How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multi-talker listening situations.

by

Meital Avivi-Reich

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

© Copyright by Meital Avivi-Reich 2015
How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multi-talker listening situations.

Meital Avivi-Reich
Doctor of Philosophy
Department of Psychology
University of Toronto
2015

Abstract

Conversations in noise challenge the perceptual and cognitive capabilities of older adults and those listening in their second language (L2), and might force them to alter the balance between the contributions of bottom-up versus top-down processes involved in spoken language comprehension. We investigated the extent to which individual differences in vocabulary and reading comprehension skills are related to individual differences in spoken language comprehension.

In Experiments 1 and 2 younger and older L1s as well as young L2s listened to conversations in English played against a babble background and answered questions regarding their content. Individual hearing differences were compensated for by creating the same degree of difficulty in recognizing spoken words in babble. In Experiment, 1 two-talker conversations were played with or without either real or virtual spatial separation between the talkers and masker. The results showed that all listeners performed better when there appeared to be spatial separation. The contribution of vocabulary to dialogue comprehension was larger when spatial...
location was virtual rather than real, whereas the contribution of reading comprehension differed as a function of age and language proficiency. In Experiment 2, three-talker conversations, with or without spatial separation, were played in either quiet or against moderate or high babble. The contribution of individual differences in vocabulary and reading comprehension skills differed among the groups, and the babble level. In both experiments compensating for differences in spoken word recognition, minimized the differences in conversation comprehension among the groups. In addition, the manner in which spoken language comprehension is achieved is modulated by the auditory scene, and the listeners’ age and linguistic status.

Experiment 3 investigated how the perceived compactness of sound sources affects spoken language recognition. Younger L1s, older L1s and young L2s were asked to repeat meaningless sentences played in noise, babble or speech, either with or without contrast of diffuseness between target and masker. Results showed a release of masking due to contrast, which was greater when the masker was speech compared with noise. Individual differences in speech recognition were related to individual differences in vocabulary and reading comprehension in young L2s and older L1s only.
Acknowledgments

I am sincerely grateful to my PhD supervisor, Bruce Schneider, to whom I am greatly indebted for providing me with extensive training, guidance, mentorship and support throughout the course of this research. I am also grateful to my committee members, Meredyth Daneman and Eyal Reingold, for their advice and excellent feedback. I would also like to thank Susan Allison, Jane Carey, Lulu Li, James Qi and members of the Human Communication Laboratory (HCL) at the University of Toronto Mississauga, for all their help with the research. Finally, I would like to thank my family for the continuing encouragement and support throughout the years. I am especially grateful to my husband, Dor Reich, for always being there for me, and my children, Maayan and Naveh, for being such good kids.

As listed below, the first experiment from my thesis has been previously published. I would like to thank the reviewers and editors for their helpful comments.

Experiment 1:


Experiment 2:

Avivi-Reich, M., Jakubczyk, A., Daneman, M., & Schneider, B.A. (under revision). How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multi-talker conversations.
# Table of Contents

Acknowledgments ........................................................................................................ iv

Table of Contents ........................................................................................................ v

List of Tables ................................................................................................................ vii

List of Figures ................................................................................................................. viii

List of Appendices .......................................................................................................... ix

1 Introduction ................................................................................................................. 1

   1.1 Speech recognition in noisy environments ......................................................... 5

   1.2 Speech comprehension in noisy environments .................................................... 8

   1.3 Age-related changes in speech recognition and comprehension ....................... 12

   1.4 Nonnative speech recognition and comprehension in noise ............................. 16

   1.5 Overview of the thesis ......................................................................................... 18

2 Experiment 1 .............................................................................................................. 20

   2.1 Method ............................................................................................................... 22

   2.2 Results ............................................................................................................... 31

   2.3 Discussion .......................................................................................................... 46

3 Experiment 2 .............................................................................................................. 58

   3.1 Method ............................................................................................................... 60

   3.2 Results ............................................................................................................... 66

   3.3 Discussion .......................................................................................................... 77

4 General Discussion of Experiments 1 and 2 ............................................................. 80

5 Experiment 3 .............................................................................................................. 86

   5.1 Method ............................................................................................................... 89

   5.2 Results ............................................................................................................... 96
5.3 Discussion ................................................................................................................................. 108

6 Conclusions ............................................................................................................................... 117

References ......................................................................................................................................... 125

Appendix A: Evaluating the contribution of vocabulary knowledge and reading comprehension to R-SPIN SNR-50 thresholds and the average percentage of questions correctly answered in Experiment 1 .......................................................... 143

Appendix B: Evaluating the contribution of vocabulary knowledge and reading comprehension to performance in the quiet conditions of the Murphy et al. (2006) study ... 148

Appendix C: Sample of Trialogue and Comprehension Questions .................................................. 150

Appendix D: Evaluating the contribution of vocabulary knowledge and reading comprehension to the average percentage of questions correctly answered in Experiment 2 ................................................................................................................................. 153
List of Tables

Table 1 ........................................................................................................................................... 31
Table 2 ........................................................................................................................................... 67
Table 3 ........................................................................................................................................... 95
Table 4 ........................................................................................................................................... 97
List of Figures

Figure 1. ........................................................................................................................... 25
Figure 2. ........................................................................................................................... 28
Figure 3. ........................................................................................................................... 34
Figure 4. ........................................................................................................................... 36
Figure 5. ........................................................................................................................... 38
Figure 6. ........................................................................................................................... 40
Figure 7. ........................................................................................................................... 42
Figure 8. ........................................................................................................................... 44
Figure 9. ........................................................................................................................... 62
Figure 10. ......................................................................................................................... 69
Figure 11.. ......................................................................................................................... 71
Figure 12. ......................................................................................................................... 74
Figure 13. ......................................................................................................................... 76
Figure 14. ......................................................................................................................... 91
Figure 15.. ......................................................................................................................... 99
Figure 16.. ......................................................................................................................... 103
Figure 17.. ......................................................................................................................... 105
Figure 18.. ......................................................................................................................... 107
List of Appendices

