviations for all outcome measures. ........................................... 73
Table 3-4: F-test statistics for main effect of time for all dependent variables ......................... 76
Table 3-5: Correlations between speech tests and (select) cognitive tests. ............................. 80
Table 4-1. Studies and coordinates included in the ALE meta-analysis used to derive a spatial map for a small-volume correction in my study. ........................................... 105
Table 4-2. Peak voxels for main effects, corrected at FWE < .05.............................................. 114
Table 4-3. Peak voxels for interactions, significant at p < .05 at a whole-brain corrected level using FWE correction................................................................. 121
Table 4-4. Peak voxels for Components 12 and 13 correlated and anti-correlated activity, significant at p < .05 at a whole-brain corrected level using FWE correction............... 137
Table 4-5. Comparison of inferior frontal region coordinates across studies......................... 143
List of Common Abbreviations

3MS Modified Mini-Mental State Exam
AG angular gyrus
ALE activation likelihood estimation
AMNART American version of the National Adult Reading Test
ANOVA analysis of variance
AO auditory-only speech
AV audiovisual speech
BOLD blood oxygen level dependent
C clear speech
CANTAB Cambridge Neuropsychological Test Automated Battery
D degraded speech
dB HL decibels hearing level
DSM-IV Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
DSST Digit Symbol Substitution Test
EEG electroencephalography
FDR false-discovery rate
FFA fusiform face area
fMRI functional magnetic resonance imaging
FWE family-wise error rate
H Context high-context sentences
H-L Context high minus low context sentences
ICA independent component analysis
IFG inferior frontal gyrus
IPS inferior parietal sulcus
ITG inferior temporal gyrus
L Context low-context sentences
LGN lateral geniculate nucleus
LNS Letter-Number Sequencing Test
LNS L Letter-Number Sequencing Test, longest sequence
M matching speech
MCC mid-cingulate cortex
MCI Mild Cognitive Impairment
MFG middle frontal gyrus
MoCA Montreal Cognitive Assessment
MOFG middle orbitofrontal gyrus
MOG middle occipital gyrus
MM mismatching speech
MMSE Mini-mental State Exam
MTG middle temporal gyrus
NV noise-vocoded
PAL Paired Associate Learning Test
PET positron emission tomography
PTA pure-tone audiometry
RMS root-mean squared
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS</td>
<td>Reading Span Test</td>
</tr>
<tr>
<td>RS L</td>
<td>Reading Span Test, longest span</td>
</tr>
<tr>
<td>RTI</td>
<td>Reaction Time Test</td>
</tr>
<tr>
<td>RVP</td>
<td>Rapid Visual Processing Test</td>
</tr>
<tr>
<td>SFG</td>
<td>superior frontal gyrus</td>
</tr>
<tr>
<td>SMG</td>
<td>supramarginal gyrus</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SOG</td>
<td>superior occipital gyrus</td>
</tr>
<tr>
<td>SWM</td>
<td>Spatial Working Memory Test</td>
</tr>
<tr>
<td>SPL</td>
<td>sound pressure level</td>
</tr>
<tr>
<td>SPL</td>
<td>superior parietal lobule (chapter 5 only)</td>
</tr>
<tr>
<td>SSQ</td>
<td>Speech Spatial Qualities Questionnaire</td>
</tr>
<tr>
<td>SSP</td>
<td>Spatial Span Test</td>
</tr>
<tr>
<td>STG</td>
<td>superior temporal gyrus</td>
</tr>
<tr>
<td>STS</td>
<td>superior temporal sulcus</td>
</tr>
<tr>
<td>TA</td>
<td>acquisition time</td>
</tr>
<tr>
<td>TE</td>
<td>echo time</td>
</tr>
<tr>
<td>TMS</td>
<td>transcranial magnetic stimulation</td>
</tr>
<tr>
<td>TR</td>
<td>time to respond</td>
</tr>
<tr>
<td>WAIS-IV</td>
<td>Weschler Adult Intelligence Scale, Fourth Edition</td>
</tr>
</tbody>
</table>
1.1 Overview

Speech communication is a robust process with immense biological and evolutionary relevance. Its complexity is difficult to appreciate in ideal, quiet listening environments. It is only in difficult listening situations, such as in noisy environments, when speech is distorted, or when hearing is impaired, that we really become aware of the cognitively demanding listening process that we often perform with remarkable ease. Often, portions of the speech signal may be masked or distorted, or even missing or interrupted, leading to perceptual ambiguity. Cognitive processes play a crucial role in resolving such perceptual challenges. This is achieved through utilization of previous experience and context, as well as recruitment of linguistic knowledge and more domain-general cognitive resources in order to assist in the listening process. Moreover, most of everyday conversation takes place against a backdrop of background noise or other competing talkers. Subsequently, our ability to resolve perceptual ambiguity through supportive cognitive processes, as well as through recruitment of visual speech information (i.e., facial gestural information that accompanies acoustic speech), is crucial to the listening process.

This thesis focuses on cognitive and visual speech contributions to speech comprehension, particularly when there is ambiguity in the auditory speech signal. This line of research has several important implications: this includes informing our understanding of communication disorders, as well as applications for the large population of older adults who experience difficulty understanding speech in background noise. This next section explores various sources of ambiguity in the speech signal.
1.2 Sources of Perceptual Ambiguity Relating to the Speech Signal

1.2.1 Energetic vs. Non-energetic Masking

Before proceeding with an overview of perceptual ambiguity in the listening process, it is necessary to first distinguish between energetic vs. non-energetic masking. Noise can interfere with the speech signal in two main ways. Energetic masking refers to noise that interferes with the physical properties of the speech signal itself. Such masking occurs within the acoustic environment, such as in the case of two overlapping sounds that carry energy within the same critical acoustic frequency bands (Pollack, 1975). Alternatively, non-energetic masking, also commonly referred to as informational masking, accounts for the remaining masking sources. In this case, the physical properties of the speech signal reach the listener in their original, intact form, but masking occurs due to either cognitive (e.g., competing attentional demands) or perceptual interference (e.g., processing load due to competing perceptual demands) with the speech signal (Pollack, 1975). Note that speech in noise potentially poses both an energetic and informational masking challenge to the listener, depending on the degree of overlapping frequency information. Whereas the definition of energetic masking is quite specific, informational masking tends to be used as a broad ‘catch-all’ term. Subsequently, the precise differentiation between energetic and informational masking remains an ongoing subject of discussion (Durlach, 2006; Kidd, Mason, Richards, Gallun & Durlach, 2008).

1.2.2 Common Sources of Perceptual Ambiguity in Speech Comprehension

Speech comprehension is a demanding listening process, even in quiet, which is often confounded by multiple sources of ambiguity, as discussed below. Here, I use the term ‘perceptual ambiguity’ to refer to all sources of ambiguity that occur beyond the stimulus in perception.
A unitary speech token (e.g., the word “single”), can take an endless number of acoustic forms as a result of several factors, contributing to ambiguity in speech form. Even within individuals, the manner in which a syllable is pronounced is influenced by preceding or following syllables, a process termed co-articulation. Moreover, the same word produced by different speakers will result in variations in its spectral and temporal form as a result of differences in speaker voice characteristics. This is due to anatomical variability of the vocal tract, patterns of word or sentence segmentation, speech rate or, in more extreme cases, the presence of an accent. Variability in portions or all of a speech signal can also be the product of energetic masking. This can lead to distortion in the speech environment due to room acoustics or reverberation, or of degradation due to loss of signal fidelity (e.g., listening over a loud speaker). In other cases, portions of the speech signal may be missing altogether. For instance, the utterance “ingle” may be heard if the first sound is clipped; as this is not an identifiable word, the listener must engage in a search for more plausible options.

In addition to ambiguity of form, listeners must also resolve ambiguity of meaning. Semantic ambiguity refers to words in which the same written (and acoustic) form carries multiple meanings; for example, the sentence: “There was not a single man at the party” could refer to a party without a single male, or a party in which there were no bachelors present. Successful resolution of either lexical or semantic ambiguity, particularly at the sentence level, requires the listener to recruit linguistic knowledge (e.g., how words are parsed) and context (both within the sentence and within the larger conversation), in order to decode the speaker’s message. Thus, not only may a singular speech token take on a number of acoustic forms, the token may also have multiple meanings that must be resolved.
It is also important to consider that real-world speech comprehension often occurs in the form of a conversation. We usually do not pause while we work to resolve ambiguity in the listening process. Rather, successful comprehension of previously heard speech often occurs simultaneously while attention is directed to an ongoing, continuous stream of speech. Moreover, interfering background noise or competing speech is a major source of ambiguity that can compromise the fidelity of the speech signal. It may also impose additional attentional demands, requiring the listener to inhibit irrelevant speech streams or information. In addition, everyday speech communication is often attached to concurrent processing demands stemming from multi-tasking (e.g., speaking while also performing other tasks, such as walking, or driving).

Importantly, information relevant to the listening process is not limited to auditory channels. Visual speech information provides complementary cues that can help decode the speaker’s message, particularly in the case of perceptual ambiguity, or in the case of speech in noise. This will be discussed in greater detail below (section 1.3.3.).

1.2.3 Hearing Impairment as a Source of Perceptual Ambiguity

Hearing impairment is an important consideration in a discussion of perceptual ambiguity. When hearing is impaired, there is greater and more frequent uncertainty regarding the identity of the bottom-up speech signal, rendering the listening process consistently more difficult. Acquired hearing loss can occur for a variety of reasons, with over 1000 genes contributing to congenital hearing loss (e.g., Petit, 2006). Sources of acquired hearing loss include but are not limited to: medical illness, structural damage to the outer ear, noise-induced hearing loss, as well as normal aging. Below, I give special attention to age-related hearing loss due to its wide prevalence, and given its topical relevance for the purpose of this thesis.
It is estimated that 25% of individuals aged 65-75 years and 70-80% of individuals over the age of 75 suffer from age-related sensorineural hearing loss (National Institutes of Health [NIH], 2015). The prevalence of hearing impairment and cognitive decline is expected to rise substantially as the world shifts towards a greater population of older adults (Mathers, Smith & Concha, 2001). Age-related hearing loss has a substantial social, economic, and psychological impact. It is associated with depression, isolation and decreased quality of life (Cacciatore et al., 1999; Carabellese et al., 1993; Comijs et al., 2004; Mulrow et al., 1990), and communication difficulties place strains on family or social relationships and limit vocational participation (Kamil & Lin, 2015). It remains a leading cause of higher scores on “years lived with disability” scales in the adult years (Cohen, Labadie, & Haynes, 2005; LaForge, Spector, & Sternberg, 1992; Mulrow et al., 1990; Mathers & Loncar, 2006), second only to depression (Mathers, Smith, & Concha, 2000). Moreover, sensory impairment is also linked to decline in cognitive functioning in old age (Lin et al., 2011; 2013). Although the 20th and 21st centuries have been marked by incredible strides in life expectancy, gains in longevity that are unmatched by preservation of a reasonable standard of quality of life place a severe burden on public health systems.

Audiological assessment of hearing is crucial in the diagnosis of age-related hearing loss. It is commonly assumed that poor hearing is a threshold issue that can be addressed by increasing sensitivity through amplification. However, not all forms of hearing difficulties in older adults can be assessed with audiometric assessment, the current gold standard in assessment of hearing. Audiometric assessments involve presentation of a series of pure-tones for threshold discrimination. However, this test measures only one aspect of hearing ability; whereas audiology measures hearing sensitivity, there is currently no clinically validated tool
available for assessment of hearing acuity (i.e., the ability to cleanly resolve sounds of different frequencies on the basilar membrane). Consequently, we are, at present, unable to clinically characterize the full range of deficits in sensory impairment of hearing.

Moreover, a chief complaint amongst older adults is difficulty understanding speech in noise (Rabbitt, 1990), many of whom present with clinically normal audiograms. Such “hidden hearing loss” is thought to be related to hearing acuity and disproportionate damage to inner ear hair cells that respond to louder intensities, and are therefore not remediable with hearing aids (Plack, Barker & Prendergast, 2014). Handicap appears to be related to central auditory processing rather than audiometry (see Schneider & Pichora-Fuller, 2000 for a review), perhaps reflecting impairment of timing information (temporal coding) important for binaural speech perception (Ruggles, Bharadwaj, & Shinn-Cunningham, 2012). Furthermore, many older adults are not aware of their difficulty hearing, particularly given that sensory loss often occurs gradually. Patients may minimize hearing difficulties, blaming communication issues on factors outside the patient’s control (e.g., “I can’t hear my husband because he always mumbles”; see Heinrich, Henshaw & Ferguson, 2016 on poor reliability of self-report). Although hearing aids often provide an improvement in communication abilities for those with deficits in hearing sensitivity, compliance in hearing aid wearers is often poor for many reasons, including poor fit, perceived lack of benefit, device issues, and attitude (McCormack & Fortnum, 2013). Cochlear implants are also becoming increasingly popular in older adults, but they too, are not a panacea. Although hearing devices play an important role in management of hearing health, there is an urgent need to establish clinical tools for assessment of hearing acuity and temporal coding. In the meantime, we must look elsewhere for hearing loss rehabilitation. The next section focuses on the role of cognition in hearing.
1.3 Sources of Non-auditory Contributions to Speech Comprehension

1.3.1 Cognition in Hearing

It has been commonly assumed that hearing function is primarily determined by sensory ability. However, research is increasingly demonstrating that this is not the case. Although hearing begins at the sensory level, the perception of sound and speech involves several levels of processing in the brain. Certainly, speech perception cannot occur in the absence of cognition—after all, speech communication relies on language and meaning, a cognitive process in itself. Here, the focus is on the role of cognitive processes in resolving perceptual ambiguity.

A greater appreciation for cognitive factors in speech perception, particularly in difficult listening environments, has signified a departure from a more passive sensory process as viewed in traditional models of hearing towards a conceptualization of hearing as an active listening process (e.g., Heald & Nusbaum, 2014). Cognition may play an important role in resolving ambiguity by guiding and constraining interpretation of multiple competing perceptual hypotheses. More generally, research has explored the issue of context or linguistic knowledge in resolving various forms of semantic ambiguity (Zekveld et al., 2012; Rodd, Davis, & Johnsrupe, 2005; Rodd, Johnsrupe, & Davis, 2012; Kane & Engle, 2000). For example, knowledge about linguistic structure and sentence parsing may facilitate the use of contextual information to support speech understanding (Aydelott, Leech, & Crinion, 2011, Billig, Davis, Deeks, Monstrey & Carlyon, 2013; Rodd et al., 2005; Rodd et al., 2010). Moreover, as the fidelity of the acoustic signal becomes compromised due to masking or pathology (i.e., ambiguity of form), it becomes increasingly difficult to evaluate competing perceptual hypotheses about the identity of the bottom-up speech signal. According to Bayes’ Rule, both stimulus quality and knowledge influence perceptual decision-making; thus, more weight is allocated to knowledge-based factors.
under adverse listening conditions (Norris & McQueen, 2008), underscoring an important role for cognitive influences in speech perception.

Rabbitt (1968) was the first to highlight the role of cognitive factors in speech perception. In a pioneering study, he observed that words heard in quiet were better remembered than words heard in background noise, suggesting that background noise may interfere with rehearsal of heard items. Since then, the contribution of cognition to the listening process has more recently emerged as a highly prolific area of research (Heald & Nusbaum, 2014; Wingfield & Tun, 2007; Arlinger et al., 2009; Akeroyd, 2008; Schneider, Pichora-Fuller, & Daneman, 2010 for reviews). The majority of research to date has focused on identification of specific cognitive domains that account for individual variability in speech comprehension; these contributions typically emerge most strongly under adverse listening conditions, either as a result of noise, pathology, or normal auditory aging.

A number of distinct cognitive processes have been highlighted in the listening process. Executive functioning may play a role in attentional processes related to speech communication, including directing attention to a particular speaker, inhibiting irrelevant speech information, as well as integrating the acoustic signal with previous knowledge (e.g., Tamati, Gilbert, & Pisoni, 2013; Woods et al., 2013; Tun, O’Kane & Wingfield, 2002). Processing speed may be particularly important for understanding speech at fast rates (e.g., Wingfield, Tun, Koh, & Rosen, 1999; Gordon-Salant & Fitzgibbons, 2001), and slower processing speed is linked to difficulty understanding speech in noise (e.g., Pronk et al., 2013; Tun & Wingfield, 1999).

The domain of working memory has received the most attention in recent years, with a number of studies linking larger working memory capacity to better speech recognition scores in
noise or when competing talkers are present (Szenkovits, Peelle, Norris, & Davis, 2012; George et al., 2007; Sorqvist & Ronnberg, 2012; Rudner, Ronnberg & Lunner, 2011; Wingfield & Tun, 2007; Akeroyd, 2008; Rudner & Lunner, 2014, Heald & Nusbaum, 2014, Talmati et al., 2013; Zekveld, Rudner, Johnsrude, Heslenfeld & Ronnberg, 2012; Humes, Lee, & Coughlin, 2006). Moreover, older adults with larger working memory spans may be better able to adapt to age-related difficulties with understanding speech in noise (Gordon-Salant & Cole, 2016). It appears that working memory allows listeners to compensate for increased processing demands involved in difficult listening conditions. Working memory processes may also facilitate subvocal rehearsal to assist in word identification. Szenkovits and colleagues (2012) found that stronger working memory was correlated with increased premotor and motor cortex activation when listening to monosyllabic pseudowords.

However, there is some evidence to suggest that the contribution of working memory processes to successful speech perception, as proposed in some models (i.e., Ronnberg et al., 2013), may be overstated. A substantial proportion of this body of research has relied on a language-based version of the traditional Digit Span test, termed Reading Span, as a measure of working memory (Davies-Venn & Souza, 2014; Tun, Wingfield, & Stine, 1991; Akeroyd, 2008; Zekveld et al., 2012, Zekveld, Rudner, Johnsrude & Rönnberg, 2013, Besser, Koelewijn, Zekveld, Kramer & Festen, 2013). The Reading Span test (Daneman & Carpenter, 1980) is similar to the Digit Span test, but instead requires subjects to recall the last word for a set of sentences, while also performing a simultaneous semantic judgment task in between presentation of sentences (the sentence set size increases over the course of the test). Thus, although the reading span test has been robustly associated with speech recognition outcome scores (but see Fullgrabe & Rosen, 2016b), the test is likely also contaminated by contributions from other
cognitive processes. These processes include verbal episodic memory, particularly during longer span trials, inhibition of non-target words, as well as processing speed and executive functioning, given the complex, multi-tasked nature of the test. Moreover, a number of studies have failed to find a correlation between working memory, particularly non-reading span measures, and speech perception tasks (Schoof & Rosen, 2014; Humes & Coughlin, 2009; Fullgrabe & Rosen, 2016a; see Fullgrabe & Rosen, 2016a for a recent challenge of the documented relationship between reading span and speech in noise). Moreover, research in this area has generally relied exclusively on indirect, correlational designs. Subsequently the extent of the contribution of working memory to the speech perception process remains unclear.

In summary, while it is clear that knowledge-guided processes play an important role in speech comprehension, the precise cognitive architecture subserving speech comprehension remains an ongoing subject of debate. The relatively modular approach of cognition utilized thus far in the literature is limited in the sense that these processes are unlikely to work in isolation (as is perhaps evident in the robustness of results obtained with the cognitively impure Reading Span test). Although it is certainly important to establish the relative contributions of various cognitive processes to the listening process, cognitive processes tend to be highly correlated with one another (e.g., Engle, Tuholski, Laughlin & Conway, 1999) and thus are unlikely to operate independently of one another. Subsequently, a greater understanding of how distinct cognitive factors may operate together, either in tandem or as a composite, is necessary in order to gain a clear view of cognitive hearing science.

The role of cognition in hearing holds particularly important implications for our understanding of normal age-related hearing loss. Notably, cognitive abilities, particularly, processing speed and executive functioning, tend to decline with age in tandem with sensory
function (Salthouse, 1996; Craik & Salthouse, 2007), and research has also demonstrated slower lexical access in connection with aging (Carroll et al., 2016). There appears to be a clear link between sensory and cognitive decline, although its causal nature remains unclear (e.g., Lin et al., 2011; 2013), and these declines appear to contribute to older adults’ difficulty understanding speech in noise (Humes & Dubno, 2010). Given that such “hidden hearing loss” is not remediable with assistive listening devices, alternative means for rehabilitation of age-related hearing loss are urgently needed. Nascent avenues for exploration include both auditory (at both the sensory and perceptual level), as well as cognitive training, which is one focus of this thesis.

1.3.2 Cognition: Listening Effort

The vast majority of studies noted thus far rely on behavioural measurements of speech perception in order to assess the contribution of distinct cognitive factors to the listening process. As mentioned above, this is frequently approached through examination of how individual differences in cognitive abilities account for variability in speech identification scores. However, this approach is limited in the sense that it fails to account for the fact that the same behavioural outcome (e.g., word or sentence identification) can be achieved with varying degrees of cognitive ease (i.e., how much mental energy the listener must exert in order to arrive at the identity of the speech token). This cognitive or effort variability in how participants may arrive at the speech identification process is not necessarily captured by word report or sentence identification data.

The emergent notion of “listening effort” marks an important contribution to this field of research by providing a complementary construct for quantifying variability in task-based perceptual load, abstracted from particular cognitive domains. According to Johnsrude and Rodd (2016), listening effort should be understood as an interaction between perceptual demands (i.e.,
the degree of perceptual ambiguity), and the cognitive resources (demanded by the task) that an individual brings to the listening situation. For example, the greater the perceptual ambiguity, the greater the degree of listening effort required; individuals with superior processing speed or working memory, for example, may be able to discern the identity of the same speech token with relatively less effort than someone with a poorer profile of cognitive abilities. However, identification of the speech token may not be achieved if ambiguity is too high (e.g., highly degraded or distorted speech), or if an individual’s cognitive resources to compensate for the perceptual ambiguity have been exceeded (e.g., speech may be spoken too quickly for one listener to understand, but not another).

There are a number of ways to measure listening effort. Although dual-task approaches have been used to index listening effort (e.g., Gosselin & Gagné, 2011; Pals, Sarampalis, & Baskent, 2012), these approaches are indirect since listening effort is assessed by changes in performance on a secondary task to assess the degree of effort on the primary speech-related task. Moreover, this is confounded by uncertainty regarding the required degree of cognitive overlap between the two tasks in order to obtain an accurate measurement of listening effort (Johnsrude & Rodd, 2016). There are also two primary physiological approaches, although they are also presently somewhat limited in their application: The first approach involves use of pupilometry (and assumes that pupil dilation increases in accordance with listening effort; Kahneman & Beatty, 1966; Piquado, Isaacowitz, & Wingfield, 2010; Sevilla, Maldonado, & Shalóm, 2014; Zekveld, Kramer, & Festen, 2011; Zekveld & Kramer, 2014). However, this approach is also limited by uncertainty regarding the exact relationship between pupil changes and degree of listening effort. Moreover, the possibility that pupilometry is also confounded by responses related to task engagement or attention (Franklin, Broadway, Mrazek, Smallwood, &
Schooler, 2013) or arousal (Bradshaw, 1967; Brown, van Steenbergen, Kedar, & Nieuwenhuis, 2014), has not been sufficiently ruled out. Functional magnetic resonance imaging (fMRI) presents another way to quantify processing load, although it too, provides only an indirect measure and cannot distinguish between processing load and cognitive demands. However, fMRI may be advantageous in allowing for exploration of the manner in which listening effort may be modulated by attention (Johnsrude & Rodd, 2016; Wild et al., 2012).

At present, current applications and measurements of listening effort continue to elude their precise quantification, and it remains difficult to disentangle processing load and cognitive resources (Johnsrude & Rodd, 2016). However, continued developments in the area of listening effort may be in permitting quantification of the broader, domain-general, and ecologically relevant cognitive factors that are also instrumental to the listening process.

1.3.3 Visual Speech Contributions to Speech Perception

Visual speech contributions to speech understanding include non-linguistic sources of information, such as non-verbal gestures or accompanying hand movements. Lip movements provide a wealth of speech information, and are frequently supplemented by the accompanying movements of other parts of the face, including the eyes, cheeks, and forehead (e.g., Munhall et al., 2004). In this section, I use “visual speech” to refer to facial gestural information, that is, the facial movements that accompany speech production (also commonly referred to as ‘speechreading’), unless otherwise noted.

Speech identification is supported by the large degree of redundancy in the auditory-visual speech signal. In many languages, visual speech segments (i.e., ‘visemes’) are complementary to auditory segments (i.e., phonemes), such that distinctions that are more
difficult to hear tend to be more easily seen, and vice versa (Rosenblum, 2008; Summerfield, 1987). Visual cues are beneficial even in ideal listening conditions (Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987; Remez, 2005), but play an especially important role in helping to recover the message content and constrain interpretation when the acoustic signal is masked, degraded, or distorted (e.g., MacLeod & Summerfield, 1990; Reisberg, et al., 1987). Visual cues can boost the signal-to-noise ratio (SNR) relative to auditory-only presentations (Sumby & Pollack, 1954; Macleod & Summerfield, 1990), and a 1 dB increase in SNR can correspond to a maximum improvement of 5-10% in intelligibility (Grant & Braida, 1991; G. A. Miller, Heise, & Lichten, 1951). Moreover, auditory speech has also been shown to facilitate speechreading when the visual speech stimulus is ambiguous (Baart & Vroomen, 2010).

Visual speech is particularly important when hearing is impaired (i.e., where perceptual ambiguity occurs in greater frequency and severity), with some individuals relying near-exclusively on facial gestural information for speech identification. Interestingly, in congenitally deafened cats, the visual cortex responds to speech in a manner analogous to auditory cortex activity in normal hearing individuals (e.g., Lomber et al., 2010), indicative of extensive crossmodal reorganization. Moreover, in normal hearing individuals, there appears to be a bidirectional, crossmodal interaction between auditory and visual speech information. In one study, subjects who silently lipread from a talker and then heard that same talker in a speech-in-noise task performed better on the latter task than subjects who lipread from one talker and then heard a different talker (Rosenblum, Miller and Sanchez, 2007). This indicates that familiarity with a talker’s facial gestures, even in the absence of auditory input, can facilitate auditory speech recognition in the presence of background noise. It has also been shown that the auditory
cortex is active during silent lipreading, demonstrating prelexical influences (i.e., at the phonetic level) of visual speech on the auditory speech signal (Calvert et al., 1997).

The time-varying dimensions of the auditory and visual signals may provide a basis for audiovisual integration. This may be achieved through amodal speech representations on the basis of kinematics and dynamics of underlying articulatory behavior (Bernstein, Auer, & Moore, 2004; Liberman & Mattingly, 1985; Rosenblum, 1994; Schwartz, Robert-Ribes, & Escudier, 1998), or more specifically, dynamic temporal information obtained from spectral and temporal patterns of formant frequencies as they change over time (Lachs and Pisoni, 2004; Lander, Hill, Kamachi, and Vatikiotis-Bateson, 2007; Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Grant & Seitz, 2000; Yehia et al., 1998). Visual information appears to “speed up” neural processing of auditory speech, via reduced response latency of auditory evoked potentials (Besle, Fort, Delpuech, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005), possibly suppressing competing lexical responses.

1.4 fMRI Techniques

Functional magnetic resonance imaging (fMRI) has emerged as a very popular tool in both cognitive and hearing science. Pioneered by Ogawa and colleagues (1990), fMRI capitalizes on the hemodynamic response of neurons: neuronal regions that are active in a particular task consume more oxygen at a rate greater than that of neurons that are relatively less active in a particular task. The peak hemodynamic response occurs 5-6 seconds following stimulus presentation, coinciding with an over-recruitment of oxygen to replenish stores. As oxygenated and deoxygenated blood have different relative magnetic properties, differences in magnetic signal variation of cerebral blood flow during this rebound local tissue oxygenation process can
be captured through fMRI, referred to as the blood-oxygen level dependent (BOLD) signal. The BOLD signal is expressed as a voxel-based pattern of activity, which measures the position of activity in three-dimensional space across time.

FMRI is advantageous in providing good spatial resolution, but lacks good temporal resolution. Subsequently there has been increased interest in coupling it with electroencephalogram (EEG) or near-infrared spectroscopy (NIRS) in order to capture both dimensions with good sensitivity (e.g., Ritter and Villringer, 2006; Babiloni & Astolfi, 2014). It is also advantageous compared to positron emission tomography (PET) since it does not require ingestion of a radioactive substance. However, it is important to recognize that the BOLD response provides only an indirect, correlated measure of neural activity. Moreover, since conventional BOLD imaging is most sensitive to tissue in draining veins relative to the active tissue, the observed signal may be evident in locations that may be distal to tissue activity (Disbrow et al., 2000).

A notable confound posed for auditory imaging arises from acoustic fMRI scanner noise, which may itself stimulate auditory areas, thereby confounding interpretation of imaging results. Subsequently, Hall and colleagues (1999) developed a sparse imaging design routinely used in auditory neuroimaging research in which stimuli are presented in between scans, capitalizing on the 5-6 second delay in the peak hemodynamic response.

fMRI techniques are advantageous for exploring neural substrates of language-based processing, allowing for comparison of differences in neural activity across conditions of interest. As mentioned in the previous section, they provide an advantage over dual-task methods by allowing us to measure manipulations of interest more directly, rather than relying on changes
in performance on a secondary task. Here, I employ fMRI techniques in order to directly investigate how the same audiovisual speech stimulus is differentially processed in the presence or absence of selective attention (Chapter 4; which is further outlined in the next section).

1.5 Outline of Thesis

This thesis focuses on cognitive and visual speech contributions to speech communication, particularly in noisy or degraded listening environments. This is explored through a review paper and two original research studies. In chapter 2, I critically examine the relationship between age-related hearing loss and cognitive decline (published as Wayne & Johnsrude, 2015). More specifically, research has documented a link between declines in hearing and cognition with age; however, the causal relationship between them is unclear. In addition to reviewing the proposed causal mechanisms for the relationship, I also present methodological criticisms of existing research in the field, pointing to an urgent need to consider the ways in which measurements of cognition rely on hearing and vice versa. In addition, I argue that future developments in this area must focus on the interconnected nature of hearing and cognition in order to improve understanding of how impairments in one domain affect the other. In chapter 3, I highlight the role of working memory in speech in noise understanding (Wayne et al., 2016). Here, I administer cognitive training, specifically working memory training to older adults. I aim to directly validate the role of working memory in speech in noise perception by demonstrating that training-related improvements in working memory lead to better performance on tests of speech in noise. In chapter 4, I use neuroimaging techniques to investigate whether successful integration of the auditory and visual speech signals relies on attention, particularly when the auditory component is degraded. Taken together, the goal of these three papers is to advance our
understanding of the underlying cognitive architecture supporting speech comprehension in challenging listening conditions.
Chapter 2 - A Review of Causal Mechanisms Underlying the Link Between Age-related Hearing Loss and Cognitive Decline
2.1 Introduction

Age-related changes in sensory sensitivity and acuity, and in cognitive processing, are among the most robust findings in psychology. Such declines will become more common as the world’s population shifts towards a greater number of older adults (Mathers, Smith & Concha 2000; World Health Organization 2012). Declines in hearing and cognition are functionally interdependent, since there is no sharp division between sensation and perception, and cognition. A growing body of research highlights the role of cognitive abilities in supporting speech comprehension, particularly when the speech signal is ambiguous due to background noise, semantic ambiguity, or unusual talker characteristics (e.g., accents; see Akeroyd, 2008; Heald & Nusbaum, 2014; Wingfield & Tun, 2007 for reviews). The mapping of ambiguous speech sounds onto the corresponding linguistic representations is a knowledge-guided process, and probably depends on working memory, executive functioning, and processing speed for efficient operation. Although cognitive contributions to everyday speech comprehension are well-established, whether peripheral hearing level, and general (i.e., not specifically auditory) cognitive function are somehow linked in old age is not clear.

Both age-related hearing loss and cognitive decline are associated with communication difficulties, isolation, decreased quality of life, and depression (Bozeat, Ralph, Patterson, Garrard & Hodges 2000; Cacciatore et al. 1999; Carabellese et al. 1993; Comijs et al. 2004; Holmen, Ericsson & Winblad 2000; Mulrow et al. 1990). Recently, Lin and colleagues have suggested that hearing loss may play a causal role in precipitating cognitive decline (F.R. Lin et al. 2013; F.R. Lin, Thorpe, Gordon-Salant & Ferrucci 2011; F.R. Lin 2011), and the relationship between hearing loss and cognitive decline has attracted increased attention in recent months (Albers et al. 2015; Panza, Solfrizzi & Logroscino 2015; Barnabei et al 2014). A clearer understanding of
the nature of the relationship between hearing loss and cognitive decline is critical if we are to minimize their impact, either in isolation or together, on quality of life, and to develop effective prevention and rehabilitation strategies (F.R. Lin et al. 2013; Pichora-Fuller 2003). If hearing loss does contribute to cognitive decline, a case may be made to offer hearing aids or other rehabilitative strategies much earlier in the course of auditory decline, and to promote their use more aggressively; this has important health care and public policy implications. Further, the degree to which such interventions are effective will depend on the size of the effect for the relationship between age-related hearing loss and cognitive decline; a larger effect size makes it more likely that such interventions will have a clinically significant impact.

In this paper, I critically examine the evidence for a link between hearing loss and cognitive decline. I begin by explaining what I (and most others, in my opinion) mean by “hearing loss” and “cognitive decline.” Then, I assess empirical support for a link, taking effect sizes into account. I follow this with a discussion of the evidence for and against several different possible causal relationships. Overall, I conclude that evidence supports a significant, but reliable relationship between age-related hearing loss and cognitive decline. In addition, I conclude that the link cannot be satisfactorily explained by any single mechanism; I end with an alternative view that incorporates multiple possible mechanisms, taking the strengths of each.

2.2 Operational Definitions of Hearing Loss and Cognitive Function, and Their Limitations

The ways in which hearing and cognition are operationally defined and assessed will influence the nature of the apparent relationship between them. Hearing loss in studies that have evaluated links with cognitive function has almost exclusively been measured using pure-tone audiometry (PTA), in which detection thresholds for pure tones across a range of frequencies are measured monaurally, yielding indices of hearing sensitivity (e.g., Anstey, Luszcz & Sanchez
Hearing loss is typically defined as at least an average 25-dB HL elevation in detection thresholds across frequency regions necessary for speech comprehension (0.5 – 4 kHz; note that selected frequency regions vary between studies), and less typically as thresholds at the worst measured frequency.

Older adults commonly complain of difficulty understanding speech in noise, although audiometric measurements are often normal. Indeed, functionally relevant loss can occur in the absence of elevation of pure-tone thresholds (Frisina & Frisina 1997; He, Dubno & Mills 1998; Hopkins & Moore 2011; Schneider 1997; Snell & Frisina 2000; Kujawa & Lieberman 2009). In fact, peripheral hearing ability encompasses a number of factors other than pure-tone sensitivity, including frequency selectivity, temporal coding fidelity, intensity resolution and loudness, among others, which are not commonly measured.

Auditory filters appear to broaden with age, reducing frequency selectivity such that sounds that are cleanly resolved in the young, normally hearing ear are ‘blurred together’ in older adults with hearing loss (Glasberg, Moore, Patterson & Nimmo-Smith 1984; He, Mills & Dubno 2007; Patterson, Nimmo-Smith, Weber & Milroy 1982). Reduction in frequency selectivity is thought in part to reflect a loss of the frequency-specific gain generated by the outer hair cells, which are vulnerable to damage from noise exposure (e.g., Fernandez et al., 2015; Liberman, Liberman & Maison, 2014; see also Stamper & Johnson, 2015). High-threshold auditory-nerve fibers (medium- and low-spontaneous rate fibers), which are thought to be important for temporal coding fidelity (and hence representation of precise frequency information), also appear to be disproportionately damaged by noise (compared to high spontaneous-rate fibers and hair
cells; Kujawa & Liberman 2009; Ruggles, Bharadwaj & Shinn-Cunningham 2012; Plack, Barker & Prendergast 2014). This cochlear synaptopathy, undetectable using conventional audiometry, would particularly degrade transmission of moderate to intense sounds and may underlie the speech in noise difficulties experienced by older listeners. Although age-related deficits in suprathreshold processing and cochlear synaptopathy appear to precede threshold elevations (e.g., Surgeyenko et al., 2013; Kujawa and Liberman, 2015), pure-tone audiometry remains the gold standard for clinical assessment of hearing loss.

Non-audiometric measurements of hearing loss are relatively rare in the literature relating it to cognitive decline, but include speech-in noise-performance (Gennis et al. 1991; Gates et al. 1996; Wong, Yu, Chan, Tong 2014) and psychophysical measures of temporal resolution, including temporal gap detection, temporal order identification, and temporal masking (Humes, Busey, Craig & Kewley-Port 2013). Such measures may well pick up deficits that correlate with cognitive decline and a focus on audiometric measures may therefore underestimate the relationship with cognitive function (Humes et al. 2013).

Different cognitive processes are differentially affected by aging (Jagust 2013), with domains such as processing speed and memory more subject to decline, compared to language and general reasoning (Salthouse 1996). Studies have generally employed inconsistent definitional criteria for cognitive decline. Some studies operationalize cognitive decline (usually dementia) as a clinical impairment relative to a normal control group through clinical consensus and the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM). Other studies index cognitive decline through change in performance on tests of cognition. Note that these may not reflect only clinical declines, as the distinction between normative and clinical declines in cognition (i.e., dementia) is based on functional criteria, in addition to psychometric
deficits. Measurement tools vary across studies, ranging from general cognitive screening tests, such as the Montreal Cognitive Assessment (MoCA; Nasreddine et al. 2005), the Mini-Mental Status Exam (MMSE; Folstein et al. 1975), and Modified Mini-Mental State Test (3MS; Teng & Chiu 1987), to sampling of domain-specific processes through specific tests that assess language, memory, and/or general reasoning abilities, as well as processing speed, selective attention and other executive functions (see table 2-1 for a description of these three cognitive screening tools and information on cut-off scores, sensitivity, and specificity). Heavy reliance on cognitive screening tools such as the MoCA, MMSE and 3MS may capture only limited variability in a normally aging (non-clinical) population; ceiling effects could potentially lead to an underestimation of the true relationship between hearing loss and cognitive decline.
<table>
<thead>
<tr>
<th>Cognitive Testing Instrument</th>
<th>Description</th>
<th>Range of Scores</th>
<th>Interpretation, Sensitivity and Specificity, and Notes</th>
</tr>
</thead>
</table>
| Montreal Cognitive Assessment (MoCA) (Nasreddine et al. 2005) | 10-minute cognitive screening tool for mild cognitive impairment (MCI; which often progresses to dementia); | 0 - 30 | < 26 indicative of MCI  
Sensitivity: 100%; Specificity: 87% (for detecting early dementia)  
2 points added for 4-9 years of education, 1 point for 10-12 years of education.  
Specific age- and education-adjusted norms available (Rossetti, et al. 2011) |
| Mini-Mental Status Exam (MMSE) (Folstein et al. 1975) | 10-minute cognitive screening tool for dementia | 0 - 30 | ≤ 9 – severe impairment; 10–18 – moderate impairment; 19-24 – mild impairment  
Sensitivity: 71-92%; Specificity: 56-96% (Cullen et al. 2014)  
Age- and education-adjusted norms available (Crum, et al. 1993) |
| Modified Mini-Mental State Test (3MS) (Teng & Chiu 1987) | 15-minute cognitive screening tool for dementia; expansion of MMSE, including more detailed scoring system. | 0 - 100 | < 79 – cognitive impairment; < 48 – severe impairment  
Sensitivity: 83-93.5%; Specificity: 85-90% (Cullen et al. 2014)  
Age- and education-adjusted norms available (Jones et al. 2002) |

Table 2-1: Description, interpretation, and psychometric properties of cognitive screening tools used in assessment of the relationship between hearing loss and cognitive decline. Note that all three tools involve tasks related to orientation, attention, calculation, recall, naming, and repetition.
It is difficult to measure cognition without relying on intact hearing, and vice versa. Some authors have maintained that cognitive measures are free of influence from audition when administered by experienced clinicians (F. R. Lin et al. 2011c; F. R. Lin et al. 2013; F. R. Lin, et al. 2011a). However, difficulty in hearing test items during administration may lead to poor performance (see Schneider, Daneman & Murphy 2005) and overestimation of the level of cognitive impairment in subjects with hearing loss (Jorgensten 2012; Ohta, Carlin & Harmon 1981; Dupuis et al. 2014), particularly if cognitive measures that rely on auditory function are used (Granick, Kleban & Weiss 1976; Thomas et al. 1983; van Boxtel et al. 2000; Wong et al. 2014). Moreover, the influence of hearing on cognitive measures may be more apparent when cognitive load (or ‘listening effort’) is high, placing increased demands on shared processing resources (e.g., Murphy et al. 2000). Although evidence suggests that the link between age-related hearing loss and cognitive decline persists when auditory items are removed from cognitive screening tests (Dupuis et al. 2014), and when non-auditory tests of cognition are used (Lin et al. 2013, Wong et al. 2014), such confounds may overestimate the strength of the association.

In addition, if hearing loss and cognition do decline simultaneously (i.e., a common etiology as opposed to a causal relationship for decline), the greater sensitivity of tests in one domain (in either hearing or cognition) could result in the appearance of deficits in that domain prior to observed deficits in the other. Such differences in measurement sensitivity could lead to an illusory causal relationship.

In the following section, I review the evidence for a link between hearing loss (largely assessed using PTA) and cognitive decline. I reviewed original, peer-reviewed research articles published prior to March 2015 that were identified either through Pubmed using a conjunction of
the search terms “age-related hearing loss” and “cognitive decline” or the conjunction of “older adults”, “hearing loss” and “cognition”, or were identified in the References sections of these papers. I excluded papers in which hearing loss or cognition were not measured objectively (for example, through self-report).

2.3 Is There a Link Between Hearing Loss and Cognitive Decline?

Early research reported that hearing loss was more common in those with dementia than without after controlling for age (Herbst & Humphrey 1980; Hodkinson 1973; Peters, Potter & Scholer 1988; Uhlmann, Larson, Rees, Koepsell & Duckert 1989; see also, Kay, Roth & Beamish 1964; Weinstein & Amsel 1986, who did not control for age). Work by Gates and colleagues (1996) showed that central auditory dysfunction (as measured by a closed-set synthetic sentence identification task with an ipsilateral competing talker) significantly increased individuals’ risk for dementia and decline on the MMSE relative to controls, although this is unsurprising given that speech-in-noise performance depends both on hearing status and on cognitive function (see also, Gates et al. 2010; Jerger, Jerger & Pirozzolo 1991; Hällgren, Larsby, Lyxell & Arlinger 2001, Moore et al. 2014). The link between age-related declines in hearing and cognition were extended to non-clinical groups, using cognitive screening measures (e.g., Thomas et al. 1983). Kiely and colleagues (2012) found that probable cognitive impairment at baseline (as diagnosed through a score of <24 on the MMSE) was significantly associated with higher PTA scores in speech-relevant frequency regions (β = 3.91, 95% CI = 2.05-5.77), also predicting faster annual rates of change in PTA averages (β = 0.40, 95% CI = 0.12-0.68; see also Wong et al. 2014). In a study of 156 old and very old adults (70 – 103 years), Lindenberger and Baltes (1994) found that individual differences in hearing sensitivity (as measured through audiometry) accounted for a significant 34.5% of the variance on a battery of cognitive tests
(spanning processing speed, memory, verbal knowledge, and fluency: the correlation between audiometric sensitivity and cognitive function was 0.58; see also Baltes and Lindenberger 1997).

In a series of recent papers, Frank Lin and colleagues have suggested that hearing loss (defined as a pure-tone average threshold exceeding 25 dB hearing loss at 0.5 – 4 kHz) may cause or exacerbate cognitive decline (F. R. Lin et al. 2011b; F. R. Lin et al. 2011a). These researchers controlled for a variety of factors, including age, sex, income, education, cardiovascular events, stroke, diabetes, hearing aid use, depression, and smoking, and therefore concluded that the relationship is not fully explained by general health factors or intelligence.

Lin and colleagues (2011b) demonstrated that patients developing dementia (diagnosed through clinical consensus using established criteria) exhibited PTA average thresholds that increased, on average, 0.52 dB per year compared to 0.27 dB in the non-demented group, after adjusting for age. In another study of 1162 older individuals followed for six years, Lin and colleagues (2013) observed that a greater severity of hearing loss at baseline was associated with increased risk for cognitive impairment at follow up (as defined by a 3MS score of less than 80 or a decline of 5 points on the 3MS from baseline), after adjusting for age, sex, education, and cardiovascular risk factors. Hearing loss was associated with lower baseline scores on the 3MS (a difference of -0.75 points compared to those with normal hearing) and on the digit symbol substitution test (DSST; difference of -0.92 points). Individuals with hearing loss lost 0.19 more points per year on the 3MS and 0.2 points per year on the DSST compared to controls. This finding is mirrored in an additional study in which hearing loss was significantly associated with DSST scores \((r = -.18; the negative correlation reflects the fact that high test values reflect better performance on cognitive tests, but worse performance on sensory tests; F. R. Lin 2011, see also Suprenant & DiDonato 2014).
Another study by Lin and colleagues (2011a) evaluated the relationship between hearing loss and declines on a battery of cognitive tests as a function of age. Cognitive tests included the MMSE and tests of memory (free recall) and executive function (Stroop, Trail Making A and B, Letter and Category Fluency), as well as reading (American National Adult Reading Test; AMNART). All scores except the AMNART and letter fluency declined linearly with age, and a 25-dB hearing loss corresponded to an age difference of 6.8 years on tests of cognitive function. Similarly, in a recent study of 894 older adults, Bush and colleagues (2015) found that PTAs were weakly, but significantly, related to MMSE scores, as well as to measures of processing speed (DSST and DSST Copy, Letter and Pattern Comparison, Useful Field of View), executive function (Stroop and Trails B), working memory (Digit Span, Spatial Span), and verbal memory (Hopkins Verbal Learning Test).

Although the findings in the studies by Lin and colleagues are statistically significant, the sample sizes are large, and the effects are small. For example, Lin and colleagues (2013) defined cognitive impairment as a 3MS Test score of <80/100 or decline of >5 points from baseline. They noted that average annual score changes on the 3MS were 0.65 points for individuals with hearing loss vs. 0.46 points for individuals with normal hearing. Although this is a 41% increase in the hearing-loss group, it is less than a 0.2% difference in the total 3MS score. Furthermore, it is difficult to infer causal relationships on the basis of either cross-sectional or longitudinal correlational designs where parameters of interest are not directly manipulated. Cross-sectional designs are particularly problematic, since they are used to infer information about individual risk for cognitive decline from between-group differences. The difference in cognitive scores across individuals at a single time point may not reflect changes in cognitive scores within individuals over time (see Hofer, Berg and Era 2003, for a similar point).
Moreover, the link between hearing loss and cognition is not observed at all (M.Y. Lin et al. 2004; Anstey, Luszcz & Sanchez 2001a; Gennis et al. 1991) or is observed to be weak in other investigations (in similar magnitude to the findings of F. R. Lin and colleagues; Tay et al. 2006; Thomas et al. 1983; Valentijn et al. 2005). Gennis and colleagues (1991) did not find a correlation between baseline PTAs or speech in noise scores and change in scores on immediate and delayed memory in older adults, after adjusting for age and sex. In a longitudinal study by M. Y. Lin and colleagues (2004) hearing loss was defined as an inability to hear at 40 dB HL or lower at 2 kHz (using a hand-held portable audiometer; note that this criterion is more strict than that used in other studies reported above). After adjusting for age, BMI, education, smoking, walking speed, handgrip strength, social network and a host of health factors, hearing sensitivity at 2 kHz was not significantly associated with decline in 3MS score (as defined by the amount of change in 3MS score exceeding one standard deviation from the mean) over an average period of 4.4 years. In another longitudinal study of over 2000 individuals that controlled for age, declines in hearing (as measured through audiometry as an average increase in thresholds of 10 dB HL or more in the better ear) were not associated with declines on tests of verbal cognition or memory (Anstey, Luszcz & Sanchez 2001a). However, Anstey and colleagues (2001a) reported a relationship between hearing and memory ($\eta = .028$, equivalent to $r = -.16$, consistent with correlations noted in other studies) when a more liberal threshold of decline (5 dB threshold increase instead of 10 dB) was used.

Similar to the findings of Lin et al. (2011), a cross-sectional study by Tay et al. (2006) using the MMSE and pure tone audiometry found that cognitive function was significantly, but weakly, correlated with hearing function after controlling for age ($r = -0.12$; compared to $r = -.18$ in Lin et al. 2011). Similarly, Thomas et al. (1983) found a correlation of $r = -0.19$ between the
two measures after controlling for age in a cross-sectional study. In a longitudinal study, Valentijn and colleagues (2005) observed that change in PTA-measured thresholds over time significantly predicted change in score on the letter-digit substitution test and change in memory performance. Again, the magnitude of these changes was rather small (less than 1% of the total test score), although a larger effect between change in PTA-measured thresholds and cognitive function was seen on a test of naming speed for colours and words (approximately a 13% change in latency compared to the average latency at baseline).

An important study by Humes and colleagues (2013) provides further compelling evidence for the existence of a link between age-related declines in sensory and cognitive function. The authors computed global sensory and cognitive processing scores derived from a 60-hour battery of tests of cognitive and sensory function in 245 young, middle-aged, and older adults. Cognitive tests included measures of verbal and non-verbal intelligence, working memory, processing speed, incidental learning, and sensory measures included threshold sensitivity (including PTA), gap detection, temporal-order judgments and other measures of acuity in hearing, vision, and tactile perception. Participants were required to pass an auditory, visual, and tactile screening procedure in order to reduce the confounding effects of peripheral sensory deficits on performance of cognitive tests. The authors used factor analysis to derive global (composite) sensory and cognitive factors, and observed that component weights for auditory sensory tests in the global sensory processing common factor were reasonably high (.57 - .76), suggesting that auditory functioning played a role in subsequent analyses. They found that reduced global sensory performance was significantly associated with poorer global cognitive function, irrespective of age ($r_p = -.53$). Perhaps most interestingly, whereas the correlation between age and global cognitive function ($r = -.55$) was non-significant when global sensory
processing was accounted for ($r_p = -.05$), it remained significant when accounting only for hearing sensitivity (as measured through audiograms; $r = -.41$). Thus, declines in sensory function generally may mediate the relationship between increasing age and declining cognitive function, but this cannot be attributed only to hearing sensitivity. The authors argue that use of simple measures of hearing sensitivity (such as audiograms) are insufficient to pick up sensory decline related to cognitive decline.

Taken together, the studies discussed in this section indicate that the relationship between hearing loss and cognitive decline appears reliable, but weak when hearing is measured using PTAs ($r = -.12$ to -.18, accounting for only 1-4% of the total variance). Sample size in many of the studies reported above is large and null findings have typically been reported with the use of more conservative diagnostic criteria for hearing loss or cognitive decline. However, evidence from Humes and colleagues (2013) suggests that exclusive reliance on measures of hearing sensitivity (i.e., audiometry) may underestimate the strength of the link.

**2.4 Hypothesized Relationships between Age-Related Hearing Loss and Cognitive Decline**

The work reviewed in the previous section suggests a small link between hearing loss and cognitive function, but doesn’t speak to the nature of the link – does hearing loss cause cognitive decline, or does cognitive decline result in hearing loss, or does some other factor cause both types of decline? Several possible relationships have been postulated (see Figure 2-1). Cognitive decline may reduce the cognitive resources that are available for auditory perception, manifesting as hearing loss (the “cognitive load on perception” hypothesis). In contrast, Lin and colleagues (F.R. Lin et al. 2013; F.R. Lin et al. 2011c; F.R. Lin 2011) have suggested that hearing loss causes cognitive decline that is either permanent (the “sensory-deprivation”
hypothesis; CHABA 1988; Lindenberger & Baltes 1994), or potentially remediable (the “information-degradation” hypothesis; CHABA 1988; Schneider and Pichora-Fuller, 2000; Pichora-Fuller 2003). Another possibility is that a third factor causes both declines (the “common-cause” hypothesis; CHABA 1988; Baltes & Lindenberger 1997; Lindenberger & Baltes 1994). Below, I review the evidence for each of the possible relationships (see Figure 2-1-A-D and Table 2-2 for a summary).

Figure 2–1: A summary of the four hypotheses for the causal relationship between age-related hearing loss and cognitive decline. A) Cognitive Load on Perception Hypothesis: Cognitive decline leads to audiometric hearing loss; B) Information Degradation Hypothesis: Audiometric hearing leads to diminished (but reversible) cognitive performance; C) Sensory Deprivation Hypothesis: Audiometric hearing loss causes cognitive decline; D): Common Cause Hypothesis: Both audiometric hearing loss and cognitive decline are caused by a third, common factor.
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Explanation</th>
<th>Supporting Evidence</th>
<th>Potential Mechanisms</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Load on Perception Hypothesis</td>
<td>Cognitive decline causes sensory decline</td>
<td>Minimal</td>
<td>Unclear</td>
<td>Lack of supporting evidence</td>
</tr>
<tr>
<td>(CHABA 1988; Lindenberger &amp; Baltes 1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory Deprivation Hypothesis</td>
<td>Sensory decline causes cognitive decline</td>
<td>Consistent with age-related perceptual deficits, increased cognitive effort, and contextual benefits</td>
<td>Depression as mediator or moderator; decline-related compensation</td>
<td>Sensory declines earlier in lifespan do not result in cognitive decline; long-term mechanism unclear</td>
</tr>
<tr>
<td>(Perceptual Hypothesis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CHABA 1988; Lindenberger &amp; Baltes 1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Degradation Hypothesis</td>
<td>Impoverished perceptual input causes cognitive decline</td>
<td>Consistent with age-related perceptual deficits, increased cognitive effort, and contextual benefits</td>
<td>Depression as mediator or moderator; decline-related compensation</td>
<td>Sensory declines earlier in lifespan do not result in cognitive decline; long-term mechanism unclear</td>
</tr>
<tr>
<td>(Perceptual Hypothesis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CHABA 1988; Schneider and Pichora-Fuller, 2000; Pichora-Fuller 2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Cause Hypothesis</td>
<td>A third variable causes both hearing loss and cognitive decline</td>
<td>Consistent with cognitive aging literature and declines in multiple sensory modalities; supported by structural equation modeling</td>
<td>Cerebrovascular disease; social relations; genetics</td>
<td>Modeling data suggests that both sensory and common factors contribute to cognitive decline</td>
</tr>
<tr>
<td>(CHABA,1988; Baltes &amp; Lindenberger 1997; Lindenberger &amp; Baltes 1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-2: Summary of proposed four directional hypotheses for the relationship between hearing loss and cognitive decline.
2.4.1 Cognitive Load on Perception Hypothesis

According to this view, declining cognitive capacity places a cognitive load on perception, which is then poorer (CHABA 1988; Lindenberger & Baltes 1994; Figure 2-1-A). There is little evidence for the idea that cognitive decline precedes hearing loss, either from behavioral studies, or from structural-equation modeling work (e.g., Humes et al. 2013; Lindenberger & Baltes 1994). In a longitudinal study, Kiely et al. (2012) found that probable cognitive impairment predicted faster rates of increase in PTA-measured thresholds. However, this study used a cognitive screening tool (MMSE) that is sensitive to hearing status, which may have confounded the results. Ultimately, evidence for this hypothesis is limited.

2.4.2 Information-Degradation Hypothesis

According to the information-degradation hypothesis (CHABA 1988; Lindenberger & Baltes 1994; Pichora-Fuller 2003; Schneider & Pichora-Fuller 2000; Figure 2-1-B), declines in older adults’ cognitive performance manifest as a consequence of compensating for impaired auditory input (owing either to normal aging or to pathology; Pichora-Fuller 2003; Schneider & Pichora-Fuller 2000, see also Schneider 2005). Evidence for the information-degradation hypothesis has been extensively reviewed by Pichora-Fuller (2003). According to this view, perceptual difficulties cascade “upwards” such that higher-level cognitive processing is compromised because mental resources have been diverted to perception. Older adults may compensate for sensory deficits via increased reliance on cognitive resources (specifically, working memory), thereby reducing the cognitive resources available for other cognitive tasks, resulting in an impairment in cognitive performance relative to younger adults (e.g., Dubno, Dirks & Morgan 1984; Kalikow, Stevens & Elliott 1977; Wingfield 1996; Zekveld, Deijen, Goverts & Kramer 2007; Akeroyd 2008).
Substantial evidence supports the idea that compensating for degraded speech depletes resources available for other downstream cognitive processes critical for ongoing speech communication. Long-term memory for words and sentences heard in degraded form tends to be poorer than memory for clear ones (Burkholder, Pisoni & Svirsky 2005; Pichora-Fuller, Schneider & Daneman 1995; Pichora-Fuller & Singh 2006; Piquado, Cousins, Wingfield & Miller 2010; Rabbitt 1968 1991; Wingfield, Tun & McCoy 2005; McCoy et al. 2005; Murphy, Craik, Li & Schneider 2000). It is thought that high processing load during perception interferes with encoding, which later manifests as poorer memory (e.g., Wingfield et al. 2005). Consistent with this idea, more effortful listening, which is required when the fidelity of the auditory input is poorer due to aging or pathology, appears to place greater demands on executive function and working-memory resources (Wingfield & Tun 2007; see also Amichetti et al. 2013).

One prediction of the information-degradation hypothesis is that an increase in perceptual load in younger adults (by degrading input) should lead to (apparent) cognitive deficits. In fact, studies by Murphy and colleagues (2000) and Schneider and colleagues (2002) observed that older adults demonstrate reduced overall memory for words even when equated with younger adults for initial perceptual performance. However, it is difficult to perfectly match perceptual conditions for young and older listeners, and the listening conditions for the older listeners in this study may have been more challenging, even though performance did not differ (moreover, in my experience, it is not unusual for older adults to be more motivated in experiments, and so try harder, than younger people). The study by Murphy and colleagues (2000) cited both impoverished sensory representations and decreased processing resources as explanations for the impaired memory in older people.
Overall, the information-degradation hypothesis explains at least some of the cognitive deficits experienced by older adults, when the perceptual input is less than perfect. In particular, it draws attention to the fact that hearing can often be effortful, such that high intelligibility can be achieved in two different ways – with effort, via increased reliance on attention, executive function, and working memory, if the speech signal is noisy, or with considerably less effort if speech is perceptually clear. Indeed, research confirms that hearing loss is commonly accompanied by increased listening effort (Feuerstein 1992; Hick & Tharpe 2002; McCoy et al. 2005; Pichora-Fuller et al. 1995; Rabbitt 1991) and is related to greater fatigue (Edwards 2007; Hick & Tharpe 2002), even when amplification is provided (Rakerd, Seitz & Whearty 1996). Accordingly, cognitive decline could be the manifestation of fatigue from everyday effortful listening; such apparent decline would be temporary and reversible.

2.4.3 Sensory-Deprivation Hypothesis

The sensory-deprivation hypothesis posits that perceptual declines cause more permanent cognitive declines (CHABA 1988; Lindenberger & Baltes 1994; Humes et al. 2013; F. R. Lin et al. 2013; Lindenberger & Baltes 1994; Pichora-Fuller 2003; Schneider & Pichora-Fuller 2000; Sekuler & Blake 1987; Uhlmann, Larson, Rees, Koepsell & Duckert 1989), possibly through neuroplastic changes that disadvantage general cognition in favor of processes supporting speech perception (F. R. Lin et al. 2013; Figure 2-1-C). Whereas the information-degradation hypothesis considers changes in cognition to be a potentially reversible byproduct of devoting cognitive resources to perception, the sensory-deprivation hypothesis emphasizes that such chronic re-allocation of cognitive resources may produce permanent changes in cognitive performance over time. Direct evidence for the sensory deprivation hypothesis is limited,
although it is consistent with some structural equation models (Humes et al. 2013; Lindenberger & Baltes 1994).

A potential mechanism of the sensory-deprivation hypothesis is thought to be deafferentation and atrophy in the auditory system as well as subsequent reorganization, owing to long-term sensory deprivation. Auditory deprivation occurring early in the lifespan appears to result in a change in physiology and functional reorganization of auditory areas (see Butler & Lomber 2013; Kral 2013; Kral & Sharma 2012 for reviews). Research suggests that age-related or noise-induced hearing loss may also result in cortical changes (Gold & Bajo 2014; Allman, Keniston & Meredith 2009; Meredith, Keniston & Allman 2012; Pienkowski & Eggermont 2011; Caspary, Schatteman & Hughes 2005; Dietrich, Nieschalk, Stoll, Rajan & Pantev 2001; Syka 2002; Wang, Brozoski, Ling, Hughes & Caspary 2011; Rajan 1998), which could also affect cognitive functions that rely on the affected brain regions.

Compensation for sensory loss may also trigger neurovascular and neurophysiological changes similar to those associated with dementia and that lead to declines in cognition. Wong and colleagues (2009) found decreased activation in auditory regions in older adults relative to younger adults for speech presented in noise. The authors also observed increased activation in frontal and posterior parietal working memory and attention networks, which were positively correlated with behavioral performance in older adults. Moreover, recent evidence indicates a compelling negative relationship between gray matter density in primary auditory areas and peripheral hearing ability, possibly reflecting cortical reorganization as a consequence of reduced sensory input (Peelle, Troiani, Grossman & Wingfield 2011; Lin et al. 2014; see also, Eckert et al. 2012). However, such a causal link requires that hearing loss precede the onset of clinically significant cognitive decline, which is very difficult to demonstrate.
Alternatively, depression and social isolation may mediate the relationship between hearing loss and cognitive decline. Difficulty with communication can severely limit the social integration of those with hearing loss or cognitive impairment. Decreased social interaction is associated with depression and withdrawal (Cacciatore et al. 1999; Mener, Betz, Genther, Chen & Lin 2013), and indeed hearing loss and cognitive impairment are linked to depression (Boi et al. 2012; Cacciatore et al. 1999; Herbst & Humphrey 1980; Jones, Victor & Vetter 1984; Kalayam et al. 1995; Kiely, Anstey & Luszcz 2013; Naramura et al. 1999, but see Mener, Betz, Genther, Chen & Lin 2013; Burt, Zembar & Niederehe 1995; Herbst & Humphrey 1980; Lopez et al. 2003). However, it has yet to be directly demonstrated that depression mediates or moderates the relationship between hearing loss and cognition (although see Jones, Vetter & Victor 1984, see also Suprenant & DiDonato 2014).

However, the sensory-deprivation hypothesis is contradicted by the fact that long-term sensory impairments in younger adults appear to have negligible effects on overall level of (non-verbal) cognitive function (see Vernon 2005 for a review), indicating that impoverished sensory input cannot be the sole trigger of cognitive impairment in affected older adults. Although cognitive and sensory function are moderately to strongly correlated across the lifespan (Humes et al. 2013), reduced processing capacity (such as a general decline in processing speed or working memory) or changes in attention may also contribute to cognitive impairment.

2.4.4 Common-Cause Hypothesis

The common-cause hypothesis (Baltes & Lindenberger 1997; CHABA 1988; Lindenberger & Baltes 1994; Figure 2-1-D), proposes that a common mechanism underlies both age-related changes in cognition and hearing (and other sense modalities) through widespread
neural degeneration. In evaluating the common-cause hypothesis, I refer both to the cognitive aging literature and to literature evaluating age-related declines in other sensory modalities.

Not only do older adults experience greater perceptual challenges than their younger counterparts, they may also have cognitive deficits. As reviewed elsewhere, cognitive aging is characterized by widespread age-related neural degeneration, with reductions in volume and length of dendritic spine densities in multiple cortical regions, particularly in pre-frontal regions (see Park & McDonough 2013). Potential mechanisms of generalized decline in cognitive function are comprehensively reviewed by Surprenant and Neath (2006).

Aging is also accompanied by declines on tests sensitive to a number of executive functions, including working memory, long-term memory, reasoning, and inhibitory control (the ability to ignore irrelevant information during cognitive processing; for reviews, see Hedden & Gabrieli 2004; Jagust 2013; Nyberg, Lovden, Riklund, Lindenberger & Backman 2012). The idea that older adults have a smaller pool of cognitive resources on which to rely for compensation is supported by research demonstrating greater age differences in performance as tasks become more demanding (Myerson, Hale, Wagstaff, Poon & Smith 1990), and that older adults reach a maximum level of performance and activation at lower objective levels of difficulty than their younger peers (Cappell, Gmeindl & Reuter-Lorenz 2010; Nagel et al. 2009).

The fact that age-related cognitive decline tends to affect multiple domains of cognitive function is consistent with the common-cause hypothesis. General slowing of processing and performance on motor tasks is evident, and some have argued that slowed processing is the central deficit associated with aging (CHABA 1988; Salthouse 1996), affecting performance across multiple domains of cognitive and sensory function. In fact, in one study, information processing speed was most strongly linked to time to death, relative to other cognitive measures.
(Smits, Deeg, Kriegsman & Schmand 1999). However, the extent to which processing speed explains the link between hearing loss and cognitive decline is unclear. In one study, the link between timed tests of cognition and poorer audiometric thresholds was no longer significant after controlling for measures of processing speed (van Boxtel et al. 2000), but in a large-scale study, Anstey, Luszcz, and Sanchez (2001b) found that unique effects of age and sensory function on cognition were not fully explained by processing speed.

Many studies support a link between cognitive decline and vision loss, similar to that proposed for hearing loss (e.g., Anstey et al. 2001a; M. Y. Lin et al. 2004; Lindenberger & Baltes 1994; Uhlmann et al. 1991). For example, Lin et al. (2004) sampled 6112 women ages 69 and above, evaluating cognition using the 3MS cognitive screening test. The authors found a two-fold increase in the odds of cognitive and functional decline with visual impairment, defined as corrected binocular vision worse than 20/40\textsuperscript{1}. In a similar study, after controlling for family history of dementia, depression, number of medications and hearing loss, those with dementia were significantly more likely to have visual impairment, defined here as uncorrected visual acuity less than the median control group value (Uhlmann et al. 1991). In those with Alzheimer’s, the degree of visual impairment was significantly correlated with the severity of cognitive impairment.

Several groups have also shown links between hearing loss, and gait and vestibular function. Li and colleagues (2012) found that hearing loss was independently associated with slower gait speed, even after adjusting for demographic and cardiovascular factors (see also Viljanen et al. 2009). The reduction in gait speed associated with a 25-dB HL hearing loss was equivalent to that associated with 12 additional years of age. Lindenberger and Baltes (1994)

\textsuperscript{1} An odds ratio represents the odds that an outcome will occur (i.e., cognitive decline), given a particular exposure (i.e., vision loss), compared to the odds of the outcome occurring in the absence of that exposure.
also found that balance and gait correlated with intellectual functioning in old age to a similar extent as auditory sensitivity and visual acuity (see also, Gerson, Jarjoura & McCord 1989). Moreover, Lin and Ferrucci (2012) observed that hearing loss was independently associated with self-reported falls over the preceding 12 months in a sample of adults aged 40-69 years, even after adjusting for balance function (using standardized assessment). Degeneration in both hearing and balance could be linked to common pathological processes in the inner ear (Li, Simonsick, Ferrucci & F. R. Lin 2012).