Appendix A: Evaluating the contribution of vocabulary knowledge and reading comprehension to R-SPIN SNR-50 thresholds and the average percentage of questions correctly answered in Experiment 1 .......................................................... 143

Appendix B: Evaluating the contribution of vocabulary knowledge and reading comprehension to performance in the quiet conditions of the Murphy et al. (2006) study .......................................................... 148

Appendix C: Sample of Trialogue and Comprehension Questions .......................................................... 150

Appendix D: Evaluating the contribution of vocabulary knowledge and reading comprehension to the average percentage of questions correctly answered in Experiment 2 .......................................................... 153
1 Introduction

Conversations with friends, co-workers, family, healthcare providers, and others, often occur in noisy environments (e.g., malls, restaurants, stores, offices), in which there are a number of different sound sources that could interfere with one’s ability to communicate effectively. In particular, the presence of other talkers, who are not part of the conversation, can be distracting when one is trying to follow a conversation between two or more people. As such, the ability to communicate in a noisy multi-talker environment is an essential requirement in everyday life. Such multi-talker auditory scenes increase the complexity of both the perceptual and cognitive processes required for comprehension. To effectively follow a multi-talker conversation, the listener needs to perceptually segregate the talkers from one another, efficiently switch attention from one talker to another, keep track of what was said by whom, extract the meaning of each utterance, store this information in memory for future use, integrate incoming information with what each conversational participant has said or done in the past, and draw on the listener’s own knowledge of the conversation’s topic to extract general themes and ideas (Murphy, Daneman & Schneider, 2006; Schneider, Pichora-Fuller & Daneman, 2010). Essentially, fully comprehending what is going on in a conversation requires the smooth and rapid coordination of a number of auditory and cognitive processes.

Although following a conversation in noise is a challenging task for all listeners, this task seems to be disproportionately harder for older adults, who find such “crowded” situations particularly devastating (Hamilton-Wentworth District Health Council, 1988; CHABA Committee on Hearing, Bioacoustics, and Biomechanics, 1988; Murphy et al., 2006). Difficulties experienced by older adults could be the result of age-related declines in the auditory, cognitive,
and/or linguistic processes. In addition to age related changes occurring at different levels of the auditory system such as elevated hearing thresholds and loss of neural synchrony (see Schneider, 1997; Schneider & Pichora-Fuller, 2000 for reviews), age-related changes are also occurring in higher cognitive systems that are related to speech comprehension. These changes make it increasingly more difficult and effortful to comprehend speech in noisy situations. Indeed, such age-related changes in the ability to comprehend speech in these noisy situations can turn previously enjoyable social situations into effortful and exhausting experiences, which might lead to social withdrawal and could keep older adults from seeking information and assistance from health providers (CHABA, 1988). Therefore, investigating factors that can influence speech comprehension in noise is important for both understanding the reasons why older adults experience greater communication difficulties, as well as finding ways to alleviate or minimize these difficulties to improve their quality of life.

Moreover, young, healthy listeners who are operating in their second (L2) or even third language, may experience substantial difficulties when listening in such a complex auditory scene, because they have to comprehend a masked or distorted signal with an incomplete knowledge of the spoken language (for a review see Garcia Lecumberri, Cooke, & Cutler, 2010). Nonnative listeners of a language (L2) have lower scores than native listeners on a number of speech recognition and comprehension measures (Bradlow & Bent, 2002; Bradlow & Pisoni, 1999; Cooke, Garcia Lecumberri & Barker, 2008; Mayo, Florentine & Buus, 1997; Meador, Flege, & Mackay, 2000; Rogers & Lopez, 2008; Ezzatian, Avivi & Schneider, 2010). This difference in performance is influenced by several factors, such as duration of exposure to L2, age of acquisition of L2, degree of similarity between L1 and L2, knowledge of the L2 vocabulary and grammatical structure, frequency and extent of L2 use, etc. Considering the large number of people who need to communicate in languages other than their first on a daily basis, it
is important to further investigate the difficulties they experience in complex listening situations. Failure to successfully comprehend speech in L2 could severely restrict their ability to interact with others in business or in social situations.

Most of the studies, which have attempted to evaluate the relative contribution of perceptual and cognitive factors to how listeners process speech, have used single words or sentences as stimuli (see review by Humes et al., 2012), and have simply asked listeners to repeat what they have heard (e.g., Francis, 2010; George et al., 2007). Repetition of heard phonemes, words or sentences is referred to as speech recognition. Asking listeners to repeat what they have heard does not necessarily mean that they have processed and comprehended the information conveyed in the utterance. To assess the degree to which they have comprehended what they have heard, requires a task that probes the degree to which the information has been processed and comprehended. Therefore, there is a substantial difference between speech recognition and comprehension (Schneider, 2011), since comprehending a conversation requires much more than simply being able to repeat phonemes, words or sentences being spoken. Therefore, studies that do not require comprehension, while they provide important information that may shed light on some of the processes needed for speech recognition (e.g., stream segregation, morpheme identification, lexical access; see Houtgast & Festen, 2008 for a review), are limited in their ability to address the role played by the higher-order cognitive processes required to successfully comprehend a conversation (e.g., attention switching, information integration, memory).

When designing a study with the goal of exploring the contribution of low-order peripheral, and higher-order cognitive processes to speech comprehension, it is important not only to choose a task that requires the listener to comprehend and store the information, but also
to include additional tests, that provide information regarding individual differences in the skills and knowledge required to complete the task. Doing so, allows the experimenter not only to study the overall performance, but, to gain some insight into the processes taking place behind the scene. Two individuals, who manage to achieve the same level of speech comprehension in a situation, may have accomplished this task in different ways. A correlational approach in which individual perceptual and cognitive abilities are assessed and correlated with speech performance, can be used to tease apart the different perceptual and cognitive processes engaged by different groups of individuals. For example, when listening to speech in a noisy environment, if reading comprehension and speech comprehension are correlated in one population but not in another, we can postulate that the cognitive abilities tapped by the reading comprehension test are engaged in the first population but not in the second. Hence, studies investigating speech comprehension in older and L2 listeners using correlation analyses, could provide valuable information regarding how each of these groups is coping with the unique challenges each is facing when attempting to understand what is been said in adverse conditions. However, as far as we know, no one has yet taken this approach to assess the contribution of early perceptual versus higher cognitive processes to speech comprehension.