Olfaction is an interesting sensory modality, since it involves the entorhinal cortex of the anterior temporal lobe; this region is important for memory function and is one of the sites of pathology related to dementia. Olfactory deficits appear to be an ‘early warning sign’ of incipient dementia or Alzheimer’s, preceding the appearance of behavioral symptoms (e.g., Devanand et al. 2010; Duff, McCaffrey & Solomon 2002; Graves et al. 1999; Gray, Staples, Murren, Dhariwal & Bentham 2001; Karpa et al. 2010; McShane et al. 2001; C. Murphy 1999; Pardini, Huey, Cavanagh & Grafman 2009; Serby et al. 1996; Serby, Larson & Kalkstein 1991; Thompson, Knee & Golden 1998; Wang et al. 2010).

Evidence for concurrent changes in multiple perceptual/cognitive domains is suggestive of systemic central nervous system pathology, consistent with a common neurodegenerative etiology. A comprehensive theory linking hearing loss to cognitive decline must account for concurrent declines in other domains of sensory function; a common-cause is the most parsimonious explanation. Potential mechanisms for common declines include cerebrovascular disease and general physical health (e.g., Kuo et al. 2005; Laughlin, McEvoy, Barrett-Connor, Daniels & Ix 2013; Dubno et al. 2008; Eckert et al. 2013; Kiely et al. 2012; Mills, Schmiedt,
Schulte & Dubno 2006), and genetics (e.g., Viljanen et al. 2007; Wisdom, Callahan, Hawkins 2011).

Structural equation modeling data generally support the existence of a common factor contributing to age-related sensory and cognitive declines (Anstey, Hofer & Luszcz 2003; Humes et al. 2013; Lovden & Wahlin 2005; Tay et al. 2006; van Rooij & Plomp 1990; van Rooij & Plomp 1991; Whalley, Deary, Appleton & Starr 2004, Christensen, Mackinnon, Korten & Form 2001; Lindenberger & Ghisletta 2009, Li & Lindenberger 2002). However, according to a large-scale study by Anstey and colleagues (2001b), unique (non-common) effects of age and sensory function on cognition cannot be fully explained by another variable, such as processing speed. The idea that sensory function contributes independently to cognition is further supported by a recent study by Humes and colleagues (2013), demonstrating that decline in age-related sensory function is largely independent of age-related (or general) cognitive function, using a variety of measurements of auditory acuity, in addition to measuring auditory sensitivity. Such independent contributions would not be observed if a single common-cause was underlying all decline. Ultimately, evidence suggests that a mixture of common-cause and either sensory and or perceptual factors (in addition to age and processing speed) underlie cognitive decline (Anstey, Hofer & Luszcz 2003; Humes et al. 2013; Lovden & Wahlin 2005; Tay et al. 2006; van Rooij & Plomp 1990; van Rooij & Plomp 1991; Whalley, Deary, Appleton & Starr 2004, Christensen, Mackinnon, Korten & Form 2001; Lindenberger & Ghisletta 2009).

2.4.5 Summary of Hypotheses

In reviewing the four hypothesized relationships between hearing loss and cognitive decline, the information-degradation and common-cause hypotheses emerge as strong contenders. The information-degradation hypothesis is well-supported by the cognitive
literature, and accounts for documented independent contributions of sensory or perceptual deficits to cognitive decline (e.g., F. R. Lin et al. 2013; F. R. Lin et al. 2011b). The common-cause hypothesis accounts for the fact that multiple sensory modalities and cognition appear to decline concurrently. A single account cannot explain all the data, suggesting that multiple mechanisms are probably at play.

Although the sensory-deprivation hypothesis is a weak causal explanation on its own, I will consider the information-degradation and sensory-deprivation hypotheses together (the ‘sensory hypothesis’), since they both suggest that perceptual demands consume cognitive resources, which has a downstream impact on cognition. The difference is that this impact is implied to be short-term and potentially remediable by the information-degradation hypothesis, and more chronic and permanent by the sensory-deprivation hypothesis.

In the next section, I describe a framework that may help to conceptualize the links between declines in hearing and cognition in older people, capitalizing on the explanatory power of the information-degradation and common-cause hypotheses.

2.5 A Working Framework

The framework presented in Figure 2-1 attempts to capture the explanatory power of multiple empirically supported directional relationships between hearing loss and cognitive decline. This framework highlights the multiple, cumulative, deleterious processes to which the auditory/cognitive system is vulnerable (Jagust 2013). Sensation/perception and cognitive function are interdependent, and deficits at one point in the system will have multiple, cascading, effects at other points.
Neurodegenerative processes will cause declines in hearing, producing sensory and perceptual difficulty (Figure 2-2-1) and increased challenge to cognitive resources (2), such that auditory perceptual processes must compete for a reduced pool of shared cognitive resources (5). Sensory deficits occurring as a result of both central nervous system neurodegeneration and age-associated pathology in the inner ear (3) will result in greater perceptual challenges (4), which recruit more cognitive resources (5). Since the cognitive pool is already reduced or strained (2), this may result in functional cognitive impairments (6).

Diminished cognitive resources may also contribute to perceptual deficits (5), since deficits in attention or processing speed could compound or interact with deficits resulting from sensory declines (Harris, Eckert, Ahlstrom & Dubno 2010). The notion of “cognitive reserve” developed in the cognitive aging literature may be helpful here. This assumes a common pool of
undifferentiated cognitive resources that can be recruited to support speech comprehension (or other processes). Such a cognitive reserve is potentially different from dynamic compensatory processes, which are thought to rely on functional reorganization to maintain task performance (for a review, see Whalley et al. 2004). Because different individuals have different cognitive-reserve capacities (see Tucker & Stern 2011 for a review), they will vary in their ability to compensate for sensory or cognitive impairment. Whenever demands exceed the reserve, behavioral impairments will be noted (consistent with the information-degradation hypothesis), although these may be reversible if perceptual difficulty can be decreased (i.e., by improving speech clarity through amplification or reduction of noise).

Perceptual difficulty may affect communication to the degree that it leads to social withdrawal, and negative psychosocial outcomes such as depression (7). Depression could additionally precipitate or exacerbate cognitive decline (8).

This framework is also compatible with age-related declines in other sensory domains, although vision and balance, together with hearing, are the major sensory domains in which decline has been linked to negative psychosocial outcomes.

2.6 Directions for Future Research

Research linking age-related hearing loss to cognitive decline has largely relied on PTA, a behavioral index of hearing sensitivity, which may underestimate the link (Humes et al. 2013). As deficits in hearing acuity may precede elevations in PTAs (e.g., Frisina & Frisina 1997; He, Dubno & Mills 1998; Kujawa & Lieberman 2009), it would be useful to explore the degree to which the relationship between hearing loss and cognitive decline, as observed through audiometry, extends to aspects of hearing acuity by incorporating measures of frequency selectivity and temporal processing.
The framework shown in Figure 2-1 can serve as a useful starting point for testing directional hypotheses related to the relationship between cognition and sensory function - in hearing as well as in other sensory modalities. This framework offers multiple entry points for targeted intervention; the results of such manipulations would allow for more causal inference about the nature of the relationship. For example, one could examine whether hearing aids or other assistive listening devices retard or even ameliorate cognitive decline (reducing the effect of Figure 2-2-4; F. R. Lin et al. 2011b; F. R. Lin et al. 2011a; Pichora-Fuller 2003). If this were the case, it would support the sensory hypothesis. Currently, the evidence is mixed; some studies support the effectiveness of amplification in improving cognitive function (Mulrow et al 1990; Weinstein and Amsel 1986; Mulrow, Tuley & Aguilar 1992), whereas others do not (Hooren et al. 2005; Wong et al. 2014; Allen et al. 2003; F. R. Lin et al. 2011; but see also F. R. Lin et al. 2013). The effectiveness of hearing aids may be limited in part by poor compliance due to poor perceived value, poor fit, lack of comfort, and psychosocial factors, which generally have not been taken into account in previous studies (see McCormack & Fortnum 2013 for a review). Additionally, amplification does not resolve communication difficulties attributable to other factors such as high processing load as a result of poor hearing acuity, which remains with amplification (Hornsby 2013). Instructing individuals about how best to use their assistive listening devices may enhance the sensitivity of such studies (Boothroyd 2007).

Several studies have reported that repeated use of executive functions (e.g., working memory, processing speech, and attention) over the weeks of a training regime improves performance on functions and tasks that rely on these functions (e.g., Buschkuehl et al. 2008; Klingberg et al. 2005; Morrison & Chein 2011; Richmond, Morrison, Chein & Olson 2011; Reijnders, van Heugten & van Boxtel 2013; Kelly et al. 2014). Such cognitive training is another
intervention that can also be used to test directional hypotheses. The apparent benefits of cognitive training are thought to arise either from increasing cognitive resource capacity (cognitive reserve) or from more efficient use of existing cognitive resources.

A premise of the sensory-deprivation hypothesis is that chronic reallocation of cognitive resources to speech perception over time, as a consequence of hearing loss, may permanently deplete the pool of resources available to for other cognitive functions, manifesting as cognitive decline. This idea is at apparent odds with the premise of the cognitive training literature, which is that increased cognitive demand will lead to an increase in cognitive resources or in their efficient use, with generalized consequences for cognitive function. The cognitive challenges imposed by speech in difficult listening environments look a lot like training: in fact, how they would differ from challenges deliberately introduced in the context of cognitive training is not entirely clear. The cognitive training literature would predict that, if degraded sensory input does result in higher-level cognitive processes “working harder” to support perception, then we should see improved cognition in individuals with hearing loss, not cognitive decline. Although not all studies demonstrate a benefit of cognitive training, no study, to my knowledge, has observed such training to be harmful. Examination of the effect of cognitive training on the relationships among the variables (boxes) in Figure 2-1 would help to resolve this paradox.

Moreover, training speech perception abilities would enable testing of the information-degradation hypothesis. Training strategies targeting better utilization of sentence context to aid word-level recognition, as well as practice with multi-talker babble, speech presented in background noise, time-compressed speech (simulating rapid speech), or other forms of degraded speech may aid older adults in making the most of their remaining hearing (e.g.,
Sweetow & Sables 2006; Rubinstein and Boothroyd 2007; Dubno 2013), perhaps reducing the load on compensatory cognitive resources.

The association among depression, hearing loss, and cognitive decline might be more directly evaluated through the use of psychological therapy (such as cognitive behavioral therapy; CBT) in order to address underlying cognitions that may maintain withdrawal and isolation due to hearing loss. Psychosocial or psycho-educational interventions that teach skills related to hearing-loss advocacy for both individuals and their families would provide a powerful tool to boost participation in social, vocational, and leisure activities. If depression and isolation are indeed mediating factors in the relationship between hearing loss and cognitive decline, then a reduction in cognitive impairment should be observed following psychological or psychosocial intervention. If functional disability mediates the relationship between hearing loss and cognition (Jones, Victor & Vetter 1984), then interventions aimed at increasing everyday independence after hearing loss may also ameliorate age-related cognitive declines.

One corollary of the common-cause hypothesis is that cognitive and sensory function may be linked through general health. Associations between cardiovascular disease and sensory functioning have been noted (see, for instance, Gates et al. 2003; Kiely et al. 2012; Hofer et al. 2003) and could be investigated by examining the extent to which health-related physical and metabolic interventions help to offset both cognitive and sensory decline (see, for example, Witte, Kerti, Marguilies & Floel 2014).

2.7 Conclusion

Causal inferences about the nature of the relationship between hearing loss and cognitive decline are difficult to make, due to the correlational nature of much of the research. The nature of the link is also obscured by variability in the way that cognition is assessed, and by the limited
way in which hearing loss is typically assessed. Changes in frequency selectivity have arguably the greatest impact on everyday communication, and yet have seldom been addressed, although the recent interest in ‘hidden hearing loss’ (e.g., Plack, Barker & Prendergast 2014) may catalyze a change in the way hearing loss is measured. Increased use of non-auditory measures of cognition, as well as of a broader array of measures of peripheral hearing may better substantiate what currently appears to be a weak link between hearing loss and cognitive decline. Further, emerging evidence suggests that larger effect sizes may be seen for the relationship when using measures of temporal processing and frequency selectivity, compared to pure-tone audiometry.

The framework I have suggested recognizes that hearing and cognition are complementary and interdependent processes that rely on shared resources, and thus interrelate in complex ways. This framework capitalizes on the strengths of the various hypotheses and enables assessment of directional hypotheses by highlighting potential sites of cognitive, sensory, and metabolic intervention. I hope this framework can serve as a foundation for future rehabilitative efforts in ameliorating the effects of age-related hearing loss, as well as age-related declines in other sensory modalities.
Chapter 3- Working Memory Training and Speech in Noise Comprehension in Older Adults
3.1 Introduction

Perception and comprehension of speech heard in background noise becomes more difficult with age (Plomp & Mimpen, 1979; Pichora-Fuller & Souza, 2003; Sommers, 1997; Schneider et al 2002; van Rooij & Plomp, 1989). These difficulties are not fully explained by pure-tone audiometric thresholds (hearing sensitivity), and may be due in part to declining frequency selectivity and temporal coding with age (Humes & Dubno, 2010; Plack, Barker & Prendergast, 2014; Kujawa & Lieberman, 2009, Gordon-Salant, Frisina, Popper & Fay, 2010). Moreover, amplification devices (i.e., hearing aids), the most widely prescribed treatment for hearing difficulties, improve hearing sensitivity but not frequency selectivity or temporal coding, which are important for segregating speech from background sound (e.g., Perez, McCormack & Edmonds, 2014), and many individuals who have been prescribed hearing aids do not wear them (see McCormack & Fortnum, 2013 for a review). As communication difficulties are linked to depression, isolation and decreased quality of life (Mulrow et al., 1990; Carabellese et al., 1993; Cacciatore et al., 1999), rehabilitative strategies, used either in isolation or in combination with amplification, are urgently needed. Here, I focus on the utility of cognitive strategies for hearing loss rehabilitation.

Hearing loss leads to degradation of the incoming acoustic signal, and the resulting perceptual ambiguity places increased demands on executive processes that mediate knowledge-guided perceptual processes, such as using context in order to select the contextually appropriate meaning from among competing alternatives (Rodd, Davis, & Johnsrude, 2005; Rodd, Johnsrude, & Davis, 2012; Zekveld et al., 2012). This requires listeners to rely more heavily on top-down information, recruiting previous experience and linguistic knowledge to help evaluate perceptual hypotheses about the incoming signal (Kane & Engle, 2000).
The ability to use contextual information effectively to enhance intelligibility varies widely among individuals (Davis, Ford, Kherif and Johnsrude, 2011; Janse & Jesse, 2014). This variability may be attributable, in part, to individual differences in more domain-general cognitive abilities such as processing speed, and executive functions such as working memory and inhibition (Heald & Nusbaum, 2014; Wingfield & Tun, 2007; Arlinger et al., 2009; Schneider, Pichora-Fuller, & Daneman, 2010 for reviews). Executive functions allow listeners to direct attention to a particular speaker, integrate the acoustic signal with previous knowledge, and inhibit irrelevant information (e.g., Tamati, Gilbert, & Pisoni, 2013; Woods et al., 2013; Tun, O’Kane & Wingfield, 2003). In addition, knowledge about linguistic structure and sentence parsing may facilitate the use of contextual information to support speech understanding (Aydelott, Leech, & Crinion, 2011, Billig, Davis, Deeks, Monstrey & Carlyon, 2013; Rodd et al., 2005; Rodd et al., 2010).

Slower processing speed is linked to difficulty understanding speech in noise (e.g., Pronk et al., 2013; Tun & Wingfield, 1999) and may be particularly important for understanding speech spoken at fast rates (e.g., Wingfield, Tun, Koh, & Rosen, 1999; Gordon-Salant and Fitzgibbons, 2001). Both processing speed (Salthouse, 1996) and executive functions (Craik & Salthouse, 2007) decline with age and such declines appear to contribute to listening difficulties in older adults (Humes & Dubno, 2010).

A large body of research has linked working memory measures to speech comprehension in poor listening conditions (due to noise or pathology; e.g., Szenkovitz, Peelle, Norris, & Davis, 2012; George et al., 2007; Sorqvist & Rönnberg, 2012; Rudner, Rönnberg & Lunner, 2011; Wingfield & Tun, 2007; Akeroyd, 2008; Rudner & Lunner, 2014, Heald & Nusbaum, 2014, Talmati et al., 2013; Zekveld, Rudner, Johnsrude, Heslenfeld & Rönnberg, 2012; Humes, Lee, &
The evidence to date supporting the role of working memory in speech perception has been largely correlational. However, working memory may contribute to the ability to use sentence context to guide and constrain interpretation to compensate for increased processing demands when the signal is degraded and interpretation therefore ambiguous (Zekveld et al., 2012; Rodd, Davis, & Johnsrude, 2005; Rodd, Johnsrude, & Davis, 2012). Thus, individuals with greater working memory capacity may be better able to compensate for degraded listening conditions.

Research linking working memory capacity to speech comprehension has largely relied on the Reading Span test (Daneman & Carpenter, 1980) as a measure of working memory (Davies-Venn & Souza, 2014; Tun, Wingfield, & Stine, 1991; Akeroyd, 2008; Zekveld et al., 2012, Zekveld, Rudner, Johnsrude & Rönnberg, 2013, Besser, Koelewijn, Zekveld, Kramer & Festen, 2013, but see Schoof & Rosen, 2014; Humes & Coughlin, 2009). The Rönnberg and colleagues (1989) version (adapted from Daneman & Carpenter, 1980 and Baddeley et al., 1985, but see also Towse et al., 2008 and Conway et al., 2002 for alternate versions) is most widely used. These five versions differ in a number of ways (see Table 3-1). In the canonical Daneman and Carpenter (1980) version (and similarly in the Rönnberg et al., 1989 version), participants read or hear a set of sentences, and are required to make a yes/no semantic judgment after each sentence to prevent rehearsal of items. After each set participants are asked to recall the last word from each sentence in the set in serial order. (See also Lyxell and Rönnberg, 1993, in which subjects are cued whether to recall the first or last word in a set). As the test progresses, the set size increases. Although the Reading Span test is commonly referred to as a test of working memory, it also draws on other cognitive abilities, including processing speed, executive functioning (e.g., selective attention, inhibition, task switching), and reading skill: it is
not a “pure test” of working memory. Thus, it is possible that a correlation between working memory and speech in noise (Szenkovitz, Peelle, Norris, & Davis, 2012; George et al., 2007; Sorqvist & Rönnberg, 2012; Rudner, Rönnberg & Lunner, 2011; Wingfield & Tun, 2007; Akeroyd, 2008; Rudner & Lunner, 2014, Heald & Nusbaum, 2014, Talmati et al., 2013; Zekveld, Rudner, Johnsrude, Heslenfeld & Rönnberg, 2012; Humes, Lee, & Coughlin, 2006) may, in fact, be due to other cognitive processes.
<table>
<thead>
<tr>
<th></th>
<th>Mode of Delivery</th>
<th>Recall Item</th>
<th>Judgment Task</th>
<th>Discontinuation Criteria</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daneman &amp; Carpenter, 1980</td>
<td>oral or read from card</td>
<td>final word in series</td>
<td>semantic (yes/no)</td>
<td>Fail all 3 items in set</td>
<td>span where items correctly reported for 2/3 sentences in a set</td>
</tr>
<tr>
<td>Baddeley et al., 1985</td>
<td>oral</td>
<td>subject or object of sentence (as cued)</td>
<td>factual (true/false)</td>
<td>N/A</td>
<td># of items recalled in correct serial order</td>
</tr>
<tr>
<td>Ronnberg et al., 1989</td>
<td>read from computer</td>
<td>final word in series</td>
<td>semantic (yes/no)</td>
<td>N/A</td>
<td># correct items recalled / maximum score</td>
</tr>
<tr>
<td>Conway et al., 2002</td>
<td>read from computer</td>
<td>unrelated word presented at end of each sentence</td>
<td>none; must score better than 50% on comprehension post-test</td>
<td>N/A</td>
<td># of items recalled in correct serial order, weighted by the number of items within a series (e.g., 2 pts per correct 2-item series)</td>
</tr>
<tr>
<td>Towse et al., 2008</td>
<td>read from computer; subjects provide a word to complete sentence</td>
<td>Completion word provided by subjects (integrated word condition), or unrelated target word provided (independent word condition)</td>
<td>none</td>
<td>Fail all 3 items in set</td>
<td># of words correctly recalled</td>
</tr>
</tbody>
</table>

Table 3-1: Description of different versions of the Reading Span test.
One way to directly confirm the involvement of working memory processes in speech-in-noise performance is to demonstrate training-related improvement in performance on speech-in-noise tests after increasing working memory capacity through training. A growing body of research has examined the efficacy of working-memory training to improve working-memory capacity (for reviews, see Hindin & Zelinski, 2012; Karr, Areshenkoff & Garcia-Barrera, 2014; Melby-Lervag & Hulme, 2014). The efficacy of working memory training, and cognitive training generally, has been the subject of debate (e.g., Shipstead, Redick, & Engle, 2012; Owen et al., 2010; Lampit, Hallock & Valenzuela, 2014; Jacoby & Ahissar, 2013). However, if effective, working memory training (and cognitive training, in general), holds great promise for mitigating documented age-related declines in processing speed, episodic memory, working memory, and other domains of executive function (e.g., see Craik & Salthouse, 2007).

I examine whether working-memory training transfers to speech perception in noise in older adults, as well as to other tests of cognitive functioning. I used Cogmed Working Memory Training (Version QM: Pearson; Klingberg, Forssberg & Westerberg, 2002), an adaptive, computerized, commercial working memory-training program. I selected Cogmed since several publications have demonstrated its efficacy (Holmes, Gathercole, & Dunning, 2009; Klingberg, 2010; Klingberg et al., 2005; Olesen, Westerberg, & Klingberg, 2003; Brehmer et al., 2011; Brehmer, Westerberg & Backman, 2012), and it includes a placebo training condition (an active control group). Improvements due to Cogmed training have been shown to generalize to related working-memory tasks (near transfer), including verbal working memory and visuo-spatial working memory, in both younger (Holmes et al., 2009; Klingberg, 2010; Klingberg et al., 2005; Olesen et al., 2003) and older (Brehmer et al., 2011; Brehmer, Westerberg et al., 2012) adults. Far transfer to other domains of intellectual functioning (verbal and non-verbal reasoning) has
not been shown (Shipstead, Hicks & Engle, 2012). However, in a study of deaf children with cochlear implants, Kronenberger and colleagues (2011) observed improved verbal and non-verbal working memory capacity, as well as improved sentence-repetition ability after Cogmed training, providing some evidence of transfer to real-world speech comprehension ability, indicating suitability of the use of Cogmed for my study.

I tested for transfer to perception of speech in noise with two speech tasks. The first task was a closed-set, 5-word sentence matrix test (BUG; Kidd, Best & Mason, 2008), with two competing talkers (and one target speaker). The second task assessed perception in noise for sentences with and without supporting contextual information. The ability to use supporting contextual information to facilitate comprehension of degraded speech appears to depend on working memory (Janse & Jesse, 2014) and varies markedly among individuals (Zekveld et al., 2012; Rodd, Davis, & Johnsrude, 2005; Rodd, Johnsrude, & Davis, 2012). It is not known, however, whether this benefit from context can be improved through working memory training.

I provided older adults with five weeks (25 sessions) of both adaptive and placebo Cogmed (Klingberg, Forssberg & Westerberg, 2002) training in a cross-over design, and evaluated cognitive functions at three time points: prior to the start of training (T0), following the first five weeks of either adaptive or placebo training (T1), and following the second five weeks of training (T2; participants who received adaptive training in T1 received placebo training in T2 and vice versa; see Figure 3-1). Cognitive evaluation included tests of working memory, processing speed, fluency, and short-term memory. I also assessed speech comprehension in noise using the BUG and high-/low-context sentence tasks. Near-transfer was assessed through improvements on other working memory tasks as a function of adaptive training, whereas far-transfer was operationalized as improvements on non-trained cognitive tasks (i.e., those other
than tests of working memory, including speech tasks). Note that this design is among the first randomized active control trial for working memory training in older adults that specifically assesses for transfer to speech-in-noise function (see also, Henshaw & Ferguson, 2013 for the protocol for a forthcoming trial with hearing-aid users). Importantly, this design allows us to establish a causal link between working memory training and speech in noise function, rather than relying on correlational designs, a limitation of previous research. If working memory contributes to speech in noise perception, then improvements as a result of adaptive working memory training should lead to improved performance on tests of working memory (near-transfer), as well as on tests of speech perception, above and beyond practice effects (far-transfer; see Figure 3-2). In order to investigate the relationship between measures of working memory and other cognitive functions, and the ability to comprehend speech in noise, I examined correlations between cognitive abilities, particularly working memory ability, and speech-in-noise performance.

Figure 3-1: Study design. Participants were split into two groups and tested at three time points: at baseline, and following two 25-session training blocks. The A-P group received adaptive training followed by placebo training, whereas the P-A group received the placebo training followed by adaptive training.
Figure 3-2: Dissociating training from practice effects. My cross-over design allowed us to dissociate training effects from practice effects, within-subjects, by aggregating across the two 5-week blocks of training (T1-T0 and T2-T1). A-P refers to the group receiving adaptive training after baseline testing, followed by placebo training, whereas P-A refers to placebo training followed by adaptive training. Note that this is a hypothetical outcome.

3.2 Methods

3.2.1 Subjects

I recruited 26 subjects (13 male, 13 female) between 59-73 years of age (mean = 64.96 years, SD = 3.77 years) through local newspapers, flyers, and community groups. Subjects generally reported good health and they were screened for hearing loss and mild cognitive impairment (see below) before beginning the study. Informed consent was obtained from all subjects and they were compensated for their time in the laboratory at a rate of $10 per hour. Subjects completing the study also received a $50 gift card for their efforts spent training online.

3.2.2 Screening Procedure

Before commencing the study, all subjects received an audiogram (at 0.5, 1, 2, and 3 kHz), as well as the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). Subjects with scores of 23 or below on the MoCA were excluded from the study, as scores of 24 and
above are indicative of cognitive impairment with over 95% sensitivity and specificity (Luis, Keegan, & Mullan, 2009). I excluded two subjects (not reported here) on this basis. A total of 13 subjects were classified as having normal hearing, defined as thresholds at or below 25 dB HL at the tested frequencies. The 13 remaining subjects had some hearing loss: mild (a loss of < 40 dB in the better ear for at least one of the tested frequencies) in 6 participants; moderate (< 55 dB in the better ear) in four; moderate-severe (< 70 dB in the better ear) in two; and severe (>70 dB in the better ear) in one. These individuals were not excluded. The single participant with severe hearing loss wore hearing aids during testing, as did one participant with moderate hearing loss and one with moderate-severe hearing loss. I accounted for this heterogeneity in hearing levels by testing the effects of training in two ways: overall (collapsed across all subjects), as well as by conducting analyses for both groups separately (although this substantially reduces power).

During this initial screening session, subjects also completed portions of the Speech, Spatial and Qualities (SSQ) of Hearing scale (Gatehouse & Noble, 2004; the subset of questions related to spatial hearing were not administered), Raven’s Progressive Matrices (a measure of non-verbal intelligence), and a demographic questionnaire. Subjects also completed the Burns Anxiety Inventory (Burns, 1999); however, these scores did not correlate with other measures and were not used in subsequent analyses.

3.2.3 Cognitive Training Procedure

Subjects were instructed to train five days a week for ten weeks (i.e., 25 sessions in total for both active and placebo training) using the Cogmed working memory-training program (Klingberg, Forssberg & Westerberg, 2002). Twelve different training modules, involving remembering a sequence of numbers, letters, or objects for immediate recall, were used. Some exercises involved active manipulation of information, such as entering numbers in the reverse
order that they appeared. Subjects worked on 8 of a possible 12 modules on each day of training; the modules that each subject had to complete on a given day were pre-determined by the online training program and were consistent across subjects. Training sessions took approximately half an hour to an hour per day to complete (with shorter times for placebo training).

In adaptive training modules, the level of difficulty was adjusted according to subjects’ performance by increasing stimulus span length on the subsequent trial (conversely, span length was decreased following unsuccessful trials). In placebo training, only three items were ever presented for recall at a time. Since most individuals can easily recall three items, placebo training was not expected to improve working memory capacity (all subjects scored near 100% on placebo training). All subjects completed five blocked weeks of both adaptive and placebo training; however, the order in which the two different kinds of training (adaptive and placebo) was administered was pseudorandom and counterbalanced across subjects (see Figure 3-1). Two couples participated and in that case both partners were assigned to the same group at the same time, since I wanted to keep participants naive about the other condition they would perform. Subjects were told that they would be completing ‘brain training’, but no direct information about there being both adaptive and placebo conditions, and which condition they were completing at present, was provided.

The 25 sessions of each condition were completed on average in 33.88 days (SD = 2.86; including non-training days). The average time to completion did not differ significantly between the training groups or training type (placebo vs. adaptive) for subjects who successfully completed training. Progress was monitored remotely every week via the Cogmed Training Web to ensure training sessions were completed. Subjects received a weekly email from the
experimenter (RVW; a certified Cogmed coach), who addressed training-related concerns and questions, and offered encouragement to maintain motivation.

Dependent variables from adaptive Cogmed training included the Start Index (calculated by Cogmed based on span length from training days 2 and 3), the Maximum Index (calculated by Cogmed from the two best days during training), as well as the Index-Improvement score (calculated by Cogmed as the Subtraction of Start Index from the Maximum Index).

### 3.2.4 Testing Procedure

The cognitive and speech tests were administered before and after each block of testing in two sessions, separated by a short break. The order of the sessions was identical for each subject across all three time-points, but counterbalanced across subjects. Cognitive testing was completed in a quiet room free of distractions. All auditory tests were conducted in a sound-attenuating booth (Eckel Industries) with headphones (Grado Prestige SR225). Speech stimuli were adjusted to a comfortable listening level based on feedback from subjects (mean = 76.49 dB, SD = 6.76). Speech levels for each subject were kept constant across all three time-points. Testing was usually completed in 2.5-3.5 hours (across two sessions), and at the same time of day, where possible (i.e., due to scheduling logistics).

### 3.2.5 Cognitive Test Battery

I administered a broad cognitive test battery in order to assess the generalization of working memory training to multiple domains of cognitive function. Tests of near-transfer included measures of spatial working memory (Spatial Working Memory, CANTAB; Spatial Span Forward and Reverse, CANTAB), and verbal working memory (WAIS-IV Letter-Number}

---

2 For eight participants, the levels were changed due to experimental error at T1 and T2. However, at T2, the average change in levels was identical between the A-P and P-A groups for both speech tasks. At T1, the change in levels favored the adaptive group; this biased the results in favor of improved speech comprehension performance as a result of cognitive training, which we still did not observe.
Sequencing). Far-transfer tests included assessment of episodic memory (Paired Associate Learning Test, CANTAB), semantic fluency (Category Fluency; Strauss, Sherman & Spreen, 2009), response inhibition (Stop Signal Task, CANTAB), motor/processing speed (Reaction Time, CANTAB), and sustained visual attention (Rapid Visual Information Processing, CANTAB). I also included a computerized version of the Reading Span test (Rönnberg et al., 1989) and asked subjects to report the last word of every sentence. Tests were chosen on the basis of availability of published norms for older adults, although raw scores were used for the purposes of subsequent analyses. When multiple versions of a test existed (i.e., Paired Associate Learning, Stop Signal Task, Spatial Working Memory, fluency), all three versions were counterbalanced across subjects. Please see Table 2 for further description of these tests and the dependent variables derived from them.
<table>
<thead>
<tr>
<th>Cognitive Test</th>
<th>Domain Assessed</th>
<th>Description</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Working Memory (SWM)*</td>
<td>working memory</td>
<td>3, 4, 6, and 8 boxes are dispensed on the screen. Subjects search for blue tokens hidden inside one of the boxes. Only one blue token is hidden at a time, without replacement (subjects must remember which boxes have produced a token).</td>
<td>Between Errors** – The number of times a box in which a token has previously been found is revisited.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strategy - The number of times the subject begins a new search with a different box for 6- and 8-box trials.</td>
</tr>
<tr>
<td>Spatial Span (SSP)* (forward and reverse modes)</td>
<td>working memory</td>
<td>White squares are arranged in a variable sequence on screen. Subjects touch the boxes in the order in which they changed colour. The length of the sequence begins at 2 and increases adaptively up to 9 boxes. In reverse mode, subjects touch the boxes in the reverse order that they changed colour.</td>
<td>The longest sequence successfully recalled by the subject, calculated for both the forward and reverse mode.</td>
</tr>
<tr>
<td>WAIS-IV Letter-Number sequencing</td>
<td>working memory</td>
<td>Subjects repeat back a string of letters and numbers in numerical order, followed by alphabetical order. The number of items in a string increases from 2-8 letters and digits.</td>
<td>Total Score – Number of items correctly reported, up to a maximum of 30.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Longest – Longest string completed by a subject.</td>
</tr>
<tr>
<td>Reading Span</td>
<td>working memory</td>
<td>Subjects read aloud a series of unconnected sentences. After each sentence, subjects indicate whether the sentence made sense or not (e.g., “the girl sang a song” vs. “the train sang a song”) to prevent rehearsal of items. At the end of a series, they recall the last word of each sentence. The span of the series begins at 6 and increases to 8.</td>
<td>Number of Correct Responses – This is the sum of correct responses given for whether sentences were absurd or not. This score was used for validity purposes – a score of 85% correct or greater was deemed acceptable (which all subjects achieved). This score was not used in subsequent analyses.</td>
</tr>
<tr>
<td></td>
<td>(complex test)</td>
<td></td>
<td>Reading Span (Total) – Total number of words correctly recalled.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Longest – The longest series for which a subject was able to recall the last word of every item in the series.</td>
</tr>
<tr>
<td>Semantic/ Category Fluency</td>
<td>category fluency/ processing speed</td>
<td>Subjects name as many animals, fruits, or vegetables as possible within a 60 seconds.</td>
<td>Total number of correct items named.</td>
</tr>
<tr>
<td>Paired Associate Learning Test (PAL)*</td>
<td>episodic memory</td>
<td>Subjects are presented with 2, 3, 6, and 8 boxes displayed on the screen that open one at a time in a randomized order to reveal a pattern. Respondents must select the box in which each pattern appeared.</td>
<td>Errors Adj.**, Total number of errors made, adjusting for each stage not attempted due to previous failure (the test discontinues if 10 consecutive errors are made at a stage).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Errors, 8 Shapes, Adj.* – Total number of errors made at 8 shapes stage, adjusted if this stage is not reached.</td>
</tr>
<tr>
<td>Stop Signal Task (SST)*</td>
<td>inhibition</td>
<td>Subjects make a 2-choice button response, but withhold their response of a beep is heard on a trial. The timing of the auditory stop signal is set such that the subject is able to stop successfully approximately 50% of the time.</td>
<td>Direction Errors on Stop/Go Trials** – Number of trials in which the wrong button was pressed (left button when the total number of correct responses is displayed).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Proportion of Successful Stops (Last Half) – The number of times the subject stopped successfully divided by the total number of stop signals during the last half of sub-blocks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median Correct Reaction Time on Go Trials** – Median reaction time for Go trials (trials without a beep), in milliseconds.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stop Signal Delay (50%) (last half)** – Stop signal delay at which subject was able to stop 50% of the time.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stop-Signal Reaction Time – Time taken to respond.</td>
</tr>
<tr>
<td>Reaction Time (RTJ)*</td>
<td>motor/ processing speed</td>
<td>Subjects respond to a yellow dot appearing on the screen. In simple reaction time, the dot appears in a circle in the center of the screen, and in 5-choice reaction time, the spot appears in any one of five circles located concentrically to the center of the screen.</td>
<td>5-Choice Reaction Time** – Speed at which subject releases the press pad button in response to the appearance of the yellow dot during the 5-choice reaction time task (speed of cognitive function).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5-Choice Movement Time** – Time taken to touch the screen after the press pad button has been released during the 5-choice reaction time task (speed of motor functions).</td>
</tr>
<tr>
<td>Rapid Visual Information Processing (RVP)*</td>
<td>Sustained visual attention</td>
<td>Digits from 2-9 appear in a box in the center of the screen in a pseudo-random order, at the rate of 100 digits per minute. Subjects are required to make a button press response to all of three target sequences (2-4-6, 3-5-7, or 4-6-8).</td>
<td>A’ – A prime is the signal detection measure of sensitivity to the target, accounting for response bias.</td>
</tr>
</tbody>
</table>
Table 3-2: Summary of cognitive tests and outcome measures. This table lists cognitive tests repeated across the three time points, with a brief description of each test and outcome measures used. Bolded tests assess near-transfer (i.e., working memory ability), whereas the remainder of tests assess far-transfer to other cognitive domains. Tests with * are taken from the Cambridge Neuropsychological Test Automated battery (CANTAB). Outcome measures with ** are reverse coded (such that a lower value reflects a higher score).

3.2.6 Speech Tests

Sentence-Matrix Test (BUG). Stimuli for this task were taken from Kidd, Best and Mason (2008). The words were recorded with neutral inflection so that all possible combinations of words could be used. Although the corpus consists of both male and female speakers, only 8 female voices were used for this experiment. Every sentence had the structure <name verb number adjective noun> (e.g., ‘Bob found three green gloves’). Sentences consisted of a string of words composed by taking one word from each of the 5 categories (e.g., “Bob found three green shoes”). Stimuli were created by combining 3 distinct talkers (including two distractor voices), such that no word was repeated in any category across the three speakers. Subjects were instructed to follow the voice of the talker who said the name “Bob”; one signal with the name ‘Bob’ was present in each trial. Items were mixed at 0 dB SNR. A MATLAB script was used to present the stimuli, at +3 and +6 signal-to-noise ratios (dB SNRs) in four blocks of 25 trials. The first block was provided as practice for subjects and was subsequently removed from analyses. Following each stimulus presentation, subjects were instructed to select each of the target words from a five by eight matrix of options (i.e., 8 monosyllabic words from each of five different word-type categories), but the first column was not scored (since the target
name was always ‘Bob’). Dependent variables were percentage of words correctly reported at both +3 and +6 dB SNR.

The closed set nature of the task ensures that contextual information and the load on working memory is constant across stimuli. This task has the advantage of having excellent psychometric properties (e.g., intelligibility is not confounded by a tendency to guess), and item effects are substantially weaker than for open-set materials, reducing within-subject variability, in principle making any change over time within subjects easier to detect.

**Context Sentences.** I assessed the use of context by asking subjects to report words from sentences with and without contextual information, each presented with different sentences produced by two competing (same-sex) talkers. The task comprised of 96 Semantically coherent (high-context) sentences, which had supportive contextual information (e.g., “He always read a book before going to bed”), and 96 semantically anomalous (low-context) sentences, which were syntactically correct but nonsensical and were created by replacing the content words of coherent sentences with other words matched in part of speech and word frequency (e.g., “Her good slope was done in carrot”; Davis et al., 2011). Distractor sentences consisted of common, everyday sentences (e.g., “the student tried to move the desk”).

Sentences were recorded by three individuals who were raised in southern Ontario and had an accent judged by the experimenters to be non-atypical for the region (all were female; one individual recorded the target sentences and the other two recorded the distracter sentences). Sentences were divided into three sets of 64 sentences (32 at each level of context), matched across sets for average length. The order of the sentence sets
was counterbalanced across participants, so that approximately equal numbers heard each set at each testing time point. Care was taken to avoid repetition of distractor sentences across sets where possible, but target sentences were never repeated (distractor sentences were never repeated within a set).

Stimulus mixing and presentation was accomplished using MATLAB. Target and distractor sentences were all normalized to have the same RMS power. The two distractor sentences for each trial were first combined and mixed at 0 dB SNR, then this result was normalized with the target and combined with the target at +3 and +6 dB. Subjects were instructed to attend to a target voice, identified as the voice to which they had listened during 10 practice sentences (no distractors were presented during practice). On each trial, subjects were instructed to type all of the words they could understand from the target sentence, in the correct order. Word report was assessed as the proportion of words correctly reported in each sentence. As in Wayne and Johnsrude (2012) and Davis and colleagues (2005), words were scored as correct if the written form perfectly matched the word produced in the sentence. Morphological variants were scored as incorrect, whereas homonyms and misspellings were scored as correct. Words were scored correct if they were reported in the correct order, even if intervening words were absent or incorrectly reported. All subjects correctly reported practice sentences (which included both high- and low-context sentences), indicating that word report is probably not limited by poor memory. Dependent variables were percentage of words correctly reported for both high- and low-context sentences at +3 and +6 dB SNR. Benefit from context was operationally defined as word report for high-context sentences minus word report for low-context sentences, at each SNR.
3.3 Results

Subjects in the two training groups (Adaptive then Placebo; and Placebo then Adaptive) did not significantly differ on any of the outcome measures reported in Table 3-3 at baseline (at T0). Two subjects withdrew from the study for health reasons after completing five weeks of placebo training and the T1 test session; adaptive training data and T2 test data are unavailable for these subjects. Due to error, one subject completed 7 weeks of adaptive training followed by 3 weeks of placebo training. I included data from this subject, since more training may have increased the likelihood of finding any effect (which I did not observe anyway). Means and standard deviations for all outcome measures are reported in Table 3 (see Figures 3-3 and 3-4 for accuracy and word-report data for speech tests).

Figure 3-3: Accuracy scores for the BUG (sentence matrix) speech task across the three study time points. Error bars reflect standard error of the mean.
Figure 3-4: Percentage word scores for high- and low-context sentences across the three study time points. Error bars reflect standard error of the mean.

I first established the efficacy of Cogmed Working Memory training by examining evidence of improvement on adaptive training. Training-related changes on the measures from the cognitive and speech-in-noise tests were then assessed using repeated-measures ANOVAs (Jacoby & Ahissar, 2013). Results are reported across all subjects, and for the normally hearing and hearing-impaired groups separately. Finally, I examine the correlations between speech tests and tests of working memory, and between working memory (including Reading Span) and other cognitive domains. Correlations were computed using Spearman’s rho (unless otherwise noted) and all corrections were completed using the Benjamini-Hochberg Procedure.
<table>
<thead>
<tr>
<th></th>
<th>Adaptive - Placebo</th>
<th>Placebo-Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T0</td>
<td>T1</td>
</tr>
<tr>
<td><strong>Screening Tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montreal Cognitive Assessment</td>
<td>27.85 (1.77)</td>
<td>27.15 (1.57)</td>
</tr>
<tr>
<td>Raven's Raw Score</td>
<td>50.92 (6.51)</td>
<td>51.27 (6.02)</td>
</tr>
<tr>
<td>Speech Spatial Qualities: Speech</td>
<td>8.21 (0.77)</td>
<td>6.95 (1.24)</td>
</tr>
<tr>
<td>Speech Spatial Qualities: Qualities</td>
<td>8.79 (0.55)</td>
<td>8.04 (0.89)</td>
</tr>
<tr>
<td><strong>Near-Transfer (Working Memory) Tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial Working Memory (Between Errors)</td>
<td>23.46 (15.59)</td>
<td>20.46 (13.35)</td>
</tr>
<tr>
<td>Spatial Working Memory (Strategy)</td>
<td>30.31 (7.47)</td>
<td>30.38 (5.94)</td>
</tr>
<tr>
<td>Spatial Span (Forward)</td>
<td>6.23 (1.17)</td>
<td>6.85 (1.46)</td>
</tr>
<tr>
<td>Spatial Span (Reverse)</td>
<td>5.62 (1.39)</td>
<td>6.54 (1.20)</td>
</tr>
<tr>
<td>Letter-Number Sequencing (Raw Score)</td>
<td>20.39 (2.14)</td>
<td>21.23 (2.49)</td>
</tr>
<tr>
<td>Letter-Number Sequencing (Longest)</td>
<td>6.00 (0.91)</td>
<td>6.15 (1.14)</td>
</tr>
<tr>
<td><strong>Far-Transfer Tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Span</td>
<td>21.15 (3.29)</td>
<td>24.65 (5.67)</td>
</tr>
<tr>
<td>Rea Reading Span (Longest)</td>
<td>1.62 (1.56)</td>
<td>2.77 (1.30)</td>
</tr>
<tr>
<td>Fluency</td>
<td>19.31 (3.97)</td>
<td>22.00 (7.43)</td>
</tr>
<tr>
<td>Paired Associate Learning (Errors, Adj.)</td>
<td>18.08 (16.98)</td>
<td>14.92 (13.47)</td>
</tr>
<tr>
<td>Paired Associate Learning (Errors, 8 Shapes Adj.)</td>
<td>13.92 (13.09)</td>
<td>11.15 (11.46)</td>
</tr>
<tr>
<td>Stop Signal Task (Direction Errors)</td>
<td>5.54 (12.35)</td>
<td>6.69 (10.03)</td>
</tr>
<tr>
<td>Stop Signal Task (Prop. Successful Stops)</td>
<td>0.51 (0.08)</td>
<td>0.51 (0.09)</td>
</tr>
<tr>
<td>Stop Signal Task (Median Correct, Go Trials)</td>
<td>540.77 (133.89)</td>
<td>503.85 (151.14)</td>
</tr>
</tbody>
</table>
Table 3-3: Means and standard deviations for all outcome measures. Note that reaction times are provided in milliseconds and speech scores are presented as percentage of words correctly reported (context sentences) or selected (BUG).

### 3.3.1 Improvement on Cognitive Training (Cogmed)

The average Start Index (performance on days 2 and 3) was 85.38 (SD = 9.84, min = 70, max = 105) and the Maximum Index (performance on the two best training days) was 108.75 (SD = 13.53, min = 89, max = 142). The Cogmed Index-Improvement score, which compares these two (Maximum vs. Start), was 23.5 (SD = 7.95, min = 12, max = 49; the normal range is 18-42); this improvement was significant, and all subjects’ scores improved over the course of adaptive training. The Index-Improvement score did not appear to depend on whether adaptive training was first or second (i.e., before or after placebo training).
Crucially, the average improvement score is comparable to (or larger than) those reported in studies reporting near-transfer (e.g., Gropper, Gotlieb, Kronitz, & Tannock, 2014), suggesting that training was effective.

The average Raven’s score was in the 67th percentile (minimum = 21st percentile, maximum = 99th percentile). The Raven’s score significantly correlated with the Start Index score ($r_s = .64$, $p = .001$, corrected), but not with the Index-Improvement score or the Maximum Index, indicating that intelligence was related to initial performance on working-memory training, but not to training gains. The average Letter-Number Sequencing (a widely accepted test of verbal working memory) scaled score at the start of the experiment (T0) was 10.68 ($SD = 1.92$; note that 10 is the population mean scaled score, indicating average baseline working memory ability).

3.3.2 Transfer to Cognitive Tests

For cognitive test outcome measures for which normative data were available (this excludes scores for Reading Span, Paired Associate Learning, 8 Shapes Adjusted, Spatial Span Reverse, and the Stop Signal Task), the average Z-score across tests was 0.67 ($SD = 0.19$). On all tests, on average subjects performed within 1.5 standard deviations of the mean (Min. = -0.72, Max. = 1.37), indicating that older adults participating in my study consistently performed within acceptable limits.

I analyzed the data for each cognitive test using repeated-measures ANOVAs, with Training Group (adaptive before placebo or placebo before adaptive) as a between-subjects variable, and Time as a within-subjects variable with 3 levels. The Group by Time interaction reflects both training and the interaction between order (placebo, then training or training, then placebo) and training (either of which indicate that training has had some effect). The
main effect of time reflects both practice, as well as training (see Figure 2). The main effect of training group was non-significant for all dependent variables, suggesting that participants in the two groups were drawn from the same population.

Significant practice effects (evident as a main effect of time with increasing values over time) were obtained on Reading Span, Letter-Number Sequencing score, Stop Signal Task Stop Signal Response Time, Spatial Working Memory Strategy score, Reaction Time 5-choice Movement Time and 5-choice Reaction Time, as well as Spatial Span Forward and Reverse (see Table 3-4 for statistics). The interaction between Group and Time was non-significant for all measures drawn from all the cognitive tests listed in Table 3, meaning that pre-post improvement did not differ between placebo and adaptive training, suggesting that the general improvements in performance are practice effects. Analyzing the hearing and hearing-impaired groups separately did not change the pattern of results. These results indicate no evidence for training-related cognitive improvement in older adults.

I conducted additional exploratory analyses on training-related transfer, using the baseline score and training group (adaptive vs. placebo) as separate predictors in a linear regression. Data were recoded to match a between-subjects design, such that each subject contributed two data points for each of post-training performance and baseline performance, with a dummy variable coding for adaptive vs. placebo training (e.g., For the adaptive-placebo group, T0 was taken as baseline in the adaptive condition, and T1 was baseline for the placebo condition, with T1 as the post-training performance in the adaptive condition, and T2 for the placebo condition). Although this analysis was more sensitive than the repeated-measures ANOVA, I did not find any significant effect of training group, even at an uncorrected level.
Table 3-4: F-test statistics for main effect of time for all dependent variables, and uncorrected for post-hoc comparisons.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>$F_{(2, 44)}$</th>
<th>$p$</th>
<th>$\eta^2$</th>
<th>Post-hoc comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading Span Score</td>
<td>13.26</td>
<td>&lt;0.001</td>
<td>0.38</td>
<td>T2 &gt; T0**, T1 &gt; T0**</td>
</tr>
<tr>
<td>Letter-Number Sequencing Score</td>
<td>5.81</td>
<td>&lt;0.05</td>
<td>0.21</td>
<td>T2 &gt; T0*</td>
</tr>
<tr>
<td>Stop-Signal Task Stop Reaction Time</td>
<td>4.46</td>
<td>&lt;0.05</td>
<td>0.17</td>
<td>T2 &gt; T0*</td>
</tr>
<tr>
<td>Spatial Working Memory Strategy Score</td>
<td>3.16</td>
<td>0.052</td>
<td>0.13</td>
<td>T2 &gt; T1*</td>
</tr>
<tr>
<td>Reaction Time Five-Choice Movement Time</td>
<td>6.19</td>
<td>&lt;0.05</td>
<td>0.22</td>
<td>T2 &gt; T0*</td>
</tr>
<tr>
<td>Reaction Time Five-Choice Reaction Time</td>
<td>11.23</td>
<td>&lt;0.001</td>
<td>0.34</td>
<td>T2 &gt; T0**, T2 &gt; T1**</td>
</tr>
<tr>
<td>Spatial Span Forward Score</td>
<td>3.94</td>
<td>&lt;0.05</td>
<td>0.15</td>
<td>T2 &gt; T0*</td>
</tr>
<tr>
<td>Spatial Span Reverse Score</td>
<td>4.37</td>
<td>&lt;0.05</td>
<td>0.17</td>
<td>T2 &gt; T0*</td>
</tr>
<tr>
<td>High Context Sentences +3 dB SNR</td>
<td>8.61</td>
<td>&lt;0.001</td>
<td>0.28</td>
<td>T0 &gt; T1*</td>
</tr>
</tbody>
</table>

3.3.3 Transfer to Speech Tests

Data for speech tests are presented in Figures 3 and 4. As expected, the effect of SNR was significant for all three tasks, with +6 dB SNR being more intelligible than +3 dB SNR (see Table 3). In addition, significantly more words were reported for high-context sentences (proportion of words reported correctly: $M = .78$, $SD = .14$, Min = .34, Max = .93) than for low-context sentences ($M = .52$, $SD = .15$, Min = .17, Max = .74; $t(25) = 16.81$, $p < .001$; similar to Davis et al, 2011). The average proportion of words correctly reported on the BUG at +6 dB SNR was .44 ($SD = .10$, Min = .28, Max = .70) and .36 ($SD = .09$, Min = .22, Max = .56) at +3 dB SNR. These results suggest that performance was not at ceiling or floor for either of the speech tasks. I conducted a repeated-measures ANOVA on BUG performance and on word-report scores for high- and low-context sentences separately, as well as on the
context benefit (high-low) score, with training group as a between-subjects factor and time and SNR (+3 dB and +6 dB) as within-subjects factors. Speech scores did not improve as a result of training (far-transfer); the interaction between Group and Time was non-significant for all three tasks. I also analyzed high- and low-context scores together; even with increased power, there was no effect of training group. A lack of training-related transfer to speech tests is unsurprising given the lack of transfer to other cognitive tests. The main effect of Group was non-significant across all three tasks, but there was a main effect of Time for high-context sentences presented at +3 dB SNR \(F(2,44) = 8.61, p < .001, \eta^2_p = .28\), reflecting practice effects. Conducting analyses separately on hearing and hearing-impaired subjects did not change the overall pattern of results, and, similar to cognitive tests, my exploratory analyses (see previous section for details) did not reveal any significant training-related effects.

### 3.3.4 Correlations between Speech and Cognitive Tests

Since scores on cognitive tests did not depend on training (as demonstrated in a previous section), participants’ scores on each test were averaged across all three time points (and across the two SNRs for the BUG and high- and low-context sentence tests) to yield more reliable estimates of performance.

I computed correlations between average context-benefit scores (difference in word report for high- and low-context sentences) and tests of working memory (see Table 4-5). None of the correlations with context benefit survived correction. Although there was a trend towards significance for Reading Span Longest score \(r_s = -.42\), Letter-Number Sequencing score \(r_s = -.41\), and Letter-Number Sequencing Longest score \(r_s = -.44\), the negative direction of these trends was contrary to expectations, as they appeared to be driven by a positive relationship between working memory and intelligibility of low-context sentences.
specifically. In fact, word report for low-context sentences significantly correlated with Letter-Number Sequencing Average score \( (r_s = .58, p < .05) \) and Letter-Number Sequencing Longest score \( (r_s = .48, p < .05, \text{both corrected}) \). There was also a trend towards significance for correlations between low-context sentence word report and Reading Span total \( (r_s = .45) \) and Longest scores \( (r_s = .39; \text{these did not survive correction for multiple comparisons}) \). There was no relationship between word-report for high-context sentences and tests of working memory, even when examined at only the more difficult +3 dB SNR to minimize ceiling effects, with the exception of a trend in the predicted direction for Spatial Working Memory Errors \( (r_s = -.38; \text{the apparent negative correlations reflect the fact that low values are indicative of better performance}) \). The difference between correlations for low- and high-context sentences was significant for both Letter-Number Sequencing Average score (Steiger’s Z-test; \( Z = 2.48, p < .05 \)) and Letter-Number Sequencing Longest score \( (Z = 2.70, p < .05) \).

This pattern of results suggests that working memory may facilitate intelligibility particularly when contextual information is unavailable, although this may be an artifact of the word-report intelligibility measure I used. Word report is a rather unnatural assessment of speech intelligibility. Listeners generally do not need to repeat back sentences in everyday communication, and it is possible that word report scores for low-context sentences may load more highly on working memory because they are harder to remember than high-context sentences. I evaluated this hypothesis by examining the Pearson correlation between sentence length and word-report scores, for high- and low-context sentences separately. These were then averaged, within-subjects, across test time points. A repeated-measures ANOVA, with high- vs. low-context sentences and time as within-subjects factors, and training group as a
between-subjects factor revealed a main effect of level of context \((F(1,22) = 30.17, p < .001, \eta^2 = .58)\): sentence length was significantly more negatively correlated with word report for low- compared to high- context sentences (low-context \(M_r = -.26, SE = .02\); high-context \(M_r = -.08, SE = .03\)). The interaction between Context Level and Time trended towards significance \((p = .05\), sentence word report for low-context sentences was significantly more negatively correlated with sentence length at all three time points). All other main effects and interactions were non-significant. This result is consistent with low context sentences being more difficult to maintain in memory, and may account for the correlation between low-context sentences and working memory.
Table 3-5: Correlations between speech tests and (select) cognitive tests. Note that significance levels displayed here are uncorrected for multiple comparisons. Abbreviations: RS (Reading Span), RS L (Reading Span Longest Span), LNS (Letter-Number Sequencing), LNS L (Letter-Number Sequencing, Longest Sequence) SWM (Spatial Working Memory), SSP Fwd (Spatial Span Forward), SSP Rev (Spatial Span Reverse), SST SS (Stop Signal Task, Proportion of Successful Stops), BUG (Sentence-matrix speech in noise task), H Context (High context sentence word report scores), L Context (Low context sentence word report scores), H-L Context (High – low context sentence word report scores).

I did not observe any significant correlations between working memory measures and the BUG speech task scores. Interestingly, scores on the speech tasks did not significantly correlate with the Raven’s nor with the SSQ Speech or SSQ Qualities measures, although there was a trend for a correlation between SSQ Speech and low-context sentence word-report scores ($r_s = .36$).

### 3.3.5 Correlations between Reading Span and Cognitive Measures

Correlations between Reading Span score (i.e., the total number of words recalled) and speech-in-noise performance in previous reports have generally been taken as evidence for the
involvement of working memory in speech-in-noise performance (Davies-Venn & Souza, 2014; Tun, Weingfield, & Stine, 1991; Akeroyd, 2008; Zekveld et al., 2012, Zekveld, Rudner, Johnsrude & Rönnberg, 2013, Besser, Koelewijn, Zekveld, Kramer & Festen, 2013, but see Schoof & Rosen, 2014; Humes & Coughlin, 2009). However, the Reading Span test is a complex test that relies on other cognitive domains, in addition to working memory. I examined the cognitive architecture supporting Reading Span performance by correlating measures on this test with other measures of working memory (Letter-Number Sequencing Average score, Spatial Working Memory score, Spatial Span Forward, and Spatial Span Reverse scores), non-verbal reasoning (Raven’s), processing speed (Reaction Time 5-choice Reaction Time), memory (Paired Associate Learning, 8-Shapes Corrected score) and inhibition (Stop Signal Task Proportion of Successful Stops). As seen in Table 5, the Reading Span score correlated with the Paired-Associate Learning, 8-Shapes Total Errors Adjusted score ($r_s = -.54, p < .05$, corrected for multiple comparisons). There was a trend towards significance for the correlations between Reading Span score and Letter-Number Sequencing score ($r_s = .43$), Spatial Span Reverse ($r_s = .37$), as well as Stop Signal Task Proportion of Successful Stops ($r_s = -.34, p > .05$, after correction for multiple comparisons). The lack of significance here is probably due to insufficient power, but the pattern of results indicates, not surprisingly, that the Reading Span Test loaded on tests of working memory, episodic visual memory, and inhibition.

3.4 Discussion

3.4.1 Generalizability of Cognitive Training

As expected, scores on the Cogmed working-memory tests increased over time in the adaptive-training sessions, and both the adaptive and placebo training groups showed practice
effects on several test measures. However, I observed no evidence of transfer of working-memory training, even to tests that should tap the same cognitive domains as training (i.e., other working memory tests), and even when uncorrected for multiple comparisons. Ultimately, my results demonstrate no evidence that Cogmed cognitive training improved cognitive functioning as measured by my tests, or improve speech-in-noise comprehension in older adults.

My results apparently contradict those documenting near- or far-transfer of working memory training, including Cogmed, to other cognitive domains in both older and younger adults (Holmes, Gathercole, & Dunning, 2009; Klingberg, 2010; Klingberg et al., 2005; Olesen, Westerberg, & Klingberg, 2003; Brehmer et al., 2011; Brehmer, Westerberg & Backman, 2012; Hindin & Zelinski, 2012; Karr, Areshenkoff & Garcia-Barrera, 2014; Melby-Lervag & Hulme, 2014; Kronenberger et al., 2011; Karbach & Verhaeghen, 2014). In attempting to reconcile my null results for generalization of cognitive training with significant generalization noted elsewhere, it is important to note that effects of training may be specific to assessment measures used. Others have suggested that even in the absence of direct strategy instruction, cognitive strategies may be task specific (Dunning & Holmes, 2014), reflecting limited generalizability of training gains. Alternatively, Cogmed training may specifically benefit those with below-average working memory ability (Zinke et al., 2014). The vast majority of my subjects had at least average working memory ability (on Letter-Number Sequencing, all subjects had baseline scores at or above the 25th percentile (the low-average range), and subjects performed on average, half a standard deviation better than the norm on cognitive tests).
Although a lack of motivation could in principle explain a lack of efficacy, my older participants exhibited acceptable improvement scores on training, commensurate with studies reporting evidence of transfer (Gropper et al., 2014) and they appeared highly motivated. I did not compensate them for training time; only for the time spent in the testing sessions, and I experienced a very low rate of attrition. It is also possible that my study was underpowered. However, my study had more than the mean number of adults assessed in the studies reviewed in the meta-analysis of cognitive training studies in older adults by Karbach & Verhaeghen (2014; 21.34; SD = 13.98); this analysis revealed significant effects for training, although none of the studies used Cogmed (these studies also typically used a between-subjects design, a less powerful design than my within-subjects design).

Instead, my results are consistent with an emerging body of literature challenging the effectiveness and generalizability of cognitive training, including Cogmed (e.g., Gathercole, 2014; Chacko et al., 2014; Shipstead et al., 2012; Melby-Lervag & Hulme, 2012; Jacoby & Ahissar, 2013; Lampit et al., 2014). At present, the reasons for inconsistencies between these studies and those reporting evidence of generalization (e.g, Kronenberger et al., 2011) is unclear. However, a recent meta-analysis and review of cognitive training in 5000 older adults found that home-based training was ineffective compared to group-based training, and that training more than three sessions a week was less effective than training three or fewer times per week, perhaps due to cognitive fatigue (Lampit et al., 2014). My subjects trained five times per week at home, which may account for my null findings. However, it is also important to note that effect sizes reported in this meta-analysis are small (Hedge’s $g = 0.22$).

My cross-over, within-subjects design appears to be unique in the literature; to my knowledge, my design in which subjects receive both placebo and adaptive training (in a
counterbalanced manner) has not been used in studies evaluating the efficacy of Cogmed Working Memory Training (Holmes, Gathercole, & Dunning, 2009; Klingberg, 2010; Klingberg et al., 2005; Olesen, Westerberg, & Klingberg, 2003; Brehmer et al., 2011; Brehmer, Westerberg & Backman, 2012; Hindin & Zelinski, 2012; Karr, Areshenkoff & Garcia-Barrera, 2014; Melby-Lervag & Hulme, 2014; Kronenberger et al., 2011; Chacko et al., 2014), or any other form of cognitive training in older adults. This design is more rigorous than the traditional between-subjects approach because it controls for motivational/engagement effects resulting from training. More specifically, it is possible that motivation and engagement is higher in the training group, owing to the higher degree of effort necessitated by the training regimen, compared to even an active placebo group (since the task performed by the placebo group is usually easier). Thus, where adaptive and placebo training are not counterbalanced within subjects, the adaptive group may show training gains simply as a result of effort and engagement, rather than cognitive training (see also Jacoby & Ahissar, 2013 for a similar argument).

This view is supported by the apparent absence of significant differences between active (placebo training) and passive (no training) control groups in other studies (see Melby-Lervag & Hulme, 2013; Karbach & Verhaeghen, 2014 for reviews). This finding that placebo training is equivalent to no training at all suggests that the placebo condition in Cogmed is not sufficiently demanding to control for effects related to effort or engagement during training. Thus, it is possible that evidence of transfer of training to cognitive tests in studies comparing adaptive to placebo (or passive training) might reflect gains due to task engagement, rather than the content of the training regimen, suggesting that these gains might be acquired through engagement with non-specific cognitive tasks. Moreover, the relative superiority of
group-based (or lab-based) training over at-home training (Lampit et al., 2014) also suggests that task-engagement, or social interaction, may drive cognitive training gains.

### 3.4.2 Relationship between Working Memory and Speech in Noise

I observed that working memory measures correlated with the amount of benefit to word-report obtained through provision of greater context, but this relationship is explained by a positive correlation specifically with low-context word-report scores. The cognitive load imposed by low-context (semantically anomalous) sentences (as listeners strive fruitlessly after a coherent meaning) may limit processing resources available for accurate perception of further words in the utterance, reducing word report. The greater the cognitive capacity, the more resources are left over from this futile semantic integration process for accurate perception. Moreover, low-context sentences may place a greater strain on working memory since the lack of meaningful associations among the words in the sentences means that fewer retrieval cues for any given word are available. As word report was higher for high-context sentences compared to low-context sentences, it is also possible that low-context sentences were more effortful (as reflected by greater recruitment of working memory resources) as a result of being more difficult to maintain in memory for immediate report. This explanation is supported by my finding of a significant negative association between sentence length and word report for low-context sentences but not for high-context sentences. It is possible that correlations between working memory and word report for high-context sentences may emerge more strongly in longer streams of perceptual inputs (i.e., longer utterances) due to increased processing demands.
3.4.3 What Does the Reading Span Test Measure?

My results warrant some caution in using the commonly administered Reading Span test (Rönnberg et al., 1989) as an exclusive test of working memory. The Reading Span test correlated with memory, with a trend for correlation with measures of working memory and inhibition. The pattern of results suggests that Reading Span may load on both working memory and general cognitive functioning, including episodic visual memory. Given the documented contributions of more general cognitive functioning to speech perception (e.g., Heald & Nusbaum, 2014; Wingfield & Tun, 2007; Arlinger et al., 2009), as well as the high correlations between working memory and non-verbal intelligence (e.g., Engle, Tuholski, Laughlin & Conway, 1999), future research should take care to parcel contributions of working memory from other cognitive processes. This can be achieved by using more domain-specific tests of working memory (e.g., WAIS Letter-Number Sequencing and other simple span tests), as well as using multiple, converging measurements of working memory. It would also be worthwhile to compare the widely-used Rönnberg et al. (1989) version of the Reading Span test with the Daneman and Carpenter (1980) version (as well as others) to verify whether they can justifiably be used interchangeably.

3.5 Summary and Future Directions

Commercial cognitive training software, including Cogmed Working Memory Training (Klingberg, Forssberg & Westerberg, 2002), is being aggressively marketed to the general population. My study adds to the growing body of literature suggesting that cognitive training, a multi-million dollar industry, may not be as effective as initially hoped (e.g., Melby-Lervag and Hulme, 2012; Shipstead, Hicks & Engle, 2012). My study lends further support to the idea that working memory is important in speech comprehension. Although my
results suggest that individuals with better working memory capacity may be better able to compensate for degraded auditory input when contextual information is unavailable, this may be an artifact of my word report measure. Future studies should extend these findings to more naturalistic paradigms, such as through comparing reaction time as a function of cognitive load in dual-task paradigms for high- and low-context sentences. My study also suggests that exclusive reliance on the Reading Span as a measure of working memory may be problematic, since in my study Reading Span correlated significantly only with a measure of episodic visual memory, but not working memory. Future research should also aim to evaluate the impact of cognitive training on everyday cognitive functioning, also controlling for levels of effort, which are typically not matched in an active, low-level task control group (including the placebo condition of Cogmed). Working memory training, if effective, would be a cornerstone of rehabilitation programs for older adults with communication difficulties, but the evidence suggests that we have not yet found the magic ingredient.
Chapter 4 - The Role of Attention in Audiovisual Speech Integration
4.1 Introduction

Human communication in its most natural form involves both visual and acoustic correlates, and listeners frequently make use of bimodal speech cues that are readily available in face-to-face conversation. A wealth of research indicates that auditory and visual information complement each other in speech perception and that audiovisual speech is more intelligible than auditory-alone speech (e.g., Grant & Braida, 1991; Grant & Seitz, 2000; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Sumby & Pollack, 1954).

The use of visual speech information is especially advantageous when the speech signal is impoverished. Visual speech cues can be used to help recover the message content when the acoustic speech signal is degraded, enhancing its intelligibility in noisy or multi-talker environments or in the presence of an unfamiliar accent (Macleod & Summerfield, 1990; Reisberg, et al., 1987; see also, Rosen, Faulkner, & Wilkinson, 1999). Moreover, the auditory and visual components of speech are complementary such that segment distinctions that are harder to hear are easier to see, and vice versa (Summerfield, 1987; Rosenblum, 2008).