In light of the substantial differences between speech recognition and speech comprehension requirements, and the limited knowledge available with respect to the latter, the main goal of this thesis is to further study the reasons why older and L2 listeners may find it disproportionally difficult to follow and comprehend conversations in noisy situations using laboratory simulations of ecologically-relevant listening environments and tasks. In pursuit of this goal, speech recognition and comprehension performance will be investigated in more realistic auditory situations than those commonly used in research, using stimuli typically encountered in such situations (multi-talker conversations), in individuals differing in age and in
linguistic competence. In addition, the contribution of individual differences in cognitive functioning and in spoken language competence will be investigated using a correlational approach. Accordingly, I will begin by reviewing what is known about speech recognition in young nonnative and older native listeners. I will then discuss aspects of listening in real-world situations, and review studies, which have examined age-related changes in speech comprehension using longer and more complex stimuli such as monologues and dialogues. Finally, I will investigate additional challenges to speech comprehension that can occur when amplification is used in public and private settings.

1.1 Speech recognition in noisy environments

Much has been written about the difficulties listeners experience when attempting to perceive speech in the presence of other competing sound sources, that could vary in number, type, intensity, location, etc. (e.g., Brungart, Simpson, Ericson, & Scott, 2001; Brungart & Simpson, 2002; Arbogast, Mason, & Kidd, 2002). Any competing source, that spectrally overlaps the target speech signal, can interfere with the processing of the target speech at the auditory periphery, by creating overlapping excitation patterns in the cochlea and in the auditory nerve. This competition between target and masker at the periphery of the auditory system is often referred to as energetic masking or peripheral masking (Durlach et al., 2003). However, in addition to energetic masking, interference can occur at higher levels of auditory processing, even when the target and masker signals are clearly audible, because the listener fails to segregate the target stream from other competing sound sources. This failure may allow the content of irrelevant streams to intrude into working memory and interfere with the comprehension of the target message. This type of interference, which is often referred to as informational masking, can occur in addition to energetic masking (Freyman, Helfer, & McCall,
1999; Li, Daneman & Schneider, 2004; Schneider, Li & Daneman, 2007). For example, competing voices of other same-gender talkers who speak the same language, and have similar age, speech rate, accent, etc. (e.g. Brungart et al., 2001; Humes, Lee, & Coughlin, 2006; Vongpaisal & Pichora-Fuller, 2007), also are potent informational maskers, in addition to producing energetic masking. In such cases, the listener may be unable to parse the auditory scene into its components and keep the streams separated as the target speech unfolds. Moreover, even when a target speech signal is successfully segregated from other irrelevant speech signals, the involuntary phonemic, semantic, and linguistic processing elicited by the competing speech, can interfere with the processing of the target speech signal at more central cognitive levels.

A large number of cues that could alleviate or minimize the interference caused by energetic and/or informational masking, have been investigated in order to assess their potential to help release the target signal from masking during speech recognition (e.g., Arbogast et al., 2002; Murphy et al., 2006; Vongpaisal & Pichora-Fuller, 2007). The listener’s ability to successfully segregate competing sound sources into separate streams, largely depends on the perceptual similarities between the target signal and other irrelevant sound sources present in the auditory scene. Potentially, any dissimilarity between the target speech and the competing sounds could be used to achieve and maintain stream segregation, which, in turn, could enhance release from the interference caused by the presence of other masking sound sources (Bregman, 1990). Stimulus dissimilarities provide a number of cues that the auditory system can use at multiple levels of processing, beginning with, but not restricted to cochlear processes. Because the information they provide exists in the stimulus, these cues are often referred to as bottom-up cues or aids, and are likely to be assisting with auditory processing at the levels of acoustic and phonetic coding (Gierut & Pisoni, 1988). When the masking sound source is noise, asynchronous differences in the amplitude fluctuations between a speech target and a noise masker have been
found to improve speech recognition (Cooke, 2006). When the competing sound source is speech, cues include vocal differences between the characteristics of the target and competing talkers, such as differences in the fundamental frequency of the voices, their accents, speaking style, etc. Other cues may be related to the prosodic features of the target and competing speech, as well as differences in the intensity levels of the target and masking voices (e.g., Brungart et al., 2001; Darwin, Brungart, & Simpson, 2003; Humes et al., 2006). These cues are available to listeners even when listening is monaural. In addition, there are spatial cues that are available to listeners when listening is binaural. These cues enable listeners to detect differences in the input to the two ears, and to use these differences to spatially segregate the competing streams (Brungart et al., 2001). Differences in the real or perceived spatial location of sound sources in the auditory scene have been extensively studied (e.g., Freyman et al., 1999; Singh, Pichora-Fuller, & Schneider, 2008; Li et al., 2004; Murphy et al., 2006). Some of these studies focused on the importance of interaural level differences (ILDs; e.g., Glyde, Buchhol, Dillon, Cameron, & Hickson, 2013; Ihlefeld, & Litovsky, 2012), while others focused on that of interaural time differences (ITDs; e.g., Carhart, Tillman, & Johnson, 1967; Glyde et al., 2013; Darwin & Hukin, 1999; Culling, Hawley, & Litovsky, 2004).

However, our subjective experience or perception reflects not only bottom-up sensory information, but also our prior knowledge or expectations which could be applied top-down to enhance speech recognition (Sohoglu, Peelle, Carlyon, & Davis, 2012). Lexical and grammatical knowledge, as well as prior information regarding the context or the target/masker, can affect how individual speech sounds are perceived. Such top-down knowledge driven cues include knowledge of the target talker’s identity (Yonan & Sommers 2000; Newman & Evers 2007), familiarity with a talker’s voice (e.g. Brungart et al., 2001; Newman & Evers 2007; Yang et al., 2007), knowledge of a source’s location (Kidd, Arbogast, Mason, & Gallun, 2005; Singh et al.,
2008), a prior knowledge of the message’s topic (Wu et al., 2012a; Wu et al., 2012b), and a preview of the signal to be detected or a preview of the masker to be presented (e.g., Richards & Neff, 2004). Top-down cues can reduce the confusion between the target and the competing sound sources, which could result in a substantial release from informational masking. For example, Richards and Neff (2004), in reviewing several studies, found an average of approximately 15 dB improvement in speech recognition, when knowledge of the characteristics or content of either the target or the masker were provided (see their Figure 5).

1.2 Speech comprehension in noisy environments

Listening and communicating in real-world situations requires more than just speech recognition. For example, listening to a lecture requires building connections between the described events or information and existing representations in memory, in order to construct integrated mental representations (Zwaan & Rapp, 2006). In other words, comprehending several minutes of connected discourse requires a long-term engagement of cognitive resources (Gordon, Daneman, & Schneider, 2009), during which the listener has to quickly coordinate multiple levels of representation, such as phonemes, words, syntax, propositions, and global messages/narrative (Graesser, Millis, & Zwaan, 1997). The requirement for long-term cognitive engagement is added to the need for ongoing engagement of perceptual resources, in order to keep the bottom-up perpetual input coming in as intact and as clear as possible. It is reasonable to assume that as the message becomes more complex and the listening situation more difficult, the degree to which these higher cognitive processes are engaged will increase. In such a listening situation, the increased listening effort as well as the greater cognitive load could result in feelings of fatigue.
The situation becomes even more complex and quite likely more cognitively demanding when there are several people participating in a conversation. Indeed, conversation, which has been described as the primary setting for the production and comprehension of speech (Clark, 1996), introduces additional levels of complexity at both the perceptual and cognitive levels. In such listening situations, listeners need to efficiently switch attention from one talker to another as they take turns, and keep track of not only what was said but also by whom. They must also extract the meaning of each utterance, both within and across sentences, store the information for future use, draw connective inferences, integrate incoming information with previous information, and draw on the listener’s own general knowledge to extract both general and individually-related themes and ideas (Murphy et al., 2006; Schneider et al., 2010). In real life conversations, it is likely that the task of comprehending what is being said will be more demanding than the description above implies. Not only must listeners track common themes when one person stops talking and another takes over, but they must also integrate side comments made by one or more of the talkers, and recognize when a new topic is introduced. Moreover, unlike the linear and relatively organized manner in which a monologue is delivered, in a conversation, one talker might reengage the conversation before another has completed his or her sentence. This might create interruptions and inconsistencies in the information flow, which the listener has to handle quickly and efficiently, in order to not lose track of the conversation.

Given the complexities and changing demands that characterize everyday conversations, we should not be surprised that the balance between the contribution of top-down and bottom-up processes to speech comprehension is likely to be dynamic and may change, depending on the acoustic scene, and the perceptual, cognitive, and linguistic competencies of the listeners. Moreover, the nature and degree to which cognitive resources are engaged is likely to be
individually tailored to reflect the perceptual and cognitive strengths and weaknesses of the listener. For example, when listening becomes difficult, listeners might need to draw on their vocabulary knowledge to facilitate word access (Mattys, Brooks & Cooke, 2009; Mattys, Carroll, Li & Chan, 2010; Mattys & Wiget, 2011). In such a case, it might be that those who are equipped with greater vocabulary knowledge will tend to lean more towards this particular compensatory approach. Hence, we might expect to find a higher correlation between vocabulary knowledge and speech comprehension as the auditory scene becomes more complex. In addition, if attentional resources are limited (Craik & Byrd, 1982), their deployment in aid of lexical access could make it more difficult for listeners to engage other higher-order modality-independent processes (e.g., language processing, memory encoding) to help them comprehend and retain what they have heard. However, to our knowledge, there have been no attempts to investigate the degree to which listening difficulties alter the relative balance between bottom-up and top-down processing in ecologically-valid listening situations. Nor have there been attempts to determine the degree to which the contribution of different levels of processing to speech comprehension, in such situations, are modulated by the characteristics of the auditory scene, and individual characteristics such as age and linguistic competence.

To experimentally determine the reasons why people may find it difficult to follow conversations in noisy situations, we need laboratory simulations of ecologically relevant listening environments and tasks, where we can control and manipulate relevant variables such as the nature and number of competing sound sources, their spatial locations, and the signal-to-noise ratios (SNRs) under which they are presented. Several studies have attempted to shed more light on the contribution of cognitive and perceptual processes involved in speech comprehension, by examining age-related changes in performance using longer and more complex stimuli such as monologues and dialogues. Schneider et al. (2000) played monologues
to younger and older adults and asked them to answer multiple-choice questions regarding their content at the end of each one. The results showed that by individually adjusting the SNRs to compensate for differences in the speech recognition thresholds of younger and older listeners, age-related declines in the ability to process and remember a monologue could be eliminated. A later study by Murphy et al. (2006), adapted Schneider et al.’s methodology, to study how the ability to follow two-talker conversations may change with age. Participants listened to 10-15 minute dialogues between two individuals of the same gender, either in quiet or in a background of multi-talker babble, which came from the centrally located loudspeaker. After each dialogue, participants were asked to answer a set of 10 multiple-choice questions that tested the listeners’ comprehension and memory of details mentioned in the conversation. In order to simulate a typical conversation between two talkers at different spatial locations, the talkers were separated by a 9° or 45° azimuth in the first three experiments reported. In addition, a control experiment was conducted where no spatial separation between talkers was present. Their results showed that when both age groups were tested under identical conditions in the presence of spatial separation, older adults answered fewer comprehension questions correctly compared with their younger counterparts. This age difference was reduced but not completely eliminated when the listening situation was adjusted to make it equally difficult for younger and older adults to hear individual words in babble. The results of the control experiment showed that when there was no spatial separation present between talkers, the age effect was eliminated when listening situations were individually adjusted for individual differences in spoken word recognition. These results support the claim that older adults are indeed less able to successfully extract and remember information from a two-person conversation if adjustments have not been made to compensate for their poorer word recognition thresholds. Additionally, the finding that age differences persisted as long as spatial separation between the two talkers was present, even after
compensating for individual differences in word recognition, implies that older adults might benefit to a lesser extent from spatial separation compared with younger adults. This finding regarding an age difference in the ability to use spatial separation was later supported by Marrone, Mason, and Kidd (2008).