The superior temporal sulcus (STS), particularly the posterior region, is generally considered a site of AV speech binding (Calvert et al., 2000; Calvert 2001; Sekiyama et al., 2003; Beauchamp et al., 2004ab; van Atteveldt et al., 2004; 2007ab; Lee & Nopenney, 2011; but see Hocking & Price, 2008), although both the mSTS (Stevenson et al., 2010) and aSTS may be recruited. Nath and Beauchamp (2012) showed a correlation between STS response and likelihood of experiencing a McGurk effect, and that TMS to this region disrupts the McGurk effect (Beauchamp, Nath & Pasalar, 2010), suggesting that the superior temporal
region is implicated in audiovisual speech integration. The fusiform gyrus, which is robustly involved in face perception, has also been shown to be recruited in binding of audiovisual speech signals, particularly in degraded or impoverished listening conditions (McGettigan et al., 2012; Giraud & Truy, 2002; Kawase et al., 2005).

Although visual speech information can boost comprehension when the acoustic signal is degraded or when background noise is present, whether the successful binding of auditory and visual signals requires attention is unclear. Earlier behavioural studies investigating the role of attention in AV integration converged on an automatic, pre-attentive processing account of AV integration (Bertelson, Vroomen, de Gelder, & Driver, 2000; Driver, 1996; Massaro, 1987; McGurk & MacDonald, 1976; Soto-Faraco, Navarra, & Alsius, 2004; Vroomen, Bertelson, & de Gelder, 2001), which garnered further support through neurophysiological studies (Colin et al., 2002; Mottonen, Krause, Tiippana, & Sams, 2002; Samson et al., 2001).

More recent evidence, however, indicates a limit to this automaticity (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007; Lee & Noppeney, 2011; Munhall, ten Hove, Brammer, & Pare, 2009; Tiippana, Andersen, & Sams, 2004; Tuomainen, Andersen, Tiippana, & Sams, 2005). In the McGurk effect, two distinct syllables presented to the auditory and visual modality produce a unified (or ‘fused’) perceptual AV token. Tuomainen et al. (2005) demonstrated a robust McGurk effect for sine-wave speech only when participants perceived sine-wave speech as being speech-like (note that subjects are able to understand sine-wave speech if they are explicitly instructed of its speech-like nature; Remez et al., 1981). The authors argued that attention to the acoustic features relevant for phonemic classification modulated listeners’ ability to integrate the two signals, since the
presence of the McGurk effect was modulated by expectations about the speech-like nature of auditory stimuli. However, since auditory intelligibility differed between the two conditions (i.e., whether the McGurk effect was experienced), these findings could also be accounted for by top-down effects on intelligibility of unimodal auditory stimuli (see also, Lee & Noppeney, 2011), rather than attentional modulation of AV integration processes.

Others have argued that earlier studies supporting pre-attentive processing of AV integration employed distractor tasks that did not sufficiently interfere with attentional processing of audiovisual speech information (Alsius et al., 2005). Unless attentional resources are exhausted by the distractor task, it is possible that remaining resources may ‘spill over’ to secondary tasks (Lavie, 1995). Tiippana et al. (2004) employed a distractor task in which subjects attended to a partially transparent leaf floating across the face. They found that the McGurk effect was attenuated in this distractor condition relative to a baseline audiovisual condition. The authors claim that their results provide evidence for the role of attention in audiovisual integration. However, although the leaf did not actually cover the mouth of the speaker, their results are also consistent with the explanation that the artificial leaf distractor task impeded listeners’ ability to track visual speech information (the authors did not employ eye-tracking methods). Thus, their results may reflect a disruption of sensori-perceptual rather than attentional processes in audiovisual integration.

Similarly, Alsius et al. (2005) demonstrated a reduction in individuals’ susceptibility to the McGurk effect when they were asked to perform a demanding distractor task (the McGurk effect is frequently used as an index of audiovisual speech integration; e.g., Erickson et al., 2014). Alsius and colleagues observed reduced audiovisual fusion responses (McGurk) when attending to both visual and auditory distractors. Crucially, unimodal task performance
was not impeded by the dual task, indicating that the task did not interfere with unimodal processing, only bimodal integration. Subjects in the dual-task condition with the auditory distractor produced more auditory-based responses (indicating disruption of AV integration processes), rather than visually based responses (indicating effects of poorer unimodal intelligibility). In addition, the distractor task did not interfere with unimodal auditory processing (i.e., no difference in unimodal performance for single vs. dual task). Thus, these results provide compelling evidence that the McGurk effect is sensitive to high attentional loads (see also, Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010).

Thus far, studies directly examining the influence of selective attention on audiovisual integration have focused on McGurk stimuli (Nath and Beauchamp, 2011; 2012; Benoit et al., 2010; Sekiyama et al., 2003; Szycik et al., 2012). Although the McGurk effect, which occurs for artificial stimuli, may rely on different mechanisms than those employed for everyday AV speech integration (Alsius & Munhall, in prep), the data on McGurk suggest that such everyday integration may not be as automatic as initially thought, but may depend on selective attention. Neuroimaging, such as fMRI, provides a window into how the brain may differentially process audiovisual speech under different attentional conditions.

The aim of this study is to elucidate the role of selective, task-based attention in audiovisual integration and identify associated neural correlates using (non-McGurk) naturalistic, everyday sentence utterances. I compare BOLD activation in speech-sensitive regions in two main conditions using spectrally degraded stimuli: a condition in which the AV signals are congruent and a condition in which the AV signals are mismatched (i.e., the unimodal signals are incongruent). Since AV congruence is anticipated to enhance intelligibility, regions sensitive to intelligibility are expected to differentiate these two
conditions when attention is directed to speech. Whether such differences remain when attention is directed to a distractor is the question under test; if the increase in activity in speech-sensitive cortices for congruent compared to incongruent speech is smaller under distraction than under full attention, this would suggest that AV integration is sensitive to attention.

4.2 Methods

4.2.1 Participants

I tested 24 individuals between 18 - 31 years of age (mean = 21.42 years, SD = 3.35 years; 5 males). A separate group of 24 participants (mean age = 20.5 years, SD = 2.53 years; 7 males) were tested to pilot the materials and procedure. The procedure in the pilot was identical to that described below, except that it took place in a single-walled sound booth (Eckel Industries) using circumaural headphones (Sennheiser HD 280 Pro) and participants typed out all of the words they could report from degraded utterances (instead of indicating with a keypress whether they understood the gist or not).

All participants were right-handed, native-English speakers, with self-reported normal hearing, normal or corrected-to-normal vision, and no known attentional or language processing impairments. Imaging participants reported no history of seizures or psychiatric or neurological disorders, and no current use of any psychoactive medications. Participants complied with magnetic resonance imaging safety standards: they reported no prior surgeries involving metallic implants, devices, or objects. This study was cleared by the Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (Kingston, ON, Canada). Informed written consent was obtained from all participants.
4.2.2 Experimental Design

I employed a sparse-imaging design to avoid acoustic confounds associated with continuous echoplanar imaging (Edmister et al., 1999; Hall et al., 1999). Stimuli were presented in the 7 second silent gap between successive 2 second volume acquisitions (stimuli were temporally centered within this 7 second stimulus interval). Before every trial, I cued participants to attend to one of two simultaneously presented auditory stimuli: either to the sentence [speech stimulus (SP)] or to the auditory distractor sequence (AD). The AD sequence was comprised of 400-ms narrow-band noise bursts; see Materials section below for details. Participants performed a yes-no decision task associated with the attended stimulus immediately after every trial. The speech stimulus on every trial was presented at one of two levels of clarity, crossed with one of three types of simultaneous visual speech information (matching visual speech, mismatching visual speech, or no visual speech [i.e., auditory-only presentation]). Together, these yielded a factorial design with 12 conditions (2 levels of attention, 2 levels of speech clarity, and 3 levels of audiovisual type). Two baseline conditions were included: a silent baseline in which participants simply viewed a fixation cross, as well as a “scrambled” stimulus containing modified, unintelligible auditory and visual speech information (see below).

4.2.3 Materials

4.2.3.1 Speech

Sentence stimuli were identical to those used by Wild et al. (2012). I arranged 216 meaningful English sentences (e.g., “She always read a book before going to bed.”) into 12 18-item sets that were matched on mean number of syllables (mean = 12.01, SD = 0.34), words (mean = 9.25, SD = 0.17), and logarithmic sum of word frequency (mean = 6.20, SD =
0.07; van Heuven, Mandera, Keuleers & Brysbaert, 2014). Each sentence set was assigned to one of twelve conditions, counterbalanced across subjects to eliminate item-specific effects.

I recorded sentences audiovisually from a young, female native-English speaker who spoke in a way deemed not to be atypical for the region of southern Ontario. Stimuli were recorded in a single-walled sound booth (Eckel Industries) using an AKG C1000S microphone with an RME Fireface 400 audio interface (sampling at 16 bits, 44.1 khz). The video component was extracted from the audiovisual utterance and compressed with an Xvid mpeg-4 codec for compatibility with the experiment delivery program (E-prime; Psychology Software Tools).

In order to manipulate speech clarity, auditory clips were noise-vocoded (NV; Shannon et al., 1995) using a custom MATLAB (Mathworks) program. Noise-vocoding reduces the spectral clarity of the speech while preserving the temporal information in the speech envelope. I created NV stimuli by filtering each audio recording into six contiguous frequency bands (logarithmically defined so as to be approximately equally spaced along the basilar membrane; Greenwood, 1990). Filtering was performed using finite impulse response Hann bandpass filters, with a window length of 801 samples. The amplitude envelope from each frequency band was extracted by full-wave rectifying the band-limited signal and applying a low-pass filter (30 Hz cutoff, using a fourth-order Butterworth filter). Each envelope was then applied to bandpass-filtered noise from the same frequency range, and all bands were subsequently recombined to produce the final 6-band NV utterance. After processing, clear and NV auditory speech were normalized to have the same average root mean square (RMS) power.
To create the audiovisual stimuli, the clear and NV auditory stimuli were each recombined with a visual stimulus. For the matching visual condition, auditory utterances were combined with the corresponding video file. The visual stimuli for the mismatching audiovisual condition were taken from sentences of similar length assigned to the auditory-only condition and were therefore never seen by participants. Visual speech that is presented without corresponding auditory information is generally difficult for normal hearing adults to comprehend, although there is a high degree of individual variability (e.g., Summerfield, 1992; Bernstein et al., 2000). Thus, I expected that matching sentences would be more intelligible than mismatching stimuli. Mismatching and matching videos were created in Virtual Dub (a free downloadable video processing utility), with mismatching videos truncated to the length of the audio stimulus. Auditory-only stimuli were created by combining the auditory signal with a blank video for compatibility with the stimulus delivery program. Since sentence sets were counterbalanced conditions across subjects, I created 216 videos for each condition. I created baseline “scrambled” stimuli containing unintelligible auditory and visual speech information using a custom program in which the visual speech signal was scrambled into 64 x 64 pixels that were re-scrambled five times per second for the duration of the stimulus. This scrambled video was combined with signal-correlated noise (Davis and Johnsrude, 2003), such that the noise signal was modulated by the amplitude envelope of the original speech signal using Praat software (Boersma, 2002) in order to create the single final scramble stimulus.

4.2.3.2 Auditory Distractors

On half the trials, participants attended to an auditory distractor sequence, presented concurrently with a sentence stimulus, and were asked to identify a target sound that differed
from other sounds in the sequence. Auditory distractors were similar to those used by Wild et al. (2012); I created two types of sound: a target and a standard. Standards were sequences of 400 ms narrow-band noise bursts separated by a variable amount of silence (220–380 ms). The number of sounds in each sequence was selected so that the duration of the auditory distractor sequence would be approximately equal to that of the sentence with which it was paired. Each noise burst was created by passing 400 ms of broadband white noise through a filter with a fixed bandwidth of 1000 Hz and a center frequency randomly selected to be between 1000-2000 Hz. I created linear onsets of 380 ms and sharp linear offsets of 20 ms through amplitude modulation (‘approach noises’), whereas target sounds were characterized by a sharp onset and linear offset of equal duration (‘depart noises’). In half of the experimental trials the distractor sequence contained a single target sound. This was never the first sound in the sequence, and was located in the second half of the sequence 80% of the time to discourage participants from switching their attention to the speech utterance partway through the task.

4.2.4 Pilot Procedure

The procedure in the pilot was identical to that described below, except that participants typed out all of the words that they could report from degraded utterances, instead of the task used in the fMRI sessions of indicating with a keypress whether they understood the gist or not. These 24 subjects were tested with 6-band vocoded speech to ensure that speech materials were presented at a suitable level of intelligibility to avoid ceiling or floor effects. An additional 6 subjects not reported here were tested on 4-band noise-vocoded speech. See Appendix B.
4.2.5 Procedure

I presented distractor sequences and speech stimuli binaurally and simultaneously over MR-compatible high-fidelity electrostatic earphones, placed in ear defenders that attenuated the background sound of the scanner by 30 dB (Nordic Neurolab Audio System). Auditory stimuli were presented at a comfortable listening level (approximately 75-80 dB SPL). Visual stimuli were presented on a rear projection screen delivered by an MR-compatible LCD projection system (Avotec SV-6011).

On each trial, participants were cued to attend to a single stimulus stream by a visual prompt presented during the image volume acquired at the end of the previous trial (cue “speech” for the speech stimulus, and “chirps” for the auditory distractor). At the end of ‘speech’ trials, participants used a yes/no button press (using a button wand; right thumb for “yes”, left thumb for “no”) to indicate whether or not they understood the gist of the sentence, which provided a measure of the intelligibility of the attended speech signal. During each ‘chirp’ (auditory distractor) trial, participants were asked to monitor the stream for a single target stimulus (sharp onset, linear offset, rather than vice versa), using the yes/no button press (right thumb for “yes”, left thumb for “no”) at the end of each trial to indicate whether or not the target was present. A response window of 2s (prompted by the word “Respond”) occurred before the onset of the image acquisition period (see Figure 4-1).
Figure 4-1. A schematic representation of an experimental trial (TR 9000 ms). Image acquisition (2000 ms) was clustered at the end each trial. Stimuli were presented during 7000 ms of silence, with the midpoint of the speech stimulus positioned at exactly the midpoint (3500 ms). Subjects’ attention was directed to either the speech or distractor task by a cue presented during the preceding scan.

Eighteen trials were administered in each of the 12 experimental conditions (2 attention tasks, 2 speech types, and 3 audiovisual types). 18 trials of the scrambled baseline condition were administered (in keeping with the number of experimental stimuli for each condition), with 10 trials of the silent baseline (248 trials total). The 248 trials were divided into four blocks of 62 trials, each with approximately the same number of trials from each condition. To minimize task switching (Wild et al., 2012), conditions were presented in pseudorandom order such that participants performed the same attentional task (speech or auditory distractor) for between two and five consecutive trials. There were never more than five consecutive trials of one attention condition, and stimulus condition (speech clarity and
AV type) was pseudorandom within each of these sequences of trials. Control trials (silence and scrambled baseline) were pseudorandomly interspersed throughout the experiment. One extra acquisition image was added to the start of each block to allow the magnetization to reach a steady state; these dummy images were discarded from all preprocessing and analysis steps. All participants (including pilot subjects) underwent extensive training on each stimulus stream separately (i.e., clear and degraded AV speech, and distractor stimuli), and then on the two streams together in a simulation of the actual experiment, using a separate set of training stimuli. As the intelligibility of NV speech improves with experience, all participants were initially trained with 20 sentences of (auditory-only) six-band NV stimuli (spoken by the same talker used in the experiment) to ensure that intelligibility of NV speech was asymptotic prior to performing the experiment (Davis et al., 2005). I trained using the same protocol introduced by Davis et al. (2005) to facilitate learning of NV speech (subjects heard a degraded sentence for word report, after which they were presented with the clear version of the utterance (feedback), followed by the degraded utterance again).

4.2.6 Behavioral Post-test

Immediately following the experimental scanning, subjects performed a surprise recognition memory post-test; half of the sentences from each condition were randomly selected for each participant, counterbalanced across participants, with an additional 54 new foil sentences intermixed with the 108 target sentences. All sentences were presented as text (i.e., visually) on a computer screen, with the participant seated at a desk in a quiet room adjacent to the MR suite. Subjects were required to make an old/new discrimination for each stimulus, responding via a keyboard press. Sensitivity (d’) was determined for each condition,
and d’ scores were analyzed using a 2 x 2 x 3 (attention x speech clarity x AV speech type) repeated-measures ANOVA.

4.2.7 fMRI Protocol and Data Acquisition

Imaging was performed on the 3.0 Tesla Siemens Trio MRI system in the Queen’s Centre for Neuroscience Studies MR Facility (Kingston, Ontario, Canada). T2*-weighted functional images were acquired using a GE-EPI sequence (field of view, 211 mm x 211 mm; 64 x 64 matrix size; in-plane resolution, 3.3 mm x 3.3 mm; slice thickness, 3.3 mm with a 25% gap; TA, 2000 ms per volume; TR, 9000 ms; TE, 30 ms; flip angle, 78°). Acquisition was transverse oblique, angled away from the eyes, and in most cases covered the whole brain. In very few cases, slice positioning excluded the top of the superior parietal lobule. Each stimulus sequence was positioned in the silent interval such that the middle of the sequence occurred 4s before the onset of the next scan (see Figure 4-1). Due to a technical error, the image acquisition period overlapped with the stimulus presentation for the next trial for a maximum of 500 ms for four subjects. As the average sentence length was 6.2 seconds (SD = 0.28), leaving an average of 400 ms of silence at the onset of each trial, the overlap with the stimulus would have been minimal and it is unlikely that this affected stimulus intelligibility, with all conditions equally affected. Indeed, removing these four participants from the analysis degraded sensitivity but did not change the main findings. In addition to functional data, I acquired a whole-brain 3D T1-weighted anatomical image (voxel resolution, 1.0 mm³) for each participant at the start of each session.

4.2.8 fMRI Data Preprocessing

fMRI data were preprocessed and analyzed using Statistical Parametric Mapping (SPM8; Wellcome Centre for Neuroimaging, London, UK). Data preprocessing steps for each
subject included: (1) rigid realignment of each EPI volume to the first of the session; (2) coregistration of the structural image to the mean EPI; (3) normalization of the structural image to common subject space (with a subsequent affine registration to MNI space) using the group-wise DARTEL registration method included with SPM8 (Ashburner, 2007); and (4) warping of all functional volumes using deformation fields generated from the normalization step, which simultaneously resampled the images to isotropic 3 mm voxels and spatially smoothed them with a three-dimensional Gaussian kernel with a full width at half-maximum of 10 mm. Application of this smoothing kernel resulted in an estimated smoothness of ~17 mm in the group analyses.

4.2.8 fMRI Analysis

Analysis of each participant’s data was conducted using a general linear model; each scan was coded as belonging to one of 14 conditions. The four runs were modeled separately within the design matrix, and four regressors were used to remove the mean signal from each of the runs. Six realignment parameters were included to account for movement-related effects (i.e., three degrees of freedom for translational movement in the x, y, and z directions, and three degrees of freedom for rotational motion: yaw, pitch, and roll). Due to the long TR of this sparse imaging paradigm, no correction for serial autocorrelation was necessary. Due to the large number of conditions in the experiment, and the resulting low-frequency experimental signals, I did not high-pass filter the data. Models were fitted to each individual’s data using a least-mean-squares method, and parameter estimates were obtained. Contrast images for each of the 12 experimental conditions were generated by comparing each of the condition parameter estimates (i.e., 12 betas) to the scrambled baseline condition.
These images were primarily used to obtain plots of estimated signal within voxels for each condition and were entered into a group-level factorial ANOVA.

The group-level analysis was conducted using a 2 (attention) x 2 (speech clarity) x 3 (audiovisual speech type) factorial pooled-error repeated-measures ANOVA, with subject effects included. Factors were dependent, with unequal variance assumed. For whole-brain analyses of the main effects and their interaction, I used a voxelwise threshold of $p < 0.05$, FWE corrected for multiple comparisons over the whole brain as implemented in SPM.

A significant main effect or interaction in an ANOVA can be driven by many possible simple effects. In my study, for example, a main effect of speech type might indicate that activity correlates with intelligibility (i.e., high activity for clear speech and low activity for degraded speech), or that activity is increased for degraded compared with clear speech. Therefore, I parsed thresholded F-statistic images showing overall main effects (and interaction) into simple effects by inclusively masking with specific t-contrast images (i.e., simple effects), thresholded at $p < 0.05$, corrected for multiple comparisons. Significant voxels revealed by F-contrasts were labeled as being driven by one or more simple effects. Peaks were localized using the LONI probabilistic brain atlas (LPBA40; Shattuck et al., 2008) and confirmed by visual inspection of the average structural image. Results of the fMRI analysis are shown on the average normalized T1-weighted structural image for the group.

4.2.9 ALE Map for Audiovisual Speech Integration

I conducted an activation likelihood-estimation (ALE) meta-analysis (GingerALE; Eickhoff et al., 2009) on published coordinates of audiovisual speech integration for use as a small-volume correction in my analyses (to increase sensitivity). Included studies examined
effects of bimodal auditory-visual activation, or supra-additivity, localizers for audiovisual integration, or moving faces vs. baseline (see Table 4-1). A search was conducted on Pubmed, with broad inclusion criteria, since the goal was to obtain a fairly broad spatial map for audiovisual speech integration. Note that stimuli are diverse across studies and include congruent speech, McGurk stimuli, and range from syllables to sentences. I did not include coordinates from studies in which visual stimuli were not present on both sides of the contrast, such as in the case of an AV minus auditory-only contrast, since I used a baseline controlling for low-level, but non-meaningful auditory and visual information in my study. Talairach coordinates were converted to MNI space using GingerALE. The ALE spatial map was FDR (false discovery rate) thresholded at $p < .05$, with a minimum cluster size of 50 mm$^3$. The minimum ALE value was .008, with a maximum value of .027. The ALE analysis yielded seven clusters that included bilateral STG and right MTG (see Figure 4-2).

![ALE Activation Map](image)

Figure 4-2. ALE activation map for fMRI studies assessing audiovisual speech integration. The ALE analysis yielded seven bilateral clusters, which included bilateral STS and right MTG.
<table>
<thead>
<tr>
<th>Study</th>
<th>Stimuli</th>
<th>Contrast</th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beuchamp, Nath, &amp; Pasalar, 2010</td>
<td>words</td>
<td>localizer: A ∩ V</td>
<td>-56</td>
<td>-27</td>
<td>8</td>
</tr>
<tr>
<td>Calvert et al., 2000</td>
<td>passage</td>
<td>$A &gt; 0 \land V &gt; 0 \land AV &gt; A, V$ and $&lt; BOLD$ for incongruent speech</td>
<td>-52</td>
<td>-53</td>
<td>-9</td>
</tr>
<tr>
<td>Erickson et al., 2014</td>
<td>McGurk</td>
<td>Validating speech (i.e., McGurk &gt; Non-McGurk, Congruent &gt; Incongruent)</td>
<td>-62</td>
<td>-26</td>
<td>4</td>
</tr>
<tr>
<td>Fairhall &amp; Macaluso, 2009</td>
<td>30s speech segments</td>
<td>Selective attention to AV &gt; baseline</td>
<td>-66</td>
<td>-39</td>
<td>9</td>
</tr>
<tr>
<td>Hall, Fussell, &amp; Summerfield, 2005</td>
<td>sentences</td>
<td>$A \land V$</td>
<td>-60</td>
<td>-48</td>
<td>0</td>
</tr>
<tr>
<td>Lee &amp; Noppeney, 2011</td>
<td>sentences</td>
<td>Interaction for Intelligible AV speech</td>
<td>-52</td>
<td>-16</td>
<td>-6</td>
</tr>
<tr>
<td>MacSweeney et al., 2002</td>
<td>sentences</td>
<td>AV &gt; baseline</td>
<td>-56</td>
<td>-27</td>
<td>4</td>
</tr>
<tr>
<td>Miller &amp; D'Esposito, 2005</td>
<td>VCV</td>
<td>AV fused &gt; AV unfused (temporal synchrony)</td>
<td>-46</td>
<td>-28</td>
<td>0</td>
</tr>
<tr>
<td>Nath and Beauchamp, 2012</td>
<td>words</td>
<td>localizer: A ∩ V</td>
<td>-54</td>
<td>-28</td>
<td>4</td>
</tr>
<tr>
<td>Nath and Beauchamp, 2011</td>
<td>words</td>
<td>AV temporally synchronous &gt; AV asynchronous</td>
<td>-58</td>
<td>-34</td>
<td>-4</td>
</tr>
<tr>
<td>Noesselt et al., 2012</td>
<td>sentences</td>
<td>matching AV &gt; still face</td>
<td>-67</td>
<td>-16</td>
<td>-6</td>
</tr>
<tr>
<td>Ojanen et al., 2005</td>
<td>vowels</td>
<td>$V + \text{Degraded } A &gt; \text{still face}$</td>
<td>-55</td>
<td>-22</td>
<td>1</td>
</tr>
<tr>
<td>Sekiyama et al., 2003</td>
<td>McGurk</td>
<td>$AV &gt; \text{still face}$</td>
<td>-52</td>
<td>-8</td>
<td>-18</td>
</tr>
<tr>
<td>Stevenson et al., 2010</td>
<td>words</td>
<td>ROI: A ∩ V</td>
<td>-55</td>
<td>-19</td>
<td>9</td>
</tr>
<tr>
<td>Szycik et al., 2008</td>
<td>disyllabic lemmas</td>
<td>localizer: $A \cap V \cap AV &gt; V \cap AV &gt; A$</td>
<td>-65</td>
<td>-35</td>
<td>15</td>
</tr>
<tr>
<td>Szycik et al., 2012</td>
<td>McGurk</td>
<td>McGurk percept &gt; non-fused McGurk</td>
<td>-61</td>
<td>-28</td>
<td>7</td>
</tr>
<tr>
<td>van Atteveldt et al., 2007</td>
<td>letters</td>
<td>(bimodal &gt; unimodal) and unimodal &gt; 0</td>
<td>-58</td>
<td>-43</td>
<td>15</td>
</tr>
<tr>
<td>van Atteveldt et al., 2004</td>
<td>letters</td>
<td>STS ROI: $AV \cap A \cap AV &gt; V, A \cap A, V &gt; \text{baseline}$</td>
<td>-57</td>
<td>-33</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 4-1. Studies and coordinates included in the ALE meta-analysis used to derive a spatial map for a small-volume correction in my study. Note that baseline included a low-level auditory and visual stimulus. Abbreviations: $A$ = auditory speech; $V$ = visual speech; $AV$ = audiovisual speech; $VCV$ = vowel consonant vowel stimuli.
4.3 Behavioural Results

4.3.1 Auditory Distractor Task

Mean sensitivity was 2.55 (range 0.83-3.67). This was significantly greater than chance levels \(t_{(23)} = 17.48, p < .001\), indicating that participants were correctly attending to the intended auditory stream, similar to my pilot study (Appendix B; mean \(d' = 2.65\)) and to Wild et al (2012; \(d' = 2.15\)).

4.3.2 Speech Task

At the end of each trial, participants indicated whether or not they understood the gist of the sentence. A repeated-measures ANOVA on the proportion of sentences understood revealed a main effect of speech clarity \((F(1,23) = 134.49, p < .001, \eta_p^2 = .85)\) and audiovisual speech type, \((F(2,46) = 37.62, p < .001, \eta_p^2 = .62; \text{Figure 4-3})\). Pairwise comparisons (Sidak-corrected) revealed a similar pattern to the pilot (word-report) intelligibility data: clear speech \((M = 1.00, SD = 0)\) was significantly more intelligible than degraded speech \((M = .75, SD = 0.2; p < .001)\) and matching audiovisual speech \((M = .94, SD = .01)\) was significantly more intelligible than both mismatching audiovisual speech \((M = .83, SD = .02)\) and auditory-only speech \((M = .84, SD = .01; p < .001 \text{ for both}; \text{Figure 4-3})\). The speech clarity x audiovisual speech type interaction was significant \((F(2,46) = 38.23, p < .001, \eta_p^2 = .62)\). Again, as in the pilot, AV speech type significantly affected the intelligibility of degraded, but not clear, speech, with degraded matching audiovisual speech being significantly more intelligible than both degraded mismatching \(t_{(23)} = 8.03, p < .001\) and auditory-only sentences \(t_{(23)} = 7.14, p < .001\).
Figure 4-3. Behavioural data for gist task performed on attended sentence stimuli in the MR system, with pilot word report for comparison. Experimental subjects indicated via button press whether or not they understood the gist of each attended sentence, whereas pilot subjects typed out the full sentence for word report. Error bars denote standard error of the mean. Abbreviations: MM = mismatching audiovisual speech; AO = auditory-only speech; M = matching audiovisual speech.

4.3.4 Post-test

Results are presented in Figure 4-4. I found a main effect of speech clarity ($F(1,23) = 9.78, p < .005, \eta_p^2 = .30$). d' scores for clear speech ($M = 1.08, SD = 0.11$) were significantly higher than for degraded speech ($M = 0.87, SD = .10; p < .05$). I also observed a main effect of attention ($F(1,23) = 34.72, p < .001, \eta_p^2 = .61$), with greater recognition scores when participants attended to speech ($M = 1.28, SD = 0.13$) compared to the auditory distractor ($M = 0.67, SD = 0.10; p < .001$). These main effects must be understood in the context of a
significant attention x speech clarity interaction ($F(1,23) = 5.61$, $p < .05$, $\eta_p^2 = .20$): d' scores were significantly higher for clear speech than for degraded speech only when attending to the distractor, and not when attending to speech ($p < .001$). There was also a main effect of audiovisual speech type ($F(2,46) = 3.73$, $p < .05$, $\eta_p^2 = .14$). Pairwise comparisons (Sidak-corrected) indicated that sentences containing matching visual information ($M = 1.07$, $SD = 0.12$) were significantly better recognized than sentences containing only auditory information ($M = 0.91$, $SD = 0.09$, $p < .05$). There was no significant difference in d’ scores between sentences containing matching visual speech and mismatching visual speech ($M = .94$, $SD = .11$), nor between mismatching and auditory-only stimuli. Two- and three-way interactions involving audiovisual speech type were not significant.

Figure 4-4. fMRI behavioural results from old/new discrimination post-test. Error bars reflect standard error of the mean.
4.4 Imaging Results

Imaging data were checked for artifacts and to verify normalization. Only those blocks showing robust bilateral auditory cortex activity for the sound vs. silence contrast were included for analysis (this criterion resulted in one block being excluded from analysis from four subjects; effects for these subjects were estimated using the remaining three blocks). One additional block from another subject was removed due to excessive head movement (>5 mm). One subject completed all stimulus trials over 5 blocks, and another completed only 11 of 62 trials in the first block (with 62 trials in each of the subsequent 3 blocks), both due to errors with the stimulus delivery program.

4.4.1 Main effect of Speech Clarity

The contrast assessing the main effect of speech type revealed extensive activity in bilateral temporal and inferior frontal regions (see Figure 4-5 and Table 4-2). Activity was greater for clear compared to degraded speech in bilateral temporal regions, whereas it was greater for degraded compared to clear speech in bilateral anterior insular regions (similar to Wild et al., 2012), although the latter contrast is better understood as an interaction with attention (i.e., greater activation when attention is directed to speech, rather than the distractor; see below).
4.4.2 Main effect of Attention

The contrast assessing the main effect of attentional condition revealed widespread activity. When attention was directed to speech I observed greater activity in a broad network of regions (see Figure 4-6, red areas, and Table 4-2) compared to attending to the distractor. Attending to the distractor resulted in greater activity in a more right-dominant network (see Figure 4-6, blue areas, and Table 4-2).
Figure 4-6. t-contrasts for the main effect of attention.

4.4.3 Main effect of AV Type

As expected, the contrast assessing the main effect of AV type revealed robust bilateral activity in visual regions and thalamus; see Figure 4-7 and Table 4-2. This main effect is due to visual regions and the LGN of the thalamus (see Figure 4-7) being more active for audiovisual stimuli (matching and mismatching) than for auditory-only stimuli. Greater activity was observed in bilateral lingual gyri and right cuneus for auditory only speech relative to audiovisual speech. Comparing matching and mismatching stimuli did not reveal any significant differences, even when using either the clear > degraded contrast or ALE mask to enhance sensitivity. I hypothesized that mismatching AV information would engage
inferior frontal regions relative to the matching AV condition, in a manner similar to the
pattern of activity seen for degraded-clear speech, since matching was significantly more
intelligible than mismatching. Thus, to enhance sensitivity I attempted to use the degraded-
clear speech $t$-contrast (thresholded at FWE, $p < .05$) as a small volume correction on the MM
$>$ M contrast; however, this was not significant.