1.3 Age-related changes in speech recognition and comprehension

In noisy and complex auditory scenes, following a multi-talker conversation is a challenging task for all listeners, which can be disproportionally harder for older adults (Murphy et al., 2006), especially when the competing sources consist of other talkers (Duquesnoy, 1983; Pichora-Fuller, Schneider & Daneman, 1995). Difficulties experienced by older adults could reflect age-related declines in the auditory, cognitive, and/or linguistic processes supporting speech comprehension. Previous studies have shown evidence of changes occurring at different levels of the auditory system (e.g., elevated hearing thresholds, loss of neural synchrony, see Schneider, 1997; Schneider & Pichora-Fuller, 2000 for reviews), as well as age-related changes in the cognitive processes related to speech comprehension (Jerger, Jerger, & Porozzolo, 1991; Pichora-Fuller, 2003). These age-related declines could either cause them to experience greater interference from competing sound sources, and/or be less able to take advantage of bottom-up and/or top-down cues than their younger counterparts. Overall, age-related declines can impact the speech comprehension ability of older adults at one or more of the following processing stages: 1) the detection of the target speech signal 2) the segregation of the target steam from other competing streams, 3) the suppression of the semantic interference from irrelevant speech sources, and 4) the cognitive processing of the target signal (see Schneider et al., 2010 for a review). In addition, comprehending a multi-talker conversation would also require a quick and efficient attention switching from one talker to another.
It is reasonable to assume that age-related changes in auditory and cognitive processes required for the successful completion of any of these stages, may initiate a reorganization of the way information is processed along the auditory system and in the brain. Indeed, a number of studies have shown that older adults often engage different neural circuitry than that employed by younger adults, when performing a task (Harris, Dubno, Keren, Ahlstrom, & Eckert, 2009; Wong et al., 2009). Hence, it is likely that the relative contribution of different auditory and cognitive processes will be affected by aging (Schneider et al., 2010; Wingfield & Tun, 2007).

Older adults have been found to be more susceptible to the interference created by different types of background maskers compared to younger adults (e.g., Helfer & Freyman, 2008; Gordon-Salant, 2005). Age-related declines in the peripheral auditory system result in degraded representations of auditory signals at and beyond the cochlea (Chisolm, Willot, & Lister, 2003; Martin & Jerger, 2005; Mills, Schmiedt, Schulte, & Dubno 2006; for a review see Schmiedt, 2010). Typically, older adults, even when receiving compensation for age-related high-frequency hearing loss, require higher signal-to-noise ratios (SNRs) to perform equivalently to younger adults in noisy conditions (Jerger et al., 1991; Humes, 2005). As to the causes of this greater susceptibility, it has been found that, age-related elevations in audiometric thresholds contribute substantially, but not exclusively, to inferior word recognition in the presence of competing sound sources (Humes & Roberts 1990; Jerger et al., 1991; Humes et al., 1994; Frisina & Frisina 1997; Cervera, Soler, & Ruiz, 2009).

This increased sensitivity to the presence of a masker or to distortion in the speech signal, is often attributed to a loss of auditory neural synchrony and other temporal processing difficulties (e.g., Frisina el al., 2001; Schneider & Pichora-Fuller, 2001; Heinrich & Schneider, 2006; Pichora-Fuller et al., 2007, Wang, Wu, Li, & Schneider, 2011). Over the years, there has
been accumulating evidence for age-related declines in auditory temporal processing (Martin & Jerger, 2005; Pichora-Fuller & Souza, 2003; Schneider & Pichora-Fuller, 2001; Schneider, Daneman, & Pichora-Fuller, 2002). A listener’s spatial organization of the auditory scene relies on the detection and interpretation of small differences in the intensity and timing of the signals between the ears. Age-related changes in peripheral processing (e.g., hearing loss, especially at high frequencies), as well as loss of neural synchrony and other temporal processes, might reduce the spatial resolution (Glyde, Hickson, Cameron, & Dillon, 2011). The ability to selectively attend to sounds emanating from one direction in space, while suppressing simultaneous sounds emanating from another, is, in part, based on spatial resolution. This means that any loss of spatial resolution will likely have a cascading effect, leading to inferior speech processing in the presence of competing sound sources.

The cognitive aging literature notes that there are age-related reductions in the ability to focus attention selectively, and in preventing irrelevant information from interfering with the task at hand (Hasher & Zacks, 1988; for a review, see Schneider, Li, & Daneman, 2007), as well as evidence suggesting a general slowing of cognitive processing (e.g., Cohen, 1987; Hasher & Zacks, 1988; Salthouse, 1996; Wingfield & Stine-Morrow, 2000). The impact of these age-related changes, which could impede speech recognition and comprehension, is likely to be more noticeable as the listening situation and task become more complex and demanding. For example, comprehending a long discourse will require the listener to keep attention focused on the target speech and inhibit the processing of irrelevant sound sources for a longer time than would probably be required when the task is to simply repeat a sentence or a single word. Moreover, a multi-talker conversation will require the listener to quickly switch attention, which he or she is not required to do when there is only a single talker (Hogan, Kelly, & Craik, 2006; Levitt, Fugelsang, & Crossley, 2006; Tun, McCoy, & Wingfield, 2009). Recent studies show that
stream segregation may take a longer time to emerge in older adults than in younger adults in the
presence of speech and speech-like maskers (Ben-David, Tse, & Schneider, 2012). In real life
situations, slower stream segregation may magnify the consequences of interruptions in either
the target signal, or competing signals, on speech comprehension that can occur in a constantly
changing auditory scene (e.g., sound sources appearing and disappearing, switching from one
target talker to another). In other words, dynamic auditory scenes require the listener to re-
establish stream segregation again and again in order to maintain attention on the target stream.
The more often that re-establishing the stream segregation is required, and the longer it takes
each time, the more information might escape the listener. Age related slowing in either
peripheral or central processes, as well as suggested age-related declines in attention, will likely
disrupt the processing of information. In addition, although linguistic knowledge has been found
to be relatively preserved in older age (Burke & Shafto, 2008), it is possible that older adults,
under stressful and difficult listening situations, may experience reduced capability to utilize
their linguistic knowledge and skills, in order to enhance speech comprehension because of age-
related declines in executive functions (Clarys, Bugaiska, Tapia, & Baudouin, 2009). As a result
of all these age-related changes (as well as other changes yet to be investigated), older listeners
may use different strategies to cope with the disruptions and challenges that a complex auditory
scene entails.

In addition to the processes that are primarily handled at higher-order levels (e.g.,
attention, information integration and storage), cognitive resources might also be relocated to
assist intermediate level tasks, such as lexical access, when the bottom-up acoustic information is
insufficient for lexical access. In such cases, the engagement of cognitive resources in aid of
lexical access may deplete the pool of resources that are available for other higher-order
processes involved in comprehending speech. Hence, age-related declines in early sensory
processes that interfere with lexical access, might manifest themselves as age-related deficits in speech comprehension (for reviews see Schneider, 1997; Schneider & Pichora-Fuller, 2000, 2001, Schneider, Pichora-Fuller, & Daneman, 2010). Therefore, when examining the sources of speech comprehension difficulties, it is important to be especially thoughtful when attempting to distinguish between changes in performance, which are generated directly by cognitive declines, and those which are a secondary result of cascading effects which have a perceptual origin.

1.4 Nonnative speech recognition and comprehension in noise

Nonnative listeners also constitute a group that experiences considerable difficulty when attempting to follow a conversation conducted in their second language (L2) in the presence of competing sound sources. As Grosjean (1989) argued that the bilingual is not two monolinguals in one person, illustrating the idea that the presence of a second language likely changes how speech is processed. Nonnative listeners have been found to have lower scores on a number of speech recognition and comprehension measures, compared with adults of a similar age who were using their first language (Bradlow & Bent, 2002; Bradlow & Pisoni, 1999; Cooke et al., 2008; Mayo et al., 1997; Meador et al., 2000; Rogers & Lopez, 2008; Ezzatian et al., 2010). For instance, even at the level of identifying single words in L2, nonnative listeners will tend to be slower (e.g., Scarborough, Gerard & Cortese, 1984; FitzPatrick & Indefrey, 2009), as well as less confident and proficient (Schulpen, Dijkstra, Schriefers & Hasper, 2003), compared with native monolinguals listeners.

Nonnative young listeners are not likely to differ from native young listeners with respect to basic auditory and cognitive abilities. Therefore, it is reasonable to assume that their difficulties in L2 environments reflect the fact that their L2 semantic and linguistic processes
may not be fully developed, and/or be completely differentiated from the semantic and syntactic processes that occur in their L1 (Kroll & Steward, 1994). Several factors have been found to affect nonnative listeners’ performance; these include the duration of exposure to L2, degree of similarity between L1 and L2, the listener’s vocabulary size and knowledge of the grammatical structure of L2, as well as frequency and extent of L2 use, etc. At the perceptual level, studies have shown that the acoustic–phonetic characteristics of L2, may not be fully acquired (e.g. Florentine, 1985; Mayo et al., 1997), resulting in a reduced ability to discriminate fine phonemic information such as phonetic contrasts and phonemic categories, information which is crucial for successful speech perception (Bradlow & Pisoni, 1999; Meador et al., 2000). Failing to distinguish between sounds that are identified as different phonemes in L2, could result in an activation of additional lexical candidates as the word unfolds (Weber & Cutler, 2004). Two sources for additional candidates have been suggested: (1) greater competition from intralingual lexical candidates due to inefficient phonological processing or a phonemic confusion in L2, and (2) greater competition from interlingual lexical candidates, due to concurrent activation of words in both L1 and L2 (FitzPatrick & Indefrey, 2009). For example, a listener whose L1 does not distinguish between short and long vowels as English does, might activate words starting with /mee/ while the word /mint/ unfolds. In some cases, in which the difference between the two vowels (long/short) is the only difference between a minimal pair of words, both words might remain activated until the context in which the word was said provides enough semantic or grammatical information to sort the two and select the correct meaning. In addition, the unfolding word may activate lexical candidates not only in the target language used, but in other irrelevant languages, of which the listener commands (Spivey & Martin, 1999; Chambers & Cooke, 2009). With respect to our previous example, we might expect that any word starting with either/mee/or /mi/ in the listener’s L1 will be activated as well. In other words, in the case
of nonnative listeners, it is reasonable to assume that the difficulties they experience in L2 environments may be due to the fact their L2 semantic and linguistic processes may not be completely differentiated from the semantic and syntactic processes that are usually invoked when listening in their L1 (Kroll & Steward, 1994).

1.5 Overview of the thesis

The main goal of this thesis is to explore the contribution of low-order peripheral and higher-order cognitive processes to speech comprehension, in order to achieve better understanding of the reasons why, older listeners whose native language is English (EL1s), and young listeners for whom English is a second language (EL2s) may find it disproportionally difficult to follow and comprehend conversations in noisy situations. To increase the degree of ecological relevance of our experimental tasks to those that occur in real-world situations, we assessed speech comprehension, when people were listening in on two- and three-person conversations masked by a babble of voices. In addition, we assessed the degree to which the perceived compactness of auditory images (how concentrated the location of that source is in space) affect speech recognition, because of the increasing use of modern amplification systems that routinely alter this aspect of auditory images. For both speech comprehension and spoken word recognition tasks, we also examined the degree to which two measures of language competence (vocabulary knowledge and reading skill) could account for individual differences in performance. Specifically, we wanted to examine whether the relative contribution of these two measures of linguistic competence depended on the nature of the auditory scene, the age of the listener, and the degree to which he or she had mastered the language.
In Experiment 1 and 2, we explore the effect of age and linguistic status on the ability to successfully follow a two-talker conversation (Experiment 1; dialogue) or a three-talker conversation (Experiment 2; trialogue) conducted in a babble background noise. In Experiment 1 the locations of the sound sources were either virtual or real, and the two talkers were either spatially separated or co-located. In Experiment 2, the conversations were presented with or without spatial separation between the talkers, in either quiet or against moderate or high babble background. In both Experiments 1 and 2, Young EL1s, older EL1s and young EL2s were asked to listen to the conversations and to answer questions regarding their content. We compensated for individual hearing differences by creating the same degree of difficulty in identifying individual words in a babble background, when there was little or no contextual support for word recognition. Two measures of individual differences (vocabulary knowledge and reading comprehension skill) were included, in order to evaluate the degree to which the size of one’s lexicon as well as higher-level modality independent cognitive skills involved in language comprehension, contributed to the ability to comprehend and remember dialogues and trialogues in complex auditory scenes.