Figure 4-7. $t$-contrasts for the main effect of AV speech type. M = matching AV speech. Mm
= mismatching AV speech; AO = auditory-only speech.
<table>
<thead>
<tr>
<th>Main Effect</th>
<th>Contrast</th>
<th>Cluster size</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>T</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Clarity</td>
<td>Clear &gt; Degraded</td>
<td>687</td>
<td>-33</td>
<td>30</td>
<td>12</td>
<td>7.6</td>
<td>L Heschl’s gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-54</td>
<td>-12</td>
<td>3</td>
<td>7.58</td>
<td>STG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-54</td>
<td>-21</td>
<td>3</td>
<td>6.23</td>
<td>MTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>708</td>
<td>51</td>
<td>-66</td>
<td>27</td>
<td>6.45</td>
<td>MTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>-51</td>
<td>9</td>
<td>5.87</td>
<td>MTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>57</td>
<td>-54</td>
<td>3</td>
<td>5.83</td>
<td>STG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>348</td>
<td>60</td>
<td>3</td>
<td>-15</td>
<td>6.41</td>
<td>STG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57</td>
<td>-12</td>
<td>3</td>
<td>6.28</td>
<td>STG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
<td>-9</td>
<td>-24</td>
<td>5.04</td>
<td>MTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>68</td>
<td>-60</td>
<td>-57</td>
<td>-3</td>
<td>5.58</td>
<td>MTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>30</td>
<td>-6</td>
<td>-6</td>
<td>4.62</td>
<td>Putamen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>60</td>
<td>-21</td>
<td>-21</td>
<td>4.55</td>
<td>ITG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-66</td>
<td>39</td>
<td>4.53</td>
<td>L Supramarginal Gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>24</td>
<td>45</td>
<td>4.48</td>
<td>R Middle Frontal Gyrus</td>
<td></td>
</tr>
<tr>
<td>Degraded &gt; Clear</td>
<td></td>
<td>189</td>
<td>-33</td>
<td>24</td>
<td>-3</td>
<td>8.04</td>
<td>L anterior insula</td>
</tr>
<tr>
<td></td>
<td></td>
<td>324</td>
<td>-3</td>
<td>18</td>
<td>51</td>
<td>7.92</td>
<td>L SFG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>9</td>
<td>24</td>
<td>39</td>
<td>5.86</td>
<td>R SFG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>137</td>
<td>36</td>
<td>24</td>
<td>-3</td>
<td>6.92</td>
<td>R anterior insula</td>
</tr>
<tr>
<td>Attention</td>
<td>Speech &gt; Distractor</td>
<td>4639</td>
<td>-54</td>
<td>24</td>
<td>9</td>
<td>11.03</td>
<td>L IFG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-42</td>
<td>-66</td>
<td>24</td>
<td>10.84</td>
<td>MTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-54</td>
<td>-48</td>
<td>0</td>
<td>10.16</td>
<td>MTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>471</td>
<td>60</td>
<td>0</td>
<td>-12</td>
<td>9.02</td>
<td>R Medial Temporal Pole</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td>15</td>
<td>21</td>
<td>-19</td>
<td>R Medial Temporal Pole</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>39</td>
<td>36</td>
<td>-12</td>
<td>5.27</td>
<td>IFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>754</td>
<td>42</td>
<td>48</td>
<td>-18</td>
<td>7.5</td>
<td>R Fusiform Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
<td>-69</td>
<td>-3</td>
<td>6.85</td>
<td>ITG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36</td>
<td>-39</td>
<td>-21</td>
<td>6.84</td>
<td>Fusiform Gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>211</td>
<td>18</td>
<td>-81</td>
<td>-36</td>
<td>7.38</td>
<td>R Cerebellum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>93</td>
<td>15</td>
<td>51</td>
<td>39</td>
<td>6.62</td>
<td>SFG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>-3</td>
<td>42</td>
<td>-18</td>
<td>6.38</td>
<td>L Rectal Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>172</td>
<td>-6</td>
<td>-54</td>
<td>12</td>
<td>6.28</td>
<td>L Calcarine Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57</td>
<td>48</td>
<td>0</td>
<td>5.78</td>
<td>SFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>-6</td>
<td>39</td>
<td>5.13</td>
<td>L MCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>57</td>
<td>36</td>
<td>9</td>
<td>4.99</td>
<td>IFG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>-36</td>
<td>51</td>
<td>4.72</td>
<td>L Precentral Gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distra...</td>
<td>296</td>
<td>36</td>
<td>-39</td>
<td>39</td>
<td>7.99</td>
<td>IPS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>227</td>
<td>57</td>
<td>9</td>
<td>15</td>
<td>7.92</td>
<td>IFG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>133</td>
<td>54</td>
<td>9</td>
<td>42</td>
<td>5.4</td>
<td>R Precentral Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>218</td>
<td>54</td>
<td>30</td>
<td>-2</td>
<td>6.77</td>
<td>R Precentral Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>-18</td>
<td>21</td>
<td>6</td>
<td>R Rolandic Operculum</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>-30</td>
<td>18</td>
<td>6.38</td>
<td>STG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>26</td>
<td>-63</td>
<td>21</td>
<td>5.43</td>
<td>L Precentral Gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23</td>
<td>9</td>
<td>66</td>
<td>5.29</td>
<td>R Posterior-Medial Frontal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>15</td>
<td>63</td>
<td>5</td>
<td>4.95</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>60</td>
<td>21</td>
<td>4.78</td>
<td>IFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>-39</td>
<td>33</td>
<td>4.49</td>
<td>SFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Matching &gt; Auditory Only</td>
<td>3284</td>
<td>24</td>
<td>-96</td>
<td>6</td>
<td>28.3</td>
<td>R Calcarine Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>-15</td>
<td>3</td>
<td>25.05</td>
<td>L Lingual gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>84</td>
<td>-21</td>
<td>-90</td>
<td>-9</td>
<td>23.77</td>
<td>L Lingual gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>63</td>
<td>24</td>
<td>-27</td>
<td>0</td>
<td>10.75</td>
<td>Thalamus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>51</td>
<td>-36</td>
<td>9</td>
<td>4.83</td>
<td>R STG</td>
</tr>
<tr>
<td></td>
<td>Mismatching &gt; Matching</td>
<td>1</td>
<td>39</td>
<td>27</td>
<td>4.44</td>
<td>R MFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AV Speech Type</td>
<td>3160</td>
<td>24</td>
<td>-96</td>
<td>6</td>
<td>28.2</td>
<td>R Calcarine Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td>-15</td>
<td>-96</td>
<td>3</td>
<td>25.74</td>
<td>L Lingual gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45</td>
<td>24</td>
<td>-27</td>
<td>0</td>
<td>9.12</td>
<td>Thalamus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>63</td>
<td>51</td>
<td>-39</td>
<td>12</td>
<td>6.49</td>
<td>STG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>-51</td>
<td>-48</td>
<td>12</td>
<td>5</td>
<td>L MTG</td>
</tr>
<tr>
<td></td>
<td>Mismatching &gt; Auditory Only</td>
<td>3387</td>
<td>12</td>
<td>-66</td>
<td>-3</td>
<td>12.34</td>
<td>R Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-12</td>
<td>-66</td>
<td>-3</td>
<td>12.15</td>
<td>L Lingual gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>-78</td>
<td>27</td>
<td>11.86</td>
<td>Cuneus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auditory Only &gt; Matching</td>
<td>15</td>
<td>-66</td>
<td>3</td>
<td>12.94</td>
<td>R Lingual Gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-15</td>
<td>-66</td>
<td>-3</td>
<td>12.46</td>
<td>L Lingual gyrus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auditory Only &gt; Mismatching</td>
<td>6</td>
<td>-18</td>
<td>51</td>
<td>4.87</td>
<td>L SPL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>-66</td>
<td>-3</td>
<td>12.09</td>
<td>R Cuneus</td>
<td></td>
</tr>
</tbody>
</table>
Table 4-2. Peak voxels for main effects, corrected at FWE < .05. Abbreviations: AV = audiovisual; L = left; R = right; STG = superior temporal gyrus; MTG = middle temporal gyrus; ITG = inferior temporal gyrus; SFG = superior frontal gyrus; MFG = middle frontal gyrus; IFG = inferior frontal gyrus; SPL = superior parietal lobule; IPS = inferior parietal sulcus; MCC = midcingulate cortex

4.4.4 Attention x Speech Clarity

The attention x speech clarity interaction reveals areas in which sensitivity to speech depends both on acoustic quality and on attentional state. I predicted that the difference in activation in speech-sensitive cortex for clear speech relative to degraded speech would be significantly greater under distraction, compared to attention (Wild et al., 2012). I also predicted that degraded speech would produce stronger activation in inferior frontal regions compared to clear speech under full attention and that this difference would be smaller when listeners attended to the distractors (i.e., an attentionally gated effort-related response, as also seen in Wild et al., 2012).

The contrast assessing the speech clarity x attention interaction revealed significant activity in bilateral angular gyri, bilateral supramarginal gyri, right MTG, bilateral precuneus, bilateral middle frontal gyri, right insular cortex, left IFG, and right middle occipital gyrus (see Figure 4-8 and Table 4-3). In bilateral supramarginal gyri, right middle temporal gyrus, bilateral angular gyri, bilateral precuneus, bilateral middle frontal gyri, right superior frontal gyrus, left inferior temporal gyrus, left middle temporal gyrus, and cerebellum, this interaction appeared to be driven by a greater activity difference between clear compared to degraded speech when attending to speech compared to when the distractor was attended. A 2x2 repeated-measures ANOVA on the contrast values for all peak voxel clusters confirmed that the difference between clear and degraded speech was significantly larger under full attention,
compared to under distraction, in each of these regions. However, these contrast values are largely negative, indicating decreased activity in these regions for these conditions relative to the unintelligible scrambled baseline, limiting interpretation (e.g., see Figure 4-9; in the right MFG, I observed a cross-over interaction such that the degraded-versus-clear difference was greater under distraction than under full attention). As shown in Figure 4-9, this contrast accounted for a large portion of the interaction in speech-sensitive cortices.

In addition, as predicted, the difference in activity for degraded speech compared to clear speech was greater under full attention than distraction, in bilateral anterior insular cortex; (see Figure 4-8). This suggests that inferior frontal regions are recruited in an effort-related response to enhance speech intelligibility, replicating Wild et al., (2012).
Figure 4-8. Attention x speech clarity interaction. Attn = attention to speech; Dist = attention to distractor; C = clear speech; D = degraded speech.

Figure 4-9. Contrast values for the attention x speech clarity interaction in the left supramarginal gyrus. The interaction for BOLD activity in this region was characterized by greater activity for clear speech than for degraded speech under full attention, but not distraction. Error bars reflect mean standard error.
4.4.5 Speech Clarity x AV type

The speech clarity x audiovisual speech type interaction was not significant, even with use of small-volume correction (i.e., degraded > clear and clear > degraded as masks).

4.4.6 Attention x AV Speech Type

Areas in which attentional state modulates activity evoked by the three different types of AV stimuli (irrespective of speech clarity) included right middle occipital gyrus, right fusiform gyrus, bilateral lingual gyri, right cuneus, left superior occipital gyrus, right middle frontal gyrus, and right inferior frontal gyrus (See Figure 4-10 and Table 4-3). There are a number of ways in which AV speech type can interact with attention. I hypothesized that this interaction might be driven by full attention enhancing the activity difference between AV speech (matching plus mismatching) and auditory-only speech stimuli, compared to distraction, and this appears to be the case for activity in the right fusiform and middle occipital right gyri (See Figures 4-10 and 4-11). This was confirmed through a 2x3 repeated-measures ANOVA at peak voxels in the right middle occipital gyrus and right fusiform gyrus, although note that all contrast values in the former contrast are negative (Figure 4-11).

Increased activity in the right middle frontal gyrus and right anterior insula were seen only for the contrast assessing greater activity for mismatching speech relative to auditory-only speech, under full attention, compared to distraction (Figures 4-10 and 4-11; note that using the degraded > clear contrast as a mask for small-volume correction did not significantly enhance sensitivity to attentionally gated activity for AV speech relative to AO speech in this region). Bilateral lingual gyrus activation reflected greater activity for auditory-only speech relative to matching and mismatching speech under full attention, compared to distraction (Figures 4-10 and 4-11).
Figure 4-10. Attention x AV speech type interaction. Attn = attention to speech; Dist = attention to distractor; AV = audiovisual speech (matching and mismatching audiovisual speech conditions); AO = auditory-only speech.
4.4.7 AV Type x Speech Type x Attention Interaction

This interaction was not significant at the whole brain level, with the exception of a small cluster in the right medial orbitofrontal cortex (Table 4-3). A 2x2x3 repeated-measures ANOVA on the contrast values at the peak of this cluster revealed that for degraded speech, but not clear speech, activity was significantly greater for matching speech compared to both
mismatching and auditory-only speech, but only under full attention. Interestingly, the same pattern was also observed for mismatching minus auditory-only speech, for degraded but not clear speech, under full attention, but not distraction. The pattern shown here (see Figure 4-12) implies that a similar, but attenuated, pattern was seen for degraded mismatching speech as for degraded matching speech, relative to auditory-only speech in this region. I used the degraded > clear, clear > degraded contrasts as masks (thresholded at FWE, \( p < .05 \)), as well as the ALE map, for small volume corrections, but even with a smaller correction for multiple comparisons, this contrast did not reach significance.

Figure 4-12. Contrast values for the 3-way interaction in the right middle orbitofrontal gyrus.
Table 4-3. Peak voxels for interactions, significant at $p < .05$ at a whole-brain corrected level using FWE correction. The speech clarity x audiovisual speech type interaction was not significant at a whole-brain level. Abbreviations: AV = audiovisual; L = left; R = right; MTG = middle temporal gyrus; SFG = superior frontal gyrus; IFG = inferior frontal gyrus; AG = angular gyrus; SMG = supramarginal gyrus; SOG = superior occipital gyrus; MOG = middle occipital gyrus; MOFG = middle orbitofrontal gyrus.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Cluster Size</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention x Speech Clarity</strong></td>
<td>510</td>
<td>-45</td>
<td>-69</td>
<td>33</td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>-45</td>
<td>27</td>
<td>24</td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>-54</td>
<td>24</td>
<td></td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>246</td>
<td>-60</td>
<td>-36</td>
<td>48</td>
<td>L SMG</td>
</tr>
<tr>
<td></td>
<td>-60</td>
<td>-33</td>
<td>36</td>
<td></td>
<td>L SMG</td>
</tr>
<tr>
<td></td>
<td>113</td>
<td>54</td>
<td>-51</td>
<td>-6</td>
<td>R MTG</td>
</tr>
<tr>
<td></td>
<td>164</td>
<td>6</td>
<td>-57</td>
<td>48</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td>-9</td>
<td>-51</td>
<td>54</td>
<td></td>
<td>L Precuneus</td>
</tr>
<tr>
<td></td>
<td>-6</td>
<td>-60</td>
<td>51</td>
<td></td>
<td>L Precuneus</td>
</tr>
<tr>
<td></td>
<td>93</td>
<td>30</td>
<td>21</td>
<td>54</td>
<td>R MFG</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>27</td>
<td>45</td>
<td></td>
<td>R MFG</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>-36</td>
<td>-66</td>
<td>45</td>
<td>L AG</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>36</td>
<td>24</td>
<td>-3</td>
<td>R Anterior Insula</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-39</td>
<td>24</td>
<td>-3</td>
<td>L Anterior Insula</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-27</td>
<td>15</td>
<td>48</td>
<td>L MFG</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>54</td>
<td>-30</td>
<td>33</td>
<td>R SMG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>51</td>
<td>-66</td>
<td>-9</td>
<td>R MOG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-39</td>
<td>-51</td>
<td>36</td>
<td>L AG</td>
</tr>
<tr>
<td><strong>Attention x AV Speech Type</strong></td>
<td>182</td>
<td>48</td>
<td>-66</td>
<td>3</td>
<td>R MOG</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>42</td>
<td>-51</td>
<td>-18</td>
<td>R Fusiform Gyrus</td>
</tr>
<tr>
<td></td>
<td>298</td>
<td>-9</td>
<td>-75</td>
<td>9</td>
<td>L Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>-21</td>
<td>-51</td>
<td>-3</td>
<td></td>
<td>L Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>116</td>
<td>18</td>
<td>-57</td>
<td>0</td>
<td>R Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>-69</td>
<td>9</td>
<td></td>
<td>R Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-27</td>
<td>-87</td>
<td>24</td>
<td>L SOG</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>45</td>
<td>15</td>
<td>30</td>
<td>R MFG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>45</td>
<td>36</td>
<td>-3</td>
<td>R Anterior Insula</td>
</tr>
<tr>
<td><strong>Attention x Speech Clarity x AV Speech Type</strong></td>
<td>1</td>
<td>15</td>
<td>42</td>
<td>-18</td>
<td>R MOFG</td>
</tr>
</tbody>
</table>
4.4.8 Does AV Speech Integration Require Attention? Additional Analyses

Given limited statistical power to detect a three-way interaction in speech-sensitive cortices (STS/STG, MTG), I conducted two additional follow-up analyses in order to examine whether attention modulates audiovisual speech integration, more when the acoustic input is degraded compared to when it is clear.

4.4.8.1 Three-way Interaction: Simple Effects Predictions and Results

It is possible that the benefit of AV speech integration for matching speech relative to mismatching or auditory-only speech may reflect differences in intelligibility. Thus, when speech is degraded, sentences containing matching visual speech information are more intelligible and should therefore elicit enhanced activity in speech-sensitive cortex compared to less-intelligible mismatching and auditory-only sentences. When speech is clear, it is perfectly intelligible even when it is auditory-only, and so no differences in speech-sensitive cortex should be observed for clear speech between matching, mismatching, and auditory-only AV speech. If attention is required for AV integration when speech is degraded, then the enhanced intelligibility (and activity) observed for matching degraded speech should be uniquely observed when attention is directed at speech, and not when attention is directed to the distractor. This pattern should manifest as a three-way interaction: enhanced activity in speech-sensitive temporal regions for degraded matching speech compared to mismatching and auditory-only speech, whereas for clear speech no effect of AV type should be evident, and this interaction of AV type by speech clarity should be more evident when attending to speech than when attending to the distractor. This was not significant, even with use of degraded > clear contrast as a mask to in order to perform small-volume correction.
4.4.8.2 In areas where the effect of speech clarity depends on attention, does the effect of AV speech type also depend on attention?

Another way to ask whether audiovisual speech processing relies on attention is to examine the overlap between the attention x speech clarity and attention x audiovisual speech type interactions. That is, are there any regions in which attentional state modulates the effects of both speech clarity and audiovisual speech type? To investigate this, I used the attention x speech clarity interaction (thresholded at FWE .05; Wild et al., 2012) as a small-volume correction for the attention x AV type F-contrast. I observed significant activity in the right MOG and right FFA (see Figures 4-13 and 4-14).

In the FFA, the attention x speech clarity interaction took the form of greater activity for clear speech than for degraded speech under full attention but not under distraction. The attention x AV speech type interaction in this region arose because matching and mismatching speech both elicited greater activity under full attention than under distraction, but AO speech did not. However, note that these differences remained significant under distraction. This suggests an attentionally gated enhancement of bimodal auditory visual speech information. However, this area does not seem to be sensitive to the intelligibility of the speech (since there are no significant differences between matching and mismatching).

In the right MOG, the pattern of activity was largely characterized by differences between AV and auditory-only speech types. There was greater activity for degraded speech relative to clear speech under distraction, but not full attention (attention x speech clarity). Moreover, I observed greater activity under full attention than distraction for both matching and mismatching speech, but not auditory-only speech (attention x AV speech type).
Figure 4-13. Thresholded map (FWE \( p < .05 \)) for Attention x AV Speech Type Interaction Masked by Attention x Speech Clarity Interaction (FWE = .05 corrected).
Figure 4-14. Contrast values for masked contrasts between two 2-way interactions. Error bars reflect standard error of the mean.

4.5 ICA Analysis

The results of my SPM whole-brain analysis, and from the use of the ALE map as an ROI, revealed regions in which the effects of both AV speech type and speech clarity on activity depended somewhat on attentional state, but the results were generally rather subtle. In order to further explore my data, I conducted an independent-components analysis (ICA) using the GIFT toolbox in SPM. ICA is a data-driven approach (i.e., is not constrained by a fixed model-based hypothesis) that separates data into statistically independent components (Xu et al., 2013). The goal of ICA with my data was to provide further noise reduction by removing components that reflect effects that are unrelated to the experimental design.
I selected a total of 20 components, identified as the optimal number of components by GIFT, using MDL criteria (minimum description length; Li, Adali & Calhoun, 2007), with 50 iterations. Group ICA was conducted using temporal concatenation, which concatenates all individual data along the temporal dimension before ICA decomposition; this allows for unique time courses for each subject and assumes common spatial maps (Calhoun et al., 2009). I included 16 of 24 participants due to the fact that ICA could only be completed for participants who received the same number of trials within the same number of blocks (see Methods section).

I removed three components with the lowest reliability values (i.e., as measured through the ICASSO stability parameter) based on scree plot interpretation. I then identified an additional three components that appeared to largely overlap with ventricles and sinuses. I examined (spatial) correlations between the spatial maps of these components with cerebrospinal fluid (csf) using SPM templates; these three components correlated with csf (all $r > .3$), relative to grey matter and were subsequently removed. Thus, I retained a total of 13 components for further analysis.

To examine the degree to which the model (i.e., the SPM design matrix for each subject) fit each ICA component time course, concatenated ICA time courses were correlated with model time courses (i.e., each column of the design matrix) with separate ICA regressors for subjects and sessions. The slopes of the regressors were then averaged within-subjects for analysis using a 2x2x3 (attention x speech clarity x AV speech type) repeated-measures ANOVA, conducted for each component. I characterized components by the effects (i.e., main effects and interactions) in the design matrix that they appeared to reflect. I initially searched for components that appeared to reflect 3-way interactions; however, none were
significant following correction for multiple comparisons. Below, I summarize activity observed in component spatial maps, qualitatively characterizing their activity in terms of significant main effects and interactions from ANOVA analyses (significant at $p < .001$, corrected for multiple comparisons). Note that four components did not survive this threshold; the results of the remaining 9 components are reported below (see Figure 4-15 for spatial maps thresholded at FWE .05 for comparisons across the whole brain, as well as Figure 4-16).
Figure 4-15. ICA spatial maps for 9 retained components. Slice range: -72:3:108
4.5.1 Effects Related to Main Effects of Attention, Speech Clarity, and AV Speech Type:

Component 11 reflected greater activity for full attention to speech compared to distraction. This component was significantly expressed in bilateral superior frontal gyri, right middle frontal gyrus, right cingulate gyrus, left precuneus, bilateral MTG, bilateral angular gyri, left insular cortex, right middle and inferior occipital gyrus, right precentral and postcentral gyrus, cerebellum, bilateral parahippocampal gyri, and left hippocampus. Note that the sign of the regression value is essentially arbitrary, since the polarity of the extracted component is indeterminate.

Component 15 reflects significantly greater activity under distraction, compared to full attention, as well as greater activity for degraded speech relative to clear speech. C15 explained significant variance in the timeseries of voxels in right caudate, left putamen, right MFG, bilateral precentral and postcentral gyri, bilateral inferior frontal gyrus, right supramarginal gyrus, bilateral SFG, left cuneus, right insula, right MTG, left hippocampus, left middle orbitofrontal gyrus, brainstem, and right posterior STG, possibility reflecting sensorimotor activation (Smith et al., 2009).

Component 2 was characterized by a main effect of AV speech type: specifically, stronger activation for stimuli containing AV speech information (M and MM) relative to stimuli containing only auditory information. Unsurprisingly, this was observed in occipital cortex, extending to middle and inferior occipital gyri, right fusiform gyrus, and thalamus.
C7 reflected main effects of auditory-only speech relative to both mismatching and matching speech, as well as a main effect of attention, with greater activation for attention to speech relative to distraction. This was observed in left superior, middle, and inferior frontal gyri, left supramarginal gyrus, left superior parietal gyrus, left angular gyrus, cerebellum, left fusiform gyrus, left inferior temporal gyrus, left middle temporal gyrus, right angular gyrus, left cuneus, left precentral and postcentral gyri, and left precuneus.

4.5.2 Effects Related to Attention x Speech Clarity Interaction:

Both components 8 and 18 reflected greater activity for clear speech relative to degraded speech under full attention, but not distraction. Component 8 accounted for significant variance in voxels of precentral gyrus, left superior parietal gyrus, bilateral postcentral gyri, left supramarginal gyrus, and left angular gyrus, whereas component 18 was expressed in left precuneus, bilateral middle frontal gyri, left superior frontal gyri, and cerebellum. This component was anticorrelated with timeseries in bilateral superior frontal gyri, bilateral IFG, right lingual gyri, and right lateral orbitofrontal gyrus.

4.5.3 Effects Related to Attention x AV Type Interaction:

C3 reflected an interaction of AV speech type with attention; activity for auditory-only speech was greater under full attention than distraction, whereas this difference was not significant for matching and mismatching speech. The spatial map for C3 included a large cluster with peaks in bilateral lingual gyri and bilateral superior occipital gyri.

Component 12 was characterized by greater activity for matching and mismatching AV speech relative to auditory-only speech, for attention to speech compared to distraction. In addition, there was evidence for a cross-over interaction between attention and speech clarity such that the difference between degraded minus clear speech was greater under full attention.
compared to distraction, and the difference between clear and degraded speech was greater under distraction compared to full attention. Component 12 captured activity along the length of the left STS and MTG, which overlapped considerably with the ALE map (see Figure 4-17). The component also captured activation in right posterior and middle STG, left superior and middle frontal gyrus, left lateral orbitofrontal gyrus, bilateral middle occipital gyri, right lingual gyrus, with anti-correlated activity in bilateral insular cortex, right precentral gyrus, bilateral supramarginal gyri, bilateral lingual gyri, left postcentral gyrus, bilateral superior parietal gyri (see Table 4-4). This pattern of results suggests that AV speech integration (at least, attempted integration) is localized to speech-sensitive regions, and that this is greater under full attention compared to distraction.
Figure 4-17. Derived ALE Map for AV integration overlayed on the spatial map for Component 12. T = +1 denotes regions of correlated activity, whereas T = -1 denotes regions of anti-correlated activity.

4.5.4 Effects Related to Listening Effort:

Component 13 captured greater activity for degraded speech relative to clear speech, greater under full attention than under distraction. There was an attention x AV speech type interaction such that when attending to speech, activation was greater for mismatching speech compared to both matching and auditory only speech, but this difference was not significant under distraction (and this appears to be driven by a relatively high degree of activation in this region for clear mismatching speech under full attention, compared to the other AV speech type conditions). The component was observed in a large expanse of frontal cortex that included bilateral superior and middle frontal gyri, bilateral insular cortex, bilateral cingulate
gyrus, right caudate, and left pre- and post-central gyrus (see Table 4-4), the activated regions indicate involvement of the cingulo-opercular (frontal regions, insula, and cingulate gyrus) as well as motor cortex (pre- and post-central gyri). Cingulo-opercular activity likely reflects recruitment of an adaptive control network supporting speech-based task performance (Vaden et al., 2013), as well as more generalized executive control networks (e.g., Cieslik et al., 2015; Dosenbach et al., 2006; Smith et al., 2009), whereas recruitment of motor cortex is thought to reflect recruitment of abstracted speech motor representations supporting speech identification (Evans & Davis, 2015; Hervais-Adelman et al., 2012). Both networks are especially recruited under difficult listening conditions, such as when speech is degraded (Vaden et al., 2013; Evans & Davis, 2015; Hervais-Adelman et al., 2012). Such activity likely reflects effortful listening processes, including (unsuccessful) attempts to integrate (mismatching) visual and auditory information.
<table>
<thead>
<tr>
<th>Component</th>
<th>Cluster Size</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>12</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>L STS</td>
</tr>
<tr>
<td></td>
<td>296</td>
<td>-51</td>
<td>-42</td>
<td>6</td>
<td>L STS</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>-42</td>
<td>21</td>
<td>21</td>
<td>L IFG</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>-45</td>
<td>24</td>
<td>-9</td>
<td>L LOFG</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>33</td>
<td>-90</td>
<td>0</td>
<td>R MOG</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>48</td>
<td>-33</td>
<td>3</td>
<td>R STG</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>-42</td>
<td>-87</td>
<td>-6</td>
<td>L MOG</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-48</td>
<td>18</td>
<td>-24</td>
<td>L STG</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>18</td>
<td>-69</td>
<td>15</td>
<td>R Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>60</td>
<td>-9</td>
<td>-6</td>
<td>R STG</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-6</td>
<td>15</td>
<td>63</td>
<td>L SFG</td>
</tr>
<tr>
<td><strong>12</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R Insular Cortex</td>
</tr>
<tr>
<td></td>
<td>219</td>
<td>39</td>
<td>-12</td>
<td>-3</td>
<td>R Insular Cortex</td>
</tr>
<tr>
<td></td>
<td>188</td>
<td>60</td>
<td>-24</td>
<td>33</td>
<td>R SMG</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>-15</td>
<td>-78</td>
<td>-6</td>
<td>L Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>-54</td>
<td>-6</td>
<td>9</td>
<td>L Postcentral Gyrus</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-12</td>
<td>-48</td>
<td>57</td>
<td>L SPG</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>-36</td>
<td>-18</td>
<td>12</td>
<td>L Insular Cortex</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>-60</td>
<td>-30</td>
<td>33</td>
<td>L SMG</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9</td>
<td>-75</td>
<td>-6</td>
<td>R Lingual Gyrus</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>42</td>
<td>-42</td>
<td>51</td>
<td>R SPG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-24</td>
<td>-75</td>
<td>0</td>
<td>L Lingual Gyrus</td>
</tr>
<tr>
<td><strong>13</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>L Insular Cortex</td>
</tr>
<tr>
<td></td>
<td>619</td>
<td>-36</td>
<td>12</td>
<td>-35</td>
<td>L Insular Cortex</td>
</tr>
<tr>
<td></td>
<td>1097</td>
<td>-6</td>
<td>18</td>
<td>45</td>
<td>L SFG</td>
</tr>
<tr>
<td></td>
<td>449</td>
<td>36</td>
<td>21</td>
<td>6</td>
<td>R IFG</td>
</tr>
<tr>
<td></td>
<td>183</td>
<td>-27</td>
<td>51</td>
<td>21</td>
<td>L MFG</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>6</td>
<td>-18</td>
<td>36</td>
<td>R Cingulate Gyrus</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>30</td>
<td>54</td>
<td>18</td>
<td>R MFG</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>12</td>
<td>3</td>
<td>9</td>
<td>R Caudate</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>15</td>
<td>42</td>
<td>27</td>
<td>R SFG</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-27</td>
<td>-9</td>
<td>51</td>
<td>L MFG</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-45</td>
<td>0</td>
<td>48</td>
<td>L Precentral Gyrus</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-54</td>
<td>-24</td>
<td>0</td>
<td>L STG</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-12</td>
<td>9</td>
<td>6</td>
<td>L Caudate</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>54</td>
<td>-42</td>
<td>39</td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>51</td>
<td>-45</td>
<td>48</td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-12</td>
<td>-36</td>
<td>45</td>
<td>L Postcentral Gyrus</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-6</td>
<td>-15</td>
<td>12</td>
<td>L Postcentral Gyrus</td>
</tr>
<tr>
<td><strong>13</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LAG</td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>-36</td>
<td>-75</td>
<td>39</td>
<td>LAG</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>3</td>
<td>-66</td>
<td>27</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td>118</td>
<td>54</td>
<td>-66</td>
<td>39</td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>36</td>
<td>-51</td>
<td>18</td>
<td>R AG</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-27</td>
<td>-33</td>
<td>21</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>-48</td>
<td>-6</td>
<td>30</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>45</td>
<td>0</td>
<td>21</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>-66</td>
<td>-3</td>
<td>21</td>
<td>R Precuneus</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>63</td>
<td>-48</td>
<td>-6</td>
<td>R MTG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-18</td>
<td>-6</td>
<td>30</td>
<td>R MTG</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>60</td>
<td>-12</td>
<td>L SFG</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-63</td>
<td>-51</td>
<td>9</td>
<td>L MTG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>63</td>
<td>-9</td>
<td>45</td>
<td>R Postcentral Gyrus</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>66</td>
<td>-15</td>
<td>45</td>
<td>R Postcentral Gyrus</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-6</td>
<td>51</td>
<td>-12</td>
<td>L SFG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>66</td>
<td>-6</td>
<td>-24</td>
<td>R MTG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>36</td>
<td>-9</td>
<td>33</td>
<td>R Precentral</td>
</tr>
</tbody>
</table>
Table 4-4. Peak voxels for Components 12 and 13 correlated and anti-correlated activity, significant at p < .05 at a whole-brain corrected level using FWE correction. Abbreviations: L = left; R = right; MTG = middle temporal gyrus; STS = superior temporal gyrus; STG = superior temporal sulcus; SFG = superior frontal gyrus; AG = angular gyrus; SMG = supramarginal gyrus;; MOG = middle occipital gyrus; MOFG = middle orbitofrontal gyrus; SPG = superior parietal gyrus.

4.6 Discussion

This study investigated the role of attention in AV speech integration. Participants were instructed to attend either to speech or an auditory target detection (distractor) task. I manipulated both the availability and relevance of visual speech information (i.e., contrasting matching AV speech to auditory-only and mismatching AV speech, respectively). I also used two levels of speech clarity (i.e., clear vs. degraded) in order to examine how attention in AV speech integration is modulated by the intelligibility of the speech signal.

As expected, activity related to attention to speech was observed in a wide expanse of bilateral temporal regions. Alternatively, attention to the distractor activated a right dominant parietal network, indicating that participants were engaging with the auditory discrimination task, which was also confirmed by satisfactory behavioural performance data. Increased activity for audiovisual speech relative to auditory-only speech was observed in bilateral visual cortical areas, as well as the thalamus (LGN). I observed increased activity for clear speech relative to degraded speech in bilateral temporal regions, and this pattern of activity was more evident under full attention compared to distraction. Compared to clear speech, degraded speech activated bilateral insular cortex, and this effect was also greater under full attention compared to distraction, consistent with Wild et al.’s (2012) finding that processing
of degraded speech is supported by recruitment of higher-level processes, but only under full attention.

In addition, greater activity for auditory-only speech versus audiovisual speech was observed in bilateral lingual gyri, but only under full attention. This is contrary to predictions, as lingual gyrus is typically considered a unimodal region within visual cortex (e.g., Driver & Spence, 2000). However, activity in lingual gyri has been previously been seen in contrasts of still vs. moving faces (Calvert & Campbell, 2000), and in unimodal auditory conditions, in contrasts of low-ambiguity > noise (Rodd et al., 2005) as well as higher SNR > lower SNR (Stevenson & James, 2008). This suggests that lingual gyrus may play a greater role in auditory speech processing than commonly assumed (e.g., Zekveld et al., 2002).