Finally, in Experiment 3, we investigated the additional challenges to speech recognition that can occur when amplification is used in public and private settings, by asking younger EL1s, older EL1s and young EL2s to repeat short syntactically-correct-but-semantically-anomalous sentences (e.g. “A rose can paint a fish”) when there was either a contrast, as there often is in natural settings, between the diffuseness of the target and masker sound sources (the target sound source was compact while the masker was diffused), or not (both the target and the masker had a compact sound source). A number of studies have shown that, the greater the acoustic differences between target and masker, the easier it is to segregate the two and process the target speech (e.g. Brungart et al., 2001; Humes et al., 2006; Vongpaisal & Pichora-Fuller, 2007).
However, to date, no one has investigated the degree to which a difference in diffuseness associated with modern amplification systems may have affected the ability to segregate the target from the background.

2 Experiment 1

Studies that examined age-related changes in speech comprehension, have found that age-related declines in the ability to process and remember either a monologue (Schneider et al., 2000) or even a dialogue (Murphy et al., 2006), could be considerably reduced and even eliminated, when listening situations are individually adjusted to compensate for differences in spoken word recognition. While Schneider et al. (2000) found, that age differences in comprehending a short lecture could be eliminated by compensating for differences in spoken word recognition, Murphy et al. (2006), found that the age difference was only reduced, but not eliminated, when the two talkers were spatially separated, even when the listening situation was adjusted to make it equally difficult for younger and older adults to hear individual words.

The consistent age difference, that was found in the presence of spatial separation, suggests that older adults might not be able to benefit as much from the full range of acoustic cues available with real spatial separation. A number of studies have shown that the improvement in speech reception thresholds that occurs when targets and maskers are spatially-separated rather than co-located, is significantly larger for young normal-hearing adults than for older, hearing-impaired adults (Neher et al., 2009; Neher, Laugesen, Jensen, & Kragelund, 2011; Schneider et al., 2010). In order to further explore the sources of the age difference in the ability to comprehend and recall a dialogue when two talkers are spatially separated, we used the precedence effect to change the virtual location of each talker. A number of studies have shown
that if the same sound is played over two loudspeakers located to the left and right of the listener, with the sound on the left loudspeaker lagging behind that on the right, the listener perceives the sound as emanating from the right and vice versa (e.g., Rakerd, Aaronson, & Hartmann, 2006). Because the sound is played over both loudspeakers, a virtual spatial separation is achieved without altering the signal-to-noise ratio (SNR) at each ear. Hence, a perceived spatial separation can be achieved, even though, there is a substantial reduction in the auditory cues supporting spatial separation (e.g., no head shadow effect). Moreover, Cranford, Boose, and Moore (1990) have shown that when the precedence effect is used to specify the virtual locations of sounds, both younger and older adults experienced the sound as emanating from the side where the leading loudspeaker was positioned. The precedence effect has been used in a number of studies related to informational masking as a way of achieving a perceived (virtual) separation between sound sources, without substantially affecting the signal-to-noise ratio at each ear of the listener (e.g., Freyman et al., 1999; Li et al., 2004). It has also been shown that both younger and older adults reap the same degree of benefit from perceived separation (Li et al., 2004). However, using the precedence effect to achieve and maintain a virtual spatial separation among sound sources may require that a larger proportion of attentional resources be allocated to stream segregation since the sound sources under the precedence effect are perceived as more diffuse and cannot be precisely located in space. This, in turn, may alter the balance between the top-down and bottom-up processes involved in speech recognition, since a reduction in the fidelity of the bottom-up, acoustic information, is likely to require a greater degree of attentional investment at the perceptual level (Neher et al., 2009). It is therefore reasonable to assume that, as a result of the changes mentioned, more weight will be given to top-down processes as the use of bottom-up information becomes limited. In particular, when there is interference with the bottom-up, stimulus-driven processes leading to lexical access, the listener may come to depend
more on those top-down processes, such as vocabulary knowledge, that could be used to aid lexical access. Hence, we might expect to find the correlation between vocabulary knowledge and speech comprehension to increase, as the auditory scene becomes more complex. In addition, the deployment of attentional resources to aid lexical access could make it more difficult for listeners to engage higher-order modality-independent processes (such as those involved in reading comprehension), to help them understand and retain what they have heard. In this study, we make what we believe is the first attempt to investigate the degree to which listening difficulties alter the relative balance between bottom-up and top-down processing, and how their relative contribution to speech comprehension is modulated by the characteristics of the auditory scene, and the listener’s age and linguistic competence.

2.1 Method

Participants

The participants were 24 normal hearing younger EL1 listeners (mean age: 21.26 years; SD: 3.02), 24 normal hearing older EL1 listeners (mean age: 69.7 years; SD: 4.6), and 24 normal hearing young EL2 listeners (mean age: 21.04 years; SD: 1.71). EL1 listeners were all born and raised in a country in which the primary language was English and were not fluent in any other language at the time of participation. EL2 listeners were those who first became immersed in an English speaking environment after the age of 14. One older listener, who found the noisy background in the study to be uncomfortable, withdrew from the experiment and had to be replaced. The young participants were volunteers, recruited from the students and staff at the University of Toronto Mississauga, while the older participants were volunteers from the local community. All participants were asked to complete a questionnaire regarding their general health, hearing, vision, and cognitive status. Only participants who reported that they were in good health and had no history of serious pathology (e.g., stroke, head injury, neurological
disease, seizures, and the like) were included. All participants reported having normal or corrected vision and were asked to use their corrective lenses when necessary. None of the participants had any history of hearing disorders, and none used hearing aids. The studies reported here were approved by the Ethics Review Board of the University of Toronto.