4.6.1 Attentional Modulation of BOLD Activity in FFA

My results demonstrate that the BOLD signal in the right FFA for AV speech depends on attention. Although it has been previously shown that the FFA is sensitive to selective attention (Wojciulik et al., 1998; Downing et al., 2001; Reddy et al., 2007), this study is the first, to my knowledge, to demonstrate an attentional effect in this area for AV speech, while also adding to the body of literature documenting the role of the FFA in visual speech processing (Calvert & Campbell, 2003; Campbell et al., 2001; Kawase et al., 2005). Interestingly, this difference between AV and auditory-only speech, although attenuated, persisted under distraction, indicating that integration of AV speech stimuli is present, to some extent, even when attention is focused on another task. Given the biological relevance of speech stimuli, it is possible that remaining attentional resources not used to complete the distractor task may be allocated to speech processing. Future work could investigate whether increasing distractor task demands, either by introducing a greater number of distractors (e.g.,
Wild et al., 2012), or by increasing task difficulty, may produce a total absence of activity for AV speech stimuli in the FFA under distraction.

The FFA did not appear to be sensitive to speech clarity in my study, since I did not observe significant differences between matching and mismatching speech in this region. However, it is interesting that visual inspection of the ICA data reveals that activity in the FFA appears to be greater for clear mismatching speech relative to degraded mismatching speech under full attention, but not distraction; although this interaction was not significant (note that I had only 18 trials per condition, and therefore limited power), it suggests that activity in the FFA may be modulated by the clarity of the auditory speech signal (see Kawase et al., 2005 and Reddy et al., 2007), which would be interesting to explore in future studies.

4.6.2 The Role of Attention in AV Speech Integration

Behavioural data indicated that degraded matching speech was significantly more intelligible than either mismatching or auditory only speech. Matching AV speech was also better recalled in the recognition post-test (i.e., a main effect of AV speech type), but I did not observe an interaction with attention. I found an interaction of AV speech type with speech clarity in the pilot data (matching speech was equally well recognized in either its clear or degraded form, but this was not the case for auditory-only or mismatching speech); however, my behavioural results are somewhat equivocal, as this result was not replicated in fMRI participants.

With respect to my imaging data, comparing audiovisual speech conditions (i.e., matching and mismatching speech) relative to auditory-only speech revealed significant differences in right fusiform gyrus, as well as right middle occipital gyrus under full attention.
but not distraction. Supplementary analyses also indicated overlap between the attention x AV speech type and attention x speech clarity interactions in these regions (although the attention x speech clarity interaction was not significant). This pattern of results suggests that mechanisms supporting attentionally gated activity in the FFA and middle occipital gyrus are also active for degraded > clear speech, to a greater extent under full attention. This suggests that enhancements in this region, whether they are obtained through AV speech or speech clarity, rely on similar mechanisms.

My ICA analysis extend these results, demonstrating greater activity for the matching and mismatching conditions relative to the auditory-only condition in the left STS and MTG, as well as right STG, but only under full attention. Note that this does not reflect intelligibility, since mismatching stimuli are equivalently intelligible to auditory-only speech. There was a high degree of overlap with previous reported coordinates of AV integration (i.e., ALE map; component 12). I also observed activity in the left superior and middle frontal gyri, left lateral orbitofrontal gyrus, bilateral middle occipital gyri, and right lingual gyrus. These same regions also demonstrated more activity for degraded than clear speech under full attention, but not distraction. As both patterns of activity were seen within the same component, my results indicate that activity in these regions is enhanced by attention when visual speech information is present, as well as when speech is degraded. This suggests that the two observed phenomena (i.e., activation for speech clarity and AV integration), both of which ultimately contribute to stimulus intelligibility.

One question raised by these results is whether there is a limit to the automaticity of integration of audiovisual speech. In the FFA, I observed attenuated, but significant differences between AV speech and auditory-only stimuli, suggesting that visual speech
information was integrated under distraction. It is possible that this pattern of significant (though attenuated) activity would diminish with a more difficult distractor task. Additional work in my lab is underway examining the effect of parametric manipulations of perceptual and cognitive load on processing of spectrally degraded speech.

Although degraded matching degraded speech was over 20% more intelligible than mismatching degraded speech, this behavioural difference was not mirrored in the pattern of observed BOLD activation. The lack of activation differences between matching and mismatching (degraded) stimuli may be due to a lack of statistical power. This is supported by the fact that differences between matching and auditory-only speech were difficult to see without the use of a small-volume correction (although these differences were more evident in the ICA analysis). In addition, this pattern of activity (at an uncorrected level) appeared to be located along the middle temporal gyrus, inferior to ALE volume map, in which activity was located along the superior temporal gyrus (see Figure 4-1). The reasons for this discrepancy in spatial location of activity are unclear and do not appear to be explained by issues related to normalization of images in my study. It is also not likely that these results reflect lack of participant engagement with mismatching speech, since spatial maps derived from ICA region implicated in cognitive effort (C13) indicated comparable, and in fact, increased, activation for mismatching speech items relative to matching and auditory-only stimuli. However, there were a high number of studies containing monosyllabic and meaningless McGurk stimuli in the ALE map, and my stimuli were complete meaningful sentences; this stimulus difference may explain the difference in the location of peak sensitivity between the ALE map and my data (STG for the ALE map; MTG for my data).
Interestingly, in middle and superior temporal regions, the pattern of activity revealed by comparing degraded matching to mismatching speech (i.e., greater activity for matching than mismatching speech) appeared to be similar, but weaker, than the pattern revealed by comparing matching and auditory-only speech (again, activity for matching > auditory-only speech). While it is possible that this may reflect activity differences related to intelligibility, it is also possible that attempted, but unsuccessful, integration may at least somewhat mimic successful integration in middle and superior temporal regions; this would be interesting to explore in future studies.

In addition, it is possible that the intelligibility of degraded speech was too high to observe a benefit for matching visual speech stimuli. It has been shown that AV multisensory enhancement in the STS increases as SNR decreases in an inverse fashion, with peak enhancement around 75% intelligibility for AV stimuli (Stevenson & James, 2008). Degraded AV speech in my study was ~90% intelligible, and therefore it is possible that I would have had greater sensitivity to detect AV enhancement with a less intelligible auditory signal.

Results from my ICA analysis also suggest that mismatching speech recruited inferior frontal cortex regions previously implicated in listening effort across a range of tasks (e.g., Wild et al., 2012; Davis & Johnsrude, 2003; Davis et al., 2011; see Table 4-5 for comparison of coordinates and Euclidian distances). Similar to the results obtained here, both Wild et al. (2012) and Davis & Johnsrude (2003) demonstrated greater activity in inferior frontal regions for degraded or distorted speech, relative to clear speech. Peaks were within 14 mm (smoothing here was 17 mm), indicating that such peaks were anatomically indistinguishable from ours. Moreover, these results are also similar to Davis et al.’s (2011) finding that semantically anomalous sentences elicit greater activity in inferior frontal regions than
meaningful ones, particularly at lower SNRs (i.e., at higher levels of intelligibility). Two peaks were within the smoothing range of 17 mm (at < 12 mm), with one peak located approximately ~24 mm away). Taken together, this pattern of results observed across studies suggests that the activity observed in left (and right) inferior frontal regions reflects effortful listening related both to speech clarity and linguistic content. Here, I extend these findings to include effortful listening related to mismatch of auditory and visual signals, suggesting that recruitment of inferior frontal regions, including anterior insular cortex, may more generally reflect attempts to resolve multiple sources of perceptual ambiguity (including ambiguity related to both meaning and form) in the bimodal speech signal.

<table>
<thead>
<tr>
<th>Study</th>
<th>MNI Coordinates</th>
<th>Euclidian Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>Current Study</td>
<td>-39</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>Wild et al., 2012</td>
<td>-33</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td>Davis &amp; Johnsrude, 2003</td>
<td>-48</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>-46</td>
<td>14</td>
</tr>
<tr>
<td>Davis et al., 2011</td>
<td>-46</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>-50</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 4-5. Comparison of inferior frontal region coordinates across studies. Coordinates from previous studies assessing speech perception in challenging listening conditions are presented. Euclidian distances are calculated for comparison with the current study (smoothing = 17 mm).

4.6.3 **Cognitive Control Networks in Speech Processing**

My ICA analysis also revealed the conditions under which cognitive control processes are recruited for speech perception. Component 13 included activation in a bilateral cingulo-opercular network (i.e., greater activity for degraded > clear speech under full attention,
compared to distraction, and greater activity for mismatching > matching and auditory-only speech under full attention, compared to distraction). Vaden Jr et al., (2013) demonstrated that greater cingulo-opercular activity predicts word recognition performance in noise on subsequent trials. They concluded that the cingulo-opercular network is likely involved in adaptive control in support of optimal speech recognition performance in challenging listening conditions, and this explanation is consistent with my observation of increased activity in this network for degraded compared to clear speech. Interestingly, the same regions also demonstrated increased activity for mismatching speech relative to matching and auditory-only speech under full attention but not distraction (note that this is a result of relatively higher BOLD activation for clear mismatching speech). Conflicting visual speech information is most readily apparent in this condition (as participants may still be trying to integrate AV speech information for less intelligible degraded mismatching speech). Thus, listeners would be most likely to inhibit conflicting visual speech information in this condition (as they were instructed to attend to the auditory input). Again, increased activation in the cingulo-opercular network for the mismatching condition relative to matching and auditory-only speech is also consistent with the explanation that such activity reflects recruitment of adaptive control networks. Further, cingulo-opercular network activity has been observed in tasks relating to other sensory modalities, (e.g., Cieslik et al., 2015; Dosenbach et al., 2006); thus, effort-related processes implicated in speech perception may reflect recruitment of more general (i.e., non-speech specific) executive control networks (Doesenbach et al., 2006; Raz et al., 2006; Fritz et al., 2007).

Furthermore, the same component also displayed activity in motor regions (pre- and post-central gyrus). This may reflect recruitment of speech motor representations previously
observed in difficult listening conditions (Saito et al. 2005; Evans & Davis, 2015; Hervais-Adelman et al., 2012; Bishop & Miller, 2009). This is consistent with my observation of greater activity in this region for degraded relative to clear speech, and also, for mismatching AV speech relative to matching and auditory-only speech (likely in attempt to resolve conflicting AV speech information), both under full attention, but not distraction.

Taken together, my results demonstrate the involvement of both the cingulo-opercular and motor speech networks in processing of degraded and mismatching AV speech under full attention (but not distraction). This suggests that both domain-general executive control networks (required for selective attention and inhibition of competing alternatives) and abstracted speech representations may be recruited in order to resolve perceptual ambiguity in the bimodal speech signal.

4.7 Summary

This study extends our knowledge of cognitive processing of degraded speech through three main novel findings. The first is a demonstration that AV speech integration is sensitive to attentional demands in bilateral STS and MTG. These areas overlapped somewhat with my ALE map of previously published coordinates of AV integration, but were generally inferior to the ALE map, in MTG. A large proportion of the studies used to derive the ALE map relied on McGurk stimuli and the precise anatomical location of AV integration may vary according to speech stimuli used.

These results also highlight the importance of distinguishing between AV speech integration and intelligibility benefits related to AV integration. Although I observed clear evidence that AV integration is sensitive to attentional demands, I was unable to demonstrate
differences between matching and mismatching speech that would allow me to draw conclusions about the intelligibility benefit afforded by successful AV speech integration. My data also indicate that sensitivity to speech clarity and AV speech integration overlap in the STS suggesting that the STS is sensitive to speech intelligibility, regardless of how this benefit is achieved (i.e., through either the auditory or visual modality). A similar pattern was also seen in the FFA; although the data did not support the finding that the FFA is sensitive to intelligibility benefits due to the addition of visual speech information, I was able to demonstrate overlap between effects of AV integration and speech clarity in this region.

The second major finding is that sensitivity to visual speech information in the FFA was enhanced when attention was on speech but did not disappear entirely when attention was on the distractor. This suggests that AV speech stimuli may be integrated or processed in these regions even in the absence of attention, although in an attenuated manner. It is possible that a stronger manipulation of attention (i.e., a more distracting distractor task) may have eliminated the sensitivity to visual speech information in the absence of attention.

Third, the cingulo-opercular network and motor speech regions were observed to be recruited under difficult listening conditions (as a result of ambiguity in either the acoustic or visual speech signal), likely reflecting recruitment of cognitive processes that aid in perceptual restoration/repair.

This work exploring the conditions under which speech processing -- and integration of facial speech gestures with acoustic speech -- are sensitive to attentional state may inform about real-world performance in several populations. For example, it suggests that individuals with attentional disorders may be disproportionately disadvantaged in challenging listening
environments. Moreover, due to increased attentional demand, speech communication for individuals with communication disorders or hearing loss may be especially challenging in noisy or multi-talker listening environments where different stimuli may compete for their attention.
Chapter 5 - General Discussion
5.1 Summary and Contributions

In my exploration of cognitive and visual speech contributions to speech comprehension in difficult listening environments, I presented two experiments, and one review paper. The results of these papers and contributions are summarized below.

5.1.1 Summary of Results from Chapter 2 – Age-related Hearing Loss and Cognitive Decline

In Chapter 2, I reviewed the documented relationship between hearing loss and cognitive decline. I highlight a number of methodological concerns in the existing literature, including small effect sizes, contamination of measures of cognition by hearing status, as well as near exclusive-reliance on cognitive screening tools, which are prone to ceiling effects. I argued that any uni-directional causal explanation for the link between age-related hearing loss and cognitive decline may be oversimplified. Although declines in one area may lead to deficits in the other, a common-cause explanation is most parsimonious, particularly as we continue to gain a greater appreciation for the link between hearing and general health factors (e.g., hearing loss and type 2 diabetes; Heizner & Contrera, 2016). Moreover, I argued that it is difficult to parcel out contributions of declines in hearing from declines in cognition due to the high degree of functional overlap between the two processes. Subsequently, I presented an alternative, integrated model illustrating how each of the hypothesized causal mechanisms may operate in tandem. Although sensory and cognitive factors cannot be fully disentangled, I provided a number of testable hypotheses to further clarify the relationship between age-related declines in hearing and cognition.

Ultimately, a complete understanding of the relationship between age-related declines in cognition and hearing requires a multi-disciplinary approach. This approach must reconcile
the independent, but cumulative contributions of loss of hearing function with sensory decline in other domains (i.e., visual, vestibular, and olfactory), associated anatomical and neural changes, in addition to general health and psychosocial factors. While amplification, auditory training, and cognitive training have some benefit on cognitive function (and mental health; Choi et al., 2016), such interventions in isolation have limited or inconsistent efficacy (see Mudar & Husain, 2016 for a discussion). Ultimately, optimal management and rehabilitation of age-related hearing loss and cognitive decline will likely require a collaborative approach between audiology, psychology, medicine, and cognitive scientists (Davis et al., 2016).

5.1.2 Summary of Results from Chapter 3 – Working Memory Training and Speech-in-Noise Performance in Older Adults

In Chapter 3 (Cogmed study), I aimed to demonstrate a causal relationship between improvements in working memory through training and increased proficiency with speech comprehension in noise in older adults. However, I was unable to find support for the efficacy of working memory training. These results add to the growing body of literature challenging the effectiveness of cognitive training. Importantly, both of the groups I tested received the experimental manipulation (i.e., cognitive training), but at different points in time. This cross-over design controlled for differences in participant motivation or engagement due to group assignment (i.e., it is expected that participants in the more challenging experimental training group would be more motivated and engaged), which could account for previously observed effects of cognitive training. The null effects observed with this rigorous study design provide a unique contribution to the literature. They demonstrate that benefits of cognitive training may not be seen when motivational effects are taken into account. Cognitive training is being aggressively marketed to the general population, despite the fact that training benefits cannot
be reliably reproduced. Recently, Lumosity was charged with substantial fines for lack of evidence supporting their claims regarding the efficacy of brain training games in ameliorating the risk of age-related cognitive decline. Although Lumosity has since softened their claims, other companies, like Cogmed, continue to advertise strong claims regarding the evidence base for their working memory training despite growing concerns over its efficacy (Melby-Lervag et al., 2016). Melby-Lervag et al. echo similar concerns in their review, highlighting the need for better-controlled studies (i.e., use of an active control group) and poor generalizability of training-related effects to real-world cognitive skills.

Moreover, recent evidence in children indicates that improved working memory performance after training with Cogmed is accompanied by significant changes in resting-state connectivity, which may reflect improvement in attentional control mechanisms (Astle et al., 2015). Although working memory training may yield improvement on tests of working memory, it is still possible that observed training effects are driven by improvements in sustained attention or attentional control, which would also be expected to have a significant impact on task performance. This is also likely to explain why benefits are less likely to be observed when an active control group is used (Melby-Lervag et al., 2016). This could also account for observed benefits of group training in older adults (e.g., Lampit et al., 2014), which would be expected to result in increased training engagement and motivation. Thus, it will be important for future research in this area to parcel out effects of sustained attention from those of working memory. Although cognitive training could emerge as a fruitful contribution to rehabilitation of age-related hearing loss, further research is required to determine the factors that are most likely to produce positive and stable effects.
Notably, an integrated ‘auditory-cognitive approach’ has recently been suggested (Ferguson & Henshaw, 2016). This would hypothetically involve training cognition directly within speech perception, by using tasks that listeners are likely to encounter in challenging listening environments. This could be achieved, for example, by using increasingly longer speech streams instead of individual words or numbers. Such an approach, if effective, is likely to yield better generalizability, since it would improve the congruence between trained tasks and real-world listening demands.

This study also demonstrated a role for working memory in comprehension of sentences in noise, particularly in the presence of semantic ambiguity. However, the results of this study also challenge the ubiquitous use of the Reading Span test as pure tests of working memory (for example, as assumed in a recent review by Fullgrabe & Rosen, 2016b). Instead, my results demonstrate that the Reading Span test may be sensitive to episodic memory, as well as other cognitive processes such as inhibition. Thus, my work points to an increased need to better understand and characterize the role of working memory and other cognitive processes (such as memory, attention, and executive functioning) in challenging listening conditions. These results also highlight the contribution of multiple cognitive domains (i.e., not just working memory) to speech listening and comprehension.

5.1.3 Summary of Results from Chapter 4 – Attention in Audiovisual Speech Integration

In Chapter 4, I examined whether audiovisual speech integration requires attention, particularly when the acoustic speech signal is degraded. This study provided support that audiovisual speech integration in bilateral middle and superior temporal sulcus (MTG; STS) is sensitive to attention. These results are also consistent with the notion that the STS and MTG is sensitive to speech intelligibility (e.g., Bishop & Miller, 2009), regardless of how this
intelligibility benefit is achieved (i.e., either due to speech clarity or visual speech information).

I also replicate previous work demonstrating that the fusiform face area (FFA) is sensitive to attention in audiovisual speech integration, adding to the body of literature documenting the role of this area in visual speech processing. Interestingly, this audiovisual benefit in the FFA did not completely disappear under distraction. Although it is possible that distractor stimuli were not sufficiently demanding to demonstrate a total reduction of integration, these results also highlight the possibility that the degree of audiovisual speech information may vary parametrically in accordance with perceptual or cognitive load.

I also demonstrate activity in the cingulo-opercular network, reflecting the probable recruitment of amodal, domain-general, cognitive control networks, in addition to recruitment of motor speech areas. Both have previously been observed under difficult listening conditions; here, I show that both are recruited to resolve ambiguity related to conflicting audiovisual speech and degraded auditory input, likely reflecting recruitment of cognitive processes that aid in perceptual restoration/repair.

Taken together, the results of this study demonstrate a role for attention and domain-general cognitive control networks in supporting speech perception in suboptimal listening conditions.

5.2 Further Discussion

The aim of this thesis was to further elucidate the underlying cognitive architecture supporting speech comprehension in challenging listening environments. The findings
detailed in Chapters 2, 3 and 4 demonstrate that cognitive processes, including memory, working memory, and executive control networks, are flexibly recruited in order to meet listening demands. Further, the robust finding that global cognitive function and sensory hearing ability tend to decline in tandem in old age highlights the interconnected nature of hearing and cognition. Ultimately, my work encourages us to think about how speech perception, particularly in noise, may rely on domain-general (not speech-specific) cognitive resources, especially in challenging listening environments.

Despite clear evidence for ‘global’ cognitive contributions to speech comprehension, there is still considerable work to be done with respect to characterizing the specific cognitive domains that contribute to the listening process. However, continued progress on this front hinges upon careful and appropriate selection of measures. Below, I outline some considerations for cognitive training and hearing research.

The dangers of inconsistent or poor measure selection are perhaps most readily appreciated in the cognitive training literature debate: generally, success in reproducing previously published results tends to vary in accordance with how cognitive outcomes are measured (see Melby-Lervag et al., 2016 for a recent review). In other words, there is a lack of convergent validity, as effects of training are typically seen with only some measures of a particular cognitive construct (e.g., working memory), but not others. This hinders our understanding of the cognitive processes targeted by training programs such as Cogmed, since we are unable to isolate the specific conditions under which training is most effective. More importantly, however, the ecological validity and utility of training and its commercial applications requires consistent demonstration of clinically significant, meaningful effects. This has yet to be demonstrated (Melby-Lervag et al., 2016).
Within the speech perception literature, there are similar issues at play. Again, there is a lack of convergent validity; only some tests of cognition correlate with speech-in-noise measures, with results that cannot be reliably replicated (Fullgrabe & Rosen, 2016b). The results of my Cogmed study also suggest that the field is struggling with a construct validity problem as well, since the popular Reading Span test is not a pure test of working memory, as commonly assumed (again, as reviewed in Fullgrabe & Rosen). Instead, performance on the Reading Span test likely reflects a combination of abilities, including selective attention, inhibition, dual-tasking, reading skill, as well as processing speed.

An additional limitation is that much of the research examining cognitive contributions to the listening process (including the link between age-related hearing loss and cognitive decline) has relied exclusively on correlational designs. However, while it is possible to demonstrate an association between two constructs, this relationship is necessarily indirect; for example, both may be related to a third, causal variable. This subsequently limits interpretability of findings. More direct evidence is needed, and this was the motivation for my Cogmed study (i.e., to examine the effect of direct manipulations of working memory on speech-in-noise performance). This approach is limited in feasibility, as it is not always possible nor easy to manipulate cognition. However, as noted in Chapter 2, the area of hearing loss rehabilitation perhaps provides a unique opportunity to examine the effect of direct cognitive (and auditory) training interventions on hearing, and also the effect of improving (sensory) hearing ability on cognition.
5.3 Directions for Future Research

Ultimately, four things are needed in order to advance our understanding of the relationship between hearing and cognition. Each of these concerns how we measure either hearing, or cognition. I discuss each of these in turn below. This is followed by a broader discussion of the role of cognitive hearing science and audiology within healthcare.

First, as I argued in Chapter 2 with respect to the link between age-related hearing loss and cognitive decline, we ultimately need to use a broader array of cognitive tests to measure a single construct (such as working memory), within a single study (i.e., improved convergent validity), in order to better understand observed relationships between hearing and cognition. For example, if working memory truly contributes to speech-in-noise performance, then this relationship should remain consistent across various kinds of measurements of working memory.

Second, in addition to replicability of results, hearing researchers must devote more attention to precision in their measurement selection. Any claims regarding the particular profile of cognitive functions involved in speech perception should be demonstrated through specificity. For example, in order to conclude that working memory contributes exclusively to speech in noise performance, it is not only necessary to demonstrate a correlation with working memory, but also a lack of correlation with other cognitive constructs (such as memory and attention). It is also important to consider the overlap between constructs like working memory and executive functioning; both share a large degree of common variance and appear to be related to higher-level cognitive skill (McCabe et al., 2010; Engle et al., 1999).
Moreover, as I have argued above with respect to the Reading Span test, it is crucial to verify, rather than to strictly assume, that tests are truly measuring what we think they are measuring. One way to ensure this is to rely more heavily and consistently on previously validated neuropsychological tests or tools. These also have the added benefit of standardization, both with respect to normalized performance data and test administration. As I pointed out in Chapter 3, there are many different versions of the Reading Span test, which are often used interchangeably. Such a lack of standardization is problematic if each version yields different conclusions with respect to the test’s relationship to speech-in-noise measures. As I have argued above, there is little evidence to support the use of the Reading Span test as a measure of working memory, and this practice should thus be discontinued.

Third, increased consideration should be given to ecological validity. There has been a great deal of interest in characterizing the cognitive architecture supporting speech comprehension. Indeed, it is important to understand the contribution of individual cognitive domains to speech comprehension, with good sensitivity and specificity. However, it is quite possible that cognitive constructs may not cleanly map onto everyday speech performance. This may limit the utility of this approach in the clinic; in order for these contributions to have real-world impact, research findings must translate into clinically significant and meaningful assessment tools. It is likely that a combination of cognitive functions will best predict individual differences in the real-world listening process, and thus a composite score reflecting an inventory of cognitive measures is likely to have the strongest predictive validity and clinical utility.

Fourth, the near-exclusive reliance on pure-tone audiometry to measure hearing ability is an additional limitation in the existing literature (Wilson & Margolis, 2016; McArdle &
Wilson 2008, Wilson, 2011). At present, we lack clinical tools for quantifying deficits in hearing acuity and temporal coding, which appear to be most strongly related to speech in noise difficulties in older adults. Hearing deficits that are not measurable by audiometry are commonly assumed to be cognitive in nature, but it is increasingly likely that they reflect peripheral abnormalities that are currently unmeasurable (Plack et al, 2014)

The recent increased interest in such ‘hidden hearing loss’ further highlights the limitations associated with conventional audiometry. Recent work suggests that both elevated high-frequency audiometric thresholds and cochlear synaptopathy may underlie such hidden hearing loss (Liberman et al., 2016). Further, Plack et al., (2016) have suggested that a combination of approaches, such as auditory brainstem responses, frequency-following response (i.e., a measure of central auditory processing), in addition to behavioural measures (i.e., speech-in-noise tests; see also McArdle & Wilson, 2008; Wilson, 2011), could be harnessed as diagnostic tools for hidden hearing loss. As such, the inclusion of measurements of central auditory processing and behavioural measurements of speech-in-noise performance marks a shift towards increased emphasis on perceptual (rather than strictly sensory) processes in assessment of hearing function. As we begin to venture beyond the periphery in diagnostic assessment of hearing loss, any sharp dividing line between hearing and cognition is likely to become increasingly muddled. These gaps in current clinical hearing assessment tools signifies the need to move forward in a new direction in order to more fully characterize the full range of sensory deficits associated with age-related hearing difficulties.

Growing interest in the links between cognition and hearing has been more recently accompanied by increased recognition of the need to teach cognition in the audiology curriculum (Pichora-Fuller et al., 2013; 2016). This will enable audiologists to move beyond
traditional audiometry in order to obtain a more comprehensive approximation of the complex real-world listening process. Subsequently, audiologists in the future may more routinely obtain measurements of central auditory processing, in addition to working memory (e.g., digit or listening span), processing speed, speech-in-noise functioning, and perhaps eventually, listening effort. In turn, this will allow audiologists to better assess and predict individual differences in real-world hearing function, which depends on both hearing and cognitive status or capacity (Johnsrude & Rodd, 2015; Pichora-Fuller et al., 2016). Pichora-Fuller et al. (2016) also highlight the importance of motivational factors in the listening process, which also influence psychosocial engagement and compliance with use of amplification devices.

Audiologists are well poised to play an increasingly active role in hearing loss rehabilitation, including providing interventions such as integrated auditory-cognitive training (as discussed above; Ferguson & Henshaw, 2016). Moreover, given the well-established link between age-related declines in hearing and cognition, audiologists are likely to collaborate routinely with neuropsychologists and geriatricians or family physicians, and will thus also require working knowledge of cognition and cognitive screening tools.

The increased demand for multidisciplinary and interprofessional collaboration in hearing-loss management and rehabilitation is further supported by the recent conceptualization of ‘hearing healthcare’ (Davis et al., 2016; Pichora-Fuller et al., 2015). This term recognizes the contribution of hearing-loss management to overall health and well-being across the life-span, but particularly with respect to optimal aging. This contribution highlights increased scope for hearing healthcare professionals to be providers of recommendations in their collaborations with other healthcare professionals. For example,
older adults with hearing difficulties may require accommodations to ensure optimal healthcare service delivery and compliance (e.g., during surgery or physiotherapy; Davis et al., 2016). Audiologists and other hearing healthcare professionals with working knowledge of cognition would be well positioned to provide recommendations to ensure optimal listening environments and retention of important medical information (e.g., via strategies for noise-reduction and ensuring visibility of visual speech information).

In summary, recent advancements in our field, including characterization of hidden hearing loss, and links between hearing and cognitive and general health factors, are suggestive of an emergent paradigm shift. This is an exciting time for auditory researchers, with increased scope for evidence-based study and practice through an innovative, comprehensive, and multidisciplinary approach to hearing healthcare management and rehabilitation.
References


cognitive, and sensory variables as predictors of mortality in very old adults.

theory of shared variance among age, sensory function, and cognitive function in older
adults. _Journals of Gerontology Series B-Psychological Sciences and Social Sciences_,
56(1), P3–P11. Retrieved from <Go to ISI>://000166213500001

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of
cognitive hearing science. _Scandinavian Journal of Psychology_, 50(5), 371-384. doi:
SJOP753 [pii]

speech is clearly audible and intact. _British Journal of Psychology_, 92, 339-355.

influences affecting speech and meaningful sound perception. _Trends in Amplification_,

113.

Baart, M., & Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of
speech sounds on lipreading. _Neurosci Lett_, 471(2), 100-103.

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past,
present and future. _Neuroscience & Biobehavioral Reviews_, 44, 76-93.


*Journal of Clinical Psychiatry, 48*(8), 314-318.


*Experimental Psychology, 55*(6), 371-383.


Appendix A: Copyright Permissions

Thank you for your order!

Dear Miss. Rachel Wayne,

Thank you for placing your order through Copyright Clearance Center’s RightsLink® service.

Order Summary
- Order Date: Oct 28, 2016
- Order Number: 3977621237321
- Publication: Ageing Research Reviews
- Title: A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline
- Type of Use: reuse in a thesis/dissertation
- Order Total: 0.00 CAD

View or print complete details of your order and the publisher’s terms and conditions.

Please note: You will be invoiced within 24 hours.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this survey to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com
Appendix B: Pilot Study Results from Chapter 4

Introduction

This appendix details the results of the pilot experiment for the fMRI experiment detailed in Chapter 4. My aim was to adjust the level of intelligibility of spectrally degraded (noise-vocoded) sentences in order to minimize floor and ceiling effects, when sentences were used in auditory-only and matching audiovisual speech conditions, respectively. Task was the same as in Chapter 4; I measured for each condition for each participant the average number of words reported correctly from sentences presented in that condition. I tested 6-band noise-vocoded speech, and 4-band noise vocoded speech, which is less intelligible. In all cases, intelligibility testing was followed by a surprise recognition memory post-test. I hypothesized that matching degraded speech would be more intelligible than mismatching or auditory-only degraded speech. In preparation for imaging, I wanted to ensure that stimuli intelligibility was calibrated to an optimal level where behavioural differences between these conditions would be most apparent, in both my intelligibility and speech recognition data.

Participants

I recruited a total of 6 participants for the 4-band pilot; testing for the 4-band pilot was terminated early due to concerns with floor effects in post-test recognition memory data. I recruited 24 participants for the 6-band pilot.

Methods and Materials

Sentence sets were identical to those used in Chapter 4. They were also counterbalanced in a similar manner to the fMRI study. The behavioural procedure was
identical as well with the exception that sentences were presented over headphones in a sound-attenuating booth (rather than an MRI system). Subjects were asked to type out all the words they could understand from the sentences, rather than indicate whether or not they understood the gist, in order to provide a richer estimate of intelligibility. Auditory distractor materials were identical to the fMRI study.

Results

Auditory Distractor Task

The mean d’ was comparable across the 6- and 4-band tasks, with d’s of 2.65 (SD = 0.76) and 2.40 (SD = 0.99) respectively.

Intelligibility Data

6-band

Intelligibility scores are reported as proportion of sentence reported correctly. Results are presented in Figure Appendix B-1. Clear sentences (M = 0.99, SE = 0.00) were significantly more intelligible than degraded sentences (M = 0.83, SE = 0.01; main effect of speech clarity: F (1,23) = 188.62, p < .001, ηp² = .89). There was a main effect of AV speech type (F (1,23) = 14.15, p < .05, ηp² = .70). Sentences containing matching AV information (M = 0.95, SE = 0.01; M = 0.81, SE = 0.03) were significantly better recognized than sentences containing both mismatching AV information (M = 0.87, SE= 0.01) and auditory only information (M= 0.89, SE = 0.01). The speech clarity x AV type interaction was significant (F (1,23) = 26.91, p < .001, ηp² = .71). Degraded AV matching speech (M= 2.96, SE = 0.19) was significantly more intelligible than degraded AV mismatching (M= 2.42, SE = 0.21) or AO
speech ($M = 2.51, SE = 0.23$) for degraded sentences, but not clear sentences ($M = .99; p < .001$).