**Materials, Apparatus and Procedure**

During each participant’s first session, we administrated audiometric thresholds, the Nelson-Denny reading comprehension test (Brown, Bennett, & Hanna, 1981) and the Mill Hill test of vocabulary knowledge (Raven, 1965). The dialogues, along with the babble thresholds and the low-context R-SPIN SNR-50 thresholds, were administered over the next two experimental sessions (3 dialogues per session). Sessions were typically 1.5-2 hours in duration and were completed within 2 weeks. Tests were administered in a double-walled sound-attenuating chamber. All participants were paid $10/hr for their participation.

**Hearing measures**

*Audiometric testing.* Pure-tone air-conduction thresholds were measured at nine frequencies (0.25-8kHz) for both ears, using an Interacoustics Model AC5 audiometer (Interacoustic, Assens, Denmark). All younger participants were required to have pure tone air-conduction thresholds 15 dB HL or lower, between 0.25 and 8 kHz in both ears. Young participants with a threshold of 20 dB HL at a single frequency, were not excluded from the study. Older participants were required to have pure tone thresholds 25 dB HL or lower from 0.25 to 3kHz and 35 dB HL or lower for frequencies < 6 kHz. Participants who demonstrated unbalanced hearing (more than a 15 dB difference between ears under one or more frequencies) were excluded from participation.
Older adults, with hearing in the range described, are usually considered to have normal hearing for their age. Nevertheless, it is acknowledged that older adults’ hearing changes and deteriorates with age and is not equivalent to that of younger adults (Fitzgibbons & Gordon-Salant, 1996; Schneider & Pichora-Fuller, 2001; Wingfield, 1996; Wingfield, Poon, Lombardi, & Lowe, 1985; Glyde et al., 2011). The average audiograms for the three groups of participants are shown for the right and the left ears in Figure 1. The two groups of young adults had equivalent hearing levels at all frequencies. Hearing levels for older adults were about 7 dB poorer than those of the younger adults at frequencies ≤ 3 kHz, with the younger-older difference increasing as a function of frequency for frequencies > 3 kHz.
Figure 1. Average audiograms for the three groups of participants are shown for the right and the left ears.
**Babble threshold test.** This threshold was used to individually adjust the intensity of the speech signal in order to compensate for differences in the ability to detect speech between participants. The adaptive two-interval forced choice procedure employed by Schneider et al. (2000) was used to measure individual detection thresholds for the 12-talker babble masker used in this experiment. In this procedure, a 1.5 s babble segment was randomly presented in one of two intervals which were separated by a 1.5-s silent period. Two lights on the button box indicated the occurrence of each interval, and the listener's task was to identify the interval containing the babble segment by pressing the corresponding button. Immediate feedback was provided after each press. We used an adaptive two down one up procedure (Levitt, 1971) to determine the babble threshold corresponding to the 79% point on the psychometric function. Two different babble thresholds were determined for each individual. First, a babble threshold was determined when the babble was presented over a single central loudspeaker. The sound levels of the speech signals for the condition in which the voices of both talkers were presented only over the central loudspeaker (no separation, single loudspeaker condition), were individually set to be 45 dB above this babble threshold (sensation level, SL, of 45 dB). A second babble threshold was determined when the babble was played simultaneously over two loudspeakers located 45° to the right and left of the listeners. This babble threshold was used to adjust the intensity of speech signal to 45 dB SL for the conditions in which voices were presented over both lateral loudspeakers. For a graphic illustration of the two babble conditions see Figure 2, first column on the left.

**R-SPIN test.** In this test, participants are asked to immediately repeat the last word of individual sentences presented to them in a babble background. As in Schneider et al. (2000) and Murphy et al. (2006), we used the Revised Speech Perception in Noise (R-SPIN) test (Bilger,
Nuetzel, Rabinowitz, & Rzeczkowski, 1984) to individually determine the SNR producing 50% correct identification of final words of low-context sentences (e.g., “Jane was thinking about the van.”) presented in multi-talker babble under both real and virtual location conditions. In the real no-separation condition, both the babble and R-SPIN sentences were played over the central loudspeaker. In the real separation conditions, the babble was presented over the central loudspeaker and the R-SPIN sentences were presented over the right loudspeaker. In the virtual no-separation condition, both the babble and R-SPIN sentences were presented simultaneously over both loudspeakers. In the virtual separation conditions, the babble was presented over both lateral loudspeakers simultaneously with the R-SPIN sentences also being presented over both loudspeakers, but with the sentences presented over the right loudspeaker leading the sentences presented over the left loudspeaker by 3 ms. R-SPIN SNR-50 thresholds were determined for each of the conditions tested. The R-SPIN SNR-50 thresholds estimated were rounded to units of 1 dB before being integrated into the calculation of the individually adjusted SNR that was used for presentation of the dialogues. If the estimated R-SPIN threshold was calculated to be exactly half way between two integer values, rounding was conducted towards alleviating the SNR difficulty (e.g. 3.5 dB SNR was rounded to 4dB SNR and -3.5 dB SNR to -3 dB SNR). For a graphic illustration of the four R-SPIN conditions see Figure 2, second column from the left.
Figure 2. The left column specifies the two babble thresholds collected, and under what conditions they were used to adjust the signal level; the middle column specifies the conditions under which R-SPIN SNR-50 thresholds were obtained; and the right column illustrates the four dialogue comprehension test conditions in the experiment. The top row (A) specifies the babble, R-SPIN and dialogue comprehension scenarios for the real no-separation condition. The three other conditions share the same babble threshold but differ from each other in respect to the R-SPIN and the dialogue comprehension tasks. The second row from the top (B) illustrates the R-SPIN and dialogue comprehension for the real spatial separation condition, the third row (C) illustrates the settings for the virtual no spatial separation and the fourth (D) specifies the settings for the virtual