4-band

Again, there was a main effect of speech clarity, with clear sentences ($M = 0.99, SE = 0.02$) being significantly more intelligible than degraded sentences ($M = 0.42, SE = 0.05; F(1,5) = 151.25, p < .001, \eta_p^2 = .97$). There was also a main effect of AV speech type ($F(1,5) = 151.25, p < .001, \eta_p^2 = .88$) such that sentences containing matching AV information ($M = 0.81, SE = 0.03$) were significantly better recognized than sentences containing both mismatching AV information ($M = 0.66, SE = 0.03$) and auditory only information ($M = 0.66, SE = 0.03$). The speech clarity x AV type interaction was significant ($F(1,5) = 10.84, p < .05, \eta_p^2 = 0.84$). Degraded AV matching speech ($M = 0.32, SE = 0.06$) was significantly more intelligible than degraded AV mismatching ($M = 0.33, SE = 0.06$) or AO speech ($M = 0.32, SE = 0.06$) for degraded sentences, but not clear sentences ($M = .99; p < .001$).
Figure Appendix B-1. Sentence word report scores by condition for both 6- and 4-band noise-vocoded speech. Error bars reflect standard mean error.

**Post-test Recognition Data**

**6-band**

Recognition scores were calculated using d’ Data are presented in Figure Appendix B-2. Attended sentences \( (M = 2.72, SE = 0.13) \) were better recognized than unattended ones \( (M = 1.31, SE = 0.12) \); main effect of attention: \( F(1,23) = 158.25, p < .001, \eta^2_p = .87 \). Clear speech \( (M = 2.10, SE = 0.11) \) was significantly better recognized than degraded speech \( (M = 1.93, SE = 0.12) \); main effect of speech clarity: \( F(1,23) = 8.10, p < .05, \eta^2_p = .26 \). The main effect of AV speech type was also significant \( F(1,23) = 11.94, p < .001, \eta^2_p = .53 \). Sentences containing matching AV information \( (M = 2.15, SE = 0.12) \) were significantly better recognized than sentences containing both mismatching AV information \( (M = 1.96, SE = 0.11) \) and auditory only information \( (M = 1.94, SE = 0.12) \). There were no significant differences in recognition scores between mismatching AV and auditory-only sentences. Only the AV speech type x speech clarity interaction was significant \( F(1,23) = 6.74, p < .05, \eta^2_p = .38 \). Both mismatching and auditory only sentences were better recognized when they were presented in their clear form compared to their degraded form \( (p < .05 \) for both), but matching sentences were recognized equally well regardless of how it was presented. The interaction between attention and speech clarity trended towards significance \( F(1,23) = 3.03, p = .095, \eta^2_p = .12 \); clear sentences were better recognized than degraded sentences under distraction, but not full attention \( (p < .001) \). The attention x AV type interaction was not significant, and neither was the three-way interaction.
There was a main effect of attention such that attended sentences ($M = 1.20, SE = 0.42$) were better recognized than unattended ones ($M = 0.43, SE = 0.19; F(1,5) = 7.25, p < .05, \eta^2_p = .59$). There was also a main effect of speech clarity; clear speech ($M = 1.05, SE = 0.31$) was significantly better recognized than degraded speech ($M = 0.58, SE = 0.29; F(1,5) = 75.81, p < .001, \eta^2_p = .94$). The main effect of AV speech type was not significant. All interactions, including the 3-way interaction, were not significant.

Figure Appendix B-2. d’ scores for recognition memory post-test. Error bars reflect standard error of the mean.
**Discussion**

Identical effects were seen with both 6- and 4-band speech for sentence intelligibility data; as hypothesized, degraded matching AV speech information was significantly more intelligible than either mismatching or auditory-only materials. With respect to the recognition memory post-test, for both 6- and 4-band materials, attended sentences were significantly better recognized than sentences presented when participants attended to the distractor, as expected. Clear materials were also significantly better recognized than degraded ones for both. There was a main effect of AV speech type for matching stimuli relative to auditory-only and mismatching sentences, but only for 6-band materials. The AV speech type x speech clarity interaction was significant for 6-band materials; matching stimuli was better recognized regardless of whether they were presented in clear or degraded form, whereas clear auditory-only and mismatching speech were better recognized in the clear form rather than their degraded form. There were no significant interactions for 4-band material for the recognition post-test; although this is likely attributable to lack of power due to a limited number of subjects, I was concerned about floor effects in the data (particularly the recognition post-test), in addition to smaller differences in the intelligibility of matching speech relative to other AV speech types in 4-band noise-vocoded speech. Thus, I proceeded to use 6-band speech for the imaging study.
Appendix C: Ethics Approval (for Studies Detailed in Chapters 3 & 4)

September 27, 2013

Dr. Ingrid Johnsrude
Associate Professor
Department of Psychology
Queen's University
Humphrey Hall
Kingston, ON K7L 3N6

GREB Romeo #: 6007400
Title: "GPSYC-578-12 Effects of training on perceptual learning of speech"

Dear Dr. Johnsrude:

The General Research Ethics Board (GREB) has reviewed and approved your request for renewal of ethics clearance for the above-named study. This renewal is valid for one year from September 25, 2013. Prior to the next renewal date you will be sent a reminder memo and the link to ROMEO to renew for another year.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period. An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours. Report to GREB through either ROMEO Event Report or Adverse Event Report Form at http://www.queensu.ca/ors/researchethics/GeneralREB/forms.html.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example you must report changes in study procedures or implementation of new aspects into the study procedures. Your request for protocol changes will be forwarded to the appropriate GREB reviewers and/or the GREB Chair. Please report changes to GREB through either ROMEO Event Reports or the Ethics Change Form at http://www.queensu.ca/ors/researchethics/GeneralREB/forms.html.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely,

Joan Stevenson, Ph.D.
Chair
General Research Ethics Board

c.: Dr. Julia Huyck, Ms. Cheryl Hamilton, and Miss Rachel Wayne, Co-investigators
Dr. Stanka Fitneva, Chair, Unit REB
Marie Tooley, Dept. Admin.
Miss Rachel Wayne  
Department of Psychology  
Queen’s University  

Dear Miss Wayne,

Study Title: The Role of Attention in Audiovisual Integration of Degraded Speech  
Co-Investigators: Dr. A. Alsius, Dr. K. Munhall, Dr. I. Johnsrude  
Full Board Meeting Date: October 15, 2012

The members of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board have examined the protocol, MR environmental safety checklist, telephone interview screening form, demographic questionnaire, written debriefing sheet, recruitment poster, recruitment email: PSYC 100 subject pool and the revised information/consent form (November 2012) for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of this letter. Please attend carefully to the following list of ethics requirements you must fulfill over the course of your study:

Reporting of Amendments: If there are any changes to your study (e.g. consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval.

Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information.

Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]
Chair, Research Ethics Board

Study Code: PSYC-127-12 Romeo #6007392
Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards and operates in compliance with the Tri-Council Policy Statement; Part C Division 5 of the Food and Drug Regulations, OHRP, and U.S DHHS Code of Federal Regulations Title 45, Part 46 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalwide Assurance Number: #FWA00004184, #IRB00001173

Current 2012 membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board:

Dr. A.F. Clark, Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)

Dr. H. Abdollah, Professor, Department of Medicine, Queen's University

Dr. R. Brison, Professor, Department of Emergency Medicine, Queen's University

Dr. C. Cline, Assistant Professor, Department of Medicine, Director, Office of Bioethics, Queen's University, Clinical Ethicist, Kingston General Hospital

Dr. M. Evans, Community Member

Dr. S. Horgan, Manager, Program Evaluation & Health Services Development, Geriatric Psychiatry Service, Providence Care, Mental Health Services, Assistant Professor, Department of Psychiatry

Ms. J. Hudacin, Community Member

Dr. B. Kisilevsky, Professor, School of Nursing, Departments of Psychology and Obstetrics and Gynaecology, Queen's University

Dr. J. MacKenzie, Pediatric Geneticist, Department of Paediatrics, Queen's University

Mr. D. McNaughton, Community Member

Ms. P. Newman, Pharmacist, Clinical Care Specialist and Clinical Lead, Quality and Safety, Pharmacy Services, Kingston General Hospital

Ms. S. Rohland, Privacy Officer, ICES-Queen's Health Services Research Facility, Research Associate, Division of Cancer Care and Epidemiology, Queen's Cancer Research Institute

Dr. B. Simchison, Assistant Professor, Department of Anesthesiology and Perioperative Medicine, Queen's University

Dr. J. Tang, Medical Resident, Department of Emergency Medicine, Queen's University

Ms. W. Weisbaum, LL.B. and Adjunct Instructor, Department of Family Medicine (Bioethics)
Speech Spatial Qualities

Advice about answering the questions

The following questions inquire about aspects of your ability and experience hearing and listening in different situations.

For each question, put a mark, such as a cross (x), anywhere on the scale shown against each question that runs from 0 through to 10. Putting a mark at 10 means that you would be perfectly able to do or experience what is described in the question. Putting a mark at 0 means you would be quite unable to do or experience what is described.

As an example, question 1 asks about having a conversation with someone while the TV is on at the same time. If you are well able to do this then put a mark up toward the right-hand end of the scale. If you could follow about half the conversation in this situation put the mark around the mid-point, and so on.

We expect that all the questions are relevant to your everyday experience, but if a question describes a situation that does not apply to you, put a cross in the “not applicable” box. Please also write a note next to that question explaining why it does not apply in your case.

Please answer the following questions, then go on to the questions about your hearing

Your name:

Today’s date:

Your age:

Please check one of these options:

- I have no hearing aid/s
- I use one hearing aid (left ear)
- I use one hearing aid (right ear)
- I use two hearing aids (both ears)

If you have been using hearing aid/s, for how long?

_____ years

_____ months

or

_____ weeks If you have two aids and have used them for different lengths of time, please write down both.
Speech Spatial Qualities (Part 1: Speech hearing)

1. You are talking with one other person, and there is a TV on in the same room. Without turning the TV down, can you follow what the person you’re talking to says?
   - Not at all
   - Perfectly
   - Not applicable

2. You are talking with one other person, in a quiet, carpeted lounge-room. Can you follow what the other person says?
   - Not at all
   - Perfectly
   - Not applicable

3. You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?
   - Not at all
   - Perfectly
   - Not applicable

4. You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation?
   - Not at all
   - Perfectly
   - Not applicable

5. You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?
   - Not at all
   - Perfectly
   - Not applicable
### Speech Spatial Qualities (Part 1: Speech hearing, continued)

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. You are in a group of about five people in a busy restaurant. You CANNOT see everyone else in the group. Can you follow the conversation?</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>7. You are talking to someone in a place where there are a lot of echoes, such as a church or railway-station building. Can you follow what the other person says?</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>8. Can you have a conversation with someone when another person is speaking whose voice is the same pitch as the person you’re talking to?</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>9. Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you’re talking to?</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>10. You are listening to someone talking to you, while at the same time trying to follow the news on TV. Can you follow what both people are saying?</td>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
### Speech Spatial Qualities (Part 1: Speech hearing, continued)

<table>
<thead>
<tr>
<th>Question</th>
<th>Scale</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. You are in a room where there are many people talking. Can you follow what the person you are talking to is saying?</td>
<td>Not at all</td>
<td>Perfectly</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>12. You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?</td>
<td>Not at all</td>
<td>Perfectly</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>13. Can you easily have a conversation on the telephone?</td>
<td>Not at all</td>
<td>Perfectly</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>14. You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both?</td>
<td>Not at all</td>
<td>Perfectly</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
</tbody>
</table>
### Speech Spatial Qualities (Part 3: Qualities of hearing)

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Think of when you hear two things at once, for example, water running into a basin and, at the same time, a radio playing. Do you have the impression of these as sounding separate from each other?</td>
<td><img src="#" alt="Rating Scale" /></td>
<td><img src="#" alt="Rating Scale" /></td>
</tr>
<tr>
<td>2. When you hear more than one sound at a time, do you have the impression that it seems like a single jumbled sound?</td>
<td><img src="#" alt="Rating Scale" /></td>
<td><img src="#" alt="Rating Scale" /></td>
</tr>
<tr>
<td>3. You are in a room and there is music on the radio. Someone else in the room is talking. Can you hear the voice as something separate from the music?</td>
<td><img src="#" alt="Rating Scale" /></td>
<td><img src="#" alt="Rating Scale" /></td>
</tr>
<tr>
<td>4. Do you find it easy to recognize different people you know by the sound of each one’s voice?</td>
<td><img src="#" alt="Rating Scale" /></td>
<td><img src="#" alt="Rating Scale" /></td>
</tr>
<tr>
<td>5. Do you find it easy to distinguish different pieces of music that you are familiar with?</td>
<td><img src="#" alt="Rating Scale" /></td>
<td><img src="#" alt="Rating Scale" /></td>
</tr>
</tbody>
</table>
### Speech Spatial Qualities (Part 3: Qualities of hearing, continued)

<table>
<thead>
<tr>
<th>Question</th>
<th>Scale</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Can you tell the difference between different sounds, for example, a car versus a bus, water boiling in a pot versus food cooking in a frypan?</td>
<td>Not at all, 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>7. When you listen to music, can you tell which instruments are playing?</td>
<td>Not at all, 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>8. When you listen to music, does it sound clear and natural?</td>
<td>Not at all, 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>9. Do everyday sounds that you can hear easily seem clear to you (not trained)?</td>
<td>Not at all, 10</td>
<td>Not applicable</td>
</tr>
<tr>
<td>10. Do other people's voices sound clear and natural?</td>
<td>Not at all, 10</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
### Speech Spatial Qualities (Part 3: Qualities of hearing, continued)

<table>
<thead>
<tr>
<th>Question</th>
<th>Unnatural</th>
<th>Natural</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Do everyday sounds that you hear seem to have an artificial or unnatural quality?</td>
<td>0 1 2 3 4 5 6 7 8 9 10 Not applicable</td>
<td></td>
</tr>
<tr>
<td>12. Does your own voice sound natural to you?</td>
<td>Not at all</td>
<td>Perfectly</td>
</tr>
<tr>
<td>13. Can you easily judge another person's mood from the sound of their voice?</td>
<td>Not at all</td>
<td>Perfectly</td>
</tr>
<tr>
<td>14. Do you have to concentrate very much when listening to someone or something?</td>
<td>Concentrate hard</td>
<td>Not need to concentrate</td>
</tr>
<tr>
<td>15. Do you have to put in a lot of effort to hear what is being said in conversation with others?</td>
<td>Less of effort</td>
<td>No effort</td>
</tr>
</tbody>
</table>

Not applicable
### Speech Spatial Qualities (Part 3: Qualities of hearing, continued)

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
<th>Not at all</th>
<th>Perfectly</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.</td>
<td>When you are the driver in a car, can you easily hear what someone is saying who is sitting alongside you?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>When you are a passenger can you easily hear what the driver is saying who is sitting alongside you?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Can you easily ignore other sounds when trying to listen to something?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not applicable
### Appendix E: List of Stimuli

#### Chapter 3 – Coherent/Anomalous Sentence Sets

<table>
<thead>
<tr>
<th>Anomalous</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>the gown laughed at the candle of her autumn</td>
<td>the sketch showed that the road would pass the school</td>
</tr>
<tr>
<td>the cave was signed in a young headache</td>
<td>the author wrote the book that year</td>
</tr>
<tr>
<td>the alley was large scale throughout fire</td>
<td>the rice was cooked in a large saucepan</td>
</tr>
<tr>
<td>there were tweezers and novices in her listener heat</td>
<td>it was the women that complained when the old bingo hall was closed</td>
</tr>
<tr>
<td>she offered at the fish before it was local</td>
<td>the beef was rare just as the customer had requested</td>
</tr>
<tr>
<td>the atom was meant in a fringe at the chunk</td>
<td>it was a sunny day and the children were going to the park</td>
</tr>
<tr>
<td>the badger was called in a young steamer</td>
<td>the coin was thrown onto the floor</td>
</tr>
<tr>
<td>the doll was light just as the corridor had exerted</td>
<td>she grew tomatoes in her greenhouse</td>
</tr>
<tr>
<td>the lime in the engine was quite glad</td>
<td>the award was given to the writer at the end of his career</td>
</tr>
<tr>
<td>her good slope was done in carrot</td>
<td></td>
</tr>
<tr>
<td>the country was breaking to establish the mind and leather of the balloon</td>
<td></td>
</tr>
<tr>
<td>it was the money that exclaimed when the last eagle wall was turned</td>
<td></td>
</tr>
<tr>
<td>there were weeks in the pencil</td>
<td></td>
</tr>
<tr>
<td>the cloak walked to freeze very high</td>
<td></td>
</tr>
<tr>
<td>the campus had some flame on it</td>
<td></td>
</tr>
<tr>
<td>the thing felt all of his speech at line</td>
<td></td>
</tr>
<tr>
<td>he dressed the pressure to the number in the vessel</td>
<td></td>
</tr>
<tr>
<td>the minds felt at the tile were got in stance</td>
<td></td>
</tr>
<tr>
<td>the clown on his salt was quite helpful</td>
<td></td>
</tr>
<tr>
<td>the pocket on the landlady was very single</td>
<td></td>
</tr>
<tr>
<td>he charged the lap for the niece of wheels</td>
<td></td>
</tr>
<tr>
<td>he really caught a door before going to mind</td>
<td></td>
</tr>
<tr>
<td>the thirst smiled that the wife would kill the day</td>
<td></td>
</tr>
<tr>
<td>the darling held the end that way</td>
<td></td>
</tr>
<tr>
<td>the boot was grown onto the mouth</td>
<td></td>
</tr>
<tr>
<td>it was a rusty hand and the women were getting to the inch</td>
<td></td>
</tr>
<tr>
<td>the canal was given to the title at the face of his sentence</td>
<td></td>
</tr>
<tr>
<td>some snow was agreed to the butter</td>
<td></td>
</tr>
<tr>
<td>the brain was very mild and economic to refuse</td>
<td></td>
</tr>
<tr>
<td>the studio of the county was quite high</td>
<td></td>
</tr>
<tr>
<td>she paid umbrellas in her farmyard</td>
<td></td>
</tr>
<tr>
<td>the great neck was in quiet</td>
<td></td>
</tr>
<tr>
<td>the canal was given to the title at the face of his sentence</td>
<td></td>
</tr>
</tbody>
</table>
the woman was hoping to discover the name and address of the culprit
the singer was well known throughout Europe
he always read a book before going to bed
the traffic on the expressway was very heavy
the camel was kept in a cage at the zoo
the child left all of his lunch at home
the luggage was kept in a large warehouse
there were bracelets and necklaces in her jewellery box
he searched the pack for the ace of hearts
the thief started to sprint very fast
the bride smiled at the photo of her wedding
the bruise on his knee was quite painful
the fog in the valley was quite thick
the tools found at the dig were made of bronze
some ice was added to the whisky
there were books in the cellar
the statue had some paint on it
she arrived at the shop before it was open
he guessed the answer to the question in the exam
the noise was very loud and difficult to ignore
the salary of the lawyer was quite large
the old tree was in danger
her new skirt was made of denim

his smile was rescued by the true college
the knife turned to include the volume
his zoo wrote on the blood
the disease on the mode was quite female
his room liked that his edge had kept the heart
the agenda for the soap was easy to listen
the town pointed as a coin
the bottle had been important for great eyes
the warhead trained down into the sister of the barrel
the road of the beer paid out the spiritual
a porch was called to fade the beer of gold
the corridor in the fishing word was survived when the word was penetrated
he arrested his minutes about the heart of bathroom
she spent doctors about relics ponchos and bubbles
the ice was eased by a believer
the car was early to hate the actor
the thumb was proved in a young tent
there were dimes in the bomb
he was turning at his bread in his minute
the hair of toast painted five pounds
the frailty made up over the oil of the notion
the road from the glass of the truth was appalling
anomalous  the state of months made a daisy once after they found the classmate
anomalous  the rampage stood most of his mother at the noises
anomalous  the success moved to hope the milk
anomalous  it was very national to speak his arrogance
anomalous  the high leg was clear of views
anomalous  the effect supposed to the consumer
anomalous  the box was too dead for the business
anomalous  they found the space about the cheese to the fire
anomalous  he shouted the diet of the guns
anomalous  they might that the fact was drifted
cohherent  the student tried to move the desk
cohherent  the couple had been together for three years
cohherent  they thought that the house was haunted
cohherent  the game of chess laster four hours
cohherent  she loved stories about fairies wizards and dragons
cohherent  the burglar came up over the wall of the palace
cohherent  the fireman climbed down into the bottom of the tunnel
cohherent  his wig fell on the floor
cohherent  a spoon was used to stir the cup of tea
cohherent  they told the truth about the fight to the teacher
cohherent  it was very difficult to read his handwriting
cohherent  the group of friends got a taxi home after they left the nightclub
cohherent  he enjoyed the beauty of the hills
cohherent  the drink was too hot for the baby
cohherent  the police returned to the museum
cohherent  his train was delayed by bad weather
cohherent  the pattern on the rug was quite complex
cohherent  his face showed that his team had lost the game
cohherent  the view from the top of the ridge was amazing
cohherent  the game ended as a draw
cohherent  the boy was able to climb the mountain
cohherent  the wife of the priest helped out the elderly
coherten  the guard failed to prevent the escape
coherten  he was sitting at his desk in his office
coherent  the juice was served in a large jug
coherent  the gambler lost most of his money at the races
coherent  the furniture in the dining room was removed when the room was decorated
coherent  the whole sky was full of birds
coherent  the recipe for the cake was easy to follow
coherent  he reminded his parents about the game of football
coherent  there were mice in the cave

anomalous  the war bought over the soup
anomalous  it was too dead to see itching in the coffee
anomalous  he collapsed his students by his skin of weather
anomolous the fireplace was able to follow the lips of light
anomolous bunches are often felt in the roof
anomolous her shoulder was too long for the diesel
anomolous the country pushed at the song about the leg
anomolous the arch was called in a fusion in the tart
anomolous the shoulder was famous once the salt had happened
anomolous the elite were realised to attack the brick and tax of the flamingos
anomolous he shook his task when he ran off the month
anomolous he knew day before he had seen his noses
anomolous the great election was bought down between the first form
anomolous the thing was grand when her drops were questioned
anomolous whisky is deeply moved in the window
anomolous the envelope was strange just as the biscuit had started
anomolous the slang was driven when more shops were recovered
anomolous the population husband was a practice
anomolous the art was able to propose his accident
anomolous there was a really physical runway that policy
anomolous she was standing on the collar in her engine
anomolous an upper queen was changed back at the question feeling
anomolous the day stood the secretary at grandchild
anomolous the research had a goat in its moon
anomolous his great streets were from Smith
anomolous the fees were sat in the nail
anomolous the hammer had some mother in it
anomolous the floor was threatened by the pouch
anomolous it was the rice that finished when the empty stopsign used the park
anomolous he jilted his coast before he drew it
anomolous the pressures got a mind of dress low at group
anomolous it is private for children to reduce the trumpet
cohherent the soup was kept in a carton in the fridge
coherten the man read the newspaper at lunchtime
coherten soccer is mostly played in the summer
coherten the television programme was a success
coherten the kettle had some water in it
coherten there was a really beautiful sunset that evening
coherten the neighbors made a lot of noise late at night
coherten he broke his leg when he fell off the horse
coherten the truce was broken when more guns were delivered
coherten his new clothes were from France
coherten it was too cold to go camping in winter
coherten he ironed his shirt before he wore it
coherten the woman laughed about the joke about the dog
coherten the church was destroyed by the blaze
coheren they the boy was able to conceal his cigarette
coherten she was sitting on the sofa in her bedroom
coherten the building had a nest in its roof
coherent 
the elephant was huge just as the circus had wanted

coherent 
the cows were kept in the barn

coherent 
it is common for people to avoid the dentist

coherent 
her daughter was too young for the disco

coherent 
the child was sad when her toys were damaged

coherent 
the car drove over the cliff

coherent 
an angry crowd was turned back at the government building

coherent 
the new computer was sent back after the first month

coherent 
the audience was quiet when the song was started

coherent 
the housewife was able to carry the bags of food

coherent 
he left school before he had done his exams

coherent 
the panel were supposed to ignore the height and weight of the contestants

coherent 
spiders are often found in the tub

coherent 
it was the crew that remained when the final lifeboat left the ship

coherent 
he surprised his parents by his lack of concern

Chapter 4 – Experimental Sentence Sets

he deserved the respect of his colleagues
the money for the science library was increased when the university was modernized
her children saw a snake at the picnic
roses will start to bloom very soon
the child was sad when her toys were damaged
the audience was quiet when the song was started
the students thought the museum was very boring
the public stopped attending the games after a bad start to the season
the horn was so loud that they all jumped at the noise
snow is unusual in the summer in most countries
there were many sparrows in the sky just above the trees
his face showed that his team had lost the game
their holiday was quite short and would end soon
the gate to the church was quite rusty and difficult to open
the guard tried to prevent the escape
trains are often delayed by bad weather
the care given by the nurses on the ward was very professional
the gambler lost most of his money at the races

the children thought the dolphin was beautiful
we were lucky that the hammer was kept in the toolbox
her daughter was too young for the nightclub
the thief started to sprint very fast
the child left all of his lunch at home
she was sitting on the sofa in her bedroom
she claimed that the cookies tasted much nicer
the beef was rare just as the customer had requested
the scouts and the guides always went on long hikes in the summer
he always read a book before going to bed
a splash of gin tastes really good with ice and lemon
the couple had been together for three years
there was beer and cider on the kitchen shelf
it was the women that complained when the old bingo hall was closed
they thought that the house was haunted
his boss played golf nearly every weekend
everyone was worried as the exam was much harder than expected
the flag was raised to the top of the flagpole

the singer was well known throughout Europe
it is because the ant lived under the rocks that it survived the explosion
the blunt knife was rather awkward to use
spiders are often found in the tub
she laughed at the joke about the dog
the view from the top of the ridge was amazing
there was lettuce and cucumber in the salad
the vessel was still watertight even when badly battered
we noticed that the pen shook when the man signed the form
the recipe for the cake was easy to follow
she hurt her ankle while she was cycling to the village
the boy was able to conceal his cigarette
it was too cold to go camping in winter
the group of friends got a taxi home after they left the nightclub
he enjoyed the beauty of the hills
actors normally perform at the theatre
the safety rules of the apartment were important to follow
the housewife was able to carry the bags of food

the student wrote many essays that year
the garlic and the herbs were added to the fried onion
he ironed his shirt before he wore it
the crooked tree was in danger
he added milk and sugar to his coffee
they told the truth about the fight to the teacher
the new owners of the house painted it pink
the cupboard contained ingredients he had never seen before
they thought that the stable would cost more than the house to heat
he reminded his parents about the game of football
the soldier had a map that showed him all the details
his handwriting was very difficult to read
the luggage should be kept in a large warehouse
the top of the tower had a wonderful view of the city
her new skirt was made of denim
there were forks in the drawer
the chocolates and the flowers were bought from the nearest florist
we were disappointed that the cake had not been touched

the man read the newspaper at lunchtime
the fireman climbed down into the bottom of the tunnel
the boy was able to climb the mountain
the path turned north towards the forest
the neighbors made a lot of noise last night
his girlfriend had chosen the picture on the wall
the traffic on the highway was very heavy
the elephant was huge just as the circus had wanted
If your tooth hurts that much you ought to see a dentist.
she loved stories about fairies wizards and dragons
the shoes were not the colour that the young girl wanted
He broke his leg when he fell off the horse
taking a nap can help you stay up later
it was a sunny day and the children were going to the park
the whole sky was full of birds
there were books in the cellar
the fumes from the factory are unbearable in the village
the pension payments were worth less and less every month

soccer is mostly played in the summer
the furniture in the dining room was removed when the room was decorated
the drink was too hot for the baby
the kettle had some water in it
We live a few miles from the main road.
the computer was sent back after the first month
a game of chess can last for four hours
the platform started creaking alarmingly during the speech
her backpack was full of things that she would need for her camping trip
there were bracelets and necklaces in her jewellery box
we had to be careful that the ferry was on time
there were mice in the cave
the canyon was filled with haze on sunny days
it was a cloudy week so the residents stayed in their rooms
the church was destroyed by the blaze
the competition ended as a draw
taking a hostage allowed the robbers to make their escape
the burglar came up over the wall of the palace
the lawyer has quite a large salary
the woman was hoping to discover the name and address of the culprit
he replied that the songs were quite good
some milk was borrowed from his neighbour
the building had a nest in its roof
the bride smiled at the photo of her wedding
the wife of the priest helped out the elderly
the sketch showed that the road would pass the school
they drove from the seaside to the city at the end of the day
the party began to get livelier some time later
they hoped that the pill did not have any side effects
the bait should be suitable for catching fish
he was sitting at his desk in his office
awards are given to good writers at the end of their careers
the plane flew over the buildings
the television programme was a success
the professor insisted that the students should submit their essays on time
he guessed the answer to the question in the exam

the athlete tried to win the marathon
the panel were supposed to ignore the height and weight of the contestants
the pain tempted him to abort the climb
her mother was making a cake
the juice was served in a large jug
the artefacts found at the dig were made of bronze
she cleaned the closet after she emptied it
there were tools made from gold found at the site
the pole did not support their weight as they climbed over the gate
her cousin had informed the doctor of his symptoms
he explained that the arch had been built by the Romans
airplanes are currently the best way to travel
his briefcase was brown and was made of leather
it was the crew that remained when the final lifeboat left the ship
the statue had some paint on it
the car drove over the cliff
the soldier saluted the flag with his rifle by his side
If you don’t want these old magazines, throw them out.

the police returned to the museum
the fight in the playground was over a packet of gum
the pattern on the rug was quite complex
she grew tomatoes in her greenhouse
the fog in the valley was quite thick
the carpet and the curtains were the same colour
the dock should be fairly quiet on saturdays
the tie attracted attention because of its odd appearance
the town had pubs that were quite cheap and easy to find
the shrubs are watered regularly by the gardener
the queen went on a tour of the country that summer
the win helped our team avoid elimination
the tray should have been returned to the kitchen
the dessert was put into the oven at the start of the meal
the student tried to move the desk
the garage was closed on weekends
the goat was as greedy as the family had expected
the soup was kept in a carton in the fridge

the restaurant was bought by the hotel
the mayor used cash to bribe the reporters before they exposed him to the public
the bruise on his knee was quite painful
some ice was added to the whisky
he searched the pack for the ace of hearts
the truce was broken when more guns were delivered
it would be much easier if everyone would help.
a spoon was used to stir the cup of tea
the gifts sold to the tourists in the shop were quite cheap
it is common for people to avoid the dentist
he met his father while he was walking to the shops
there was a really beautiful sunset that evening
he left school before he had done his exams
the dentist needed somewhere to relax at the end of the day
the cattle were kept in the barn
the boat drifted across the pond
opening the can takes a long time with a rusty pocketknife
the drought was eased by the arrival of the monsoon

the bishop was welcomed into the chapel
the bathroom was decorated by the family to help them sell the house
the rice was cooked in a large saucepan
daisies will begin to grow quite soon
the wax from the candle fell on the book
he surprised his parents by his lack of concern
the cake and the cookies had the same flavour
they were concerned when the kid laughed at violent movies
the gems found in the store were not worth very much money
a severe storm left the walnut tree badly damaged
thunder was heard when the children were all in their rooms
his wig fell on the floor
the soldiers thought that helmets would save their lives
it is best if the hamster stays in the shade during the summer
the coin was thrown onto the floor
his new clothes were from Buffalo
the children were hoping to play some hockey and rugby at their school
the soccer match could begin after the field was mowed

she wrote her secrets in her diary
there has been a tree towering above this house for the last fifty years
the shop was closed when she arrived there
the goal was scored by a defenseman
a new shopping mall was built last year
the kiln was hot enough to fire the pots
she thought her jacket made her look very smart
an angry crowd was turned back at the government building
they walked from the cottage down the path to the edge of the forest
the noise was very loud and difficult to ignore
his uncle had some sheep that lived out in his garden
gin was not a drink that her old man liked
the den should be an ideal place to study
it was agreed that the name of the ship would be Titanic
the author wrote the book that year
the king was making many enemies
it was obvious that the neighbourhood was dangerous to drive through
the camel was kept in a cage at the zoo

Chapter 4 – Foil Sentences (Recognition Post-test)

she saw a hare while she was skipping across the field.
a bar was used to smash the pane of glass.
there were raisins in the bread.
bark is found on the trunk of many trees.
a pen was used by the farmer to enclose the stock before he moved them to the market.
the board tried to prevent the strike.
it was free for juniors to join the club.
the plot of the story was extremely odd and difficult to follow.
the shell was fired towards the tank.
he was lying underneath the palm on the beach.
the port was used for the toast at the end of the banquet.
the weight was too much for the scales.
the waist of the jeans was very narrow.
there were dates and pears in the fruit bowl.
his presents arrived in the mail.
there was a mole on his temple just below his hairline.
the letters and the digits were the identical size.
the principal decided that the boarders should all
his new post was in china.
the flour was added to the sauce.
the pitch of the note was extremely high.
there was thyme and sage in the stuffing.
they kept a record of the events in the log.
the poll suggested that the party would lose the election.
it was the weak that suffered when the new ruler came to power.
the lock on the chest had been broken with the poker.
his calf was only strained and would heal quickly.
the prophet had a staff in his hand.
the cymbals were making a racket.
the change was meant as a tip for the waitresses.
a spade was not the suit that the card player wanted.
a bug was used to tap the apartment.
the ring was still in its case when they left the jewellers.
the ball was organized by the students to celebrate the end of the term.
the cabinet was surprisingly light and easy to carry.
the head of the local branch was replaced when the company was reorganized.
the beech and the ash were common in the local forests.
the cast learned their cues that afternoon.
a band was sewn onto the hat.
the seal came up onto the bank of the river.
the creak came from a beam in the ceiling.
she missed the company of her friends.
the boat left a wake behind it.
she filed her nails before she polished them.
the knight began to charge on his horse.
the sentence was decided by the court.
the bus started to brake too late.
the opening chord was drowned out by the bass guitar.
he walked from the dock to the cell at the end of the trial.
the break given to the guards between their watches was very short.
the blind on the window kept out the sun.
the star had many fans who came to all his concerts.
flower bulbs are normally sold during the fall.
the punch was served in a large pitcher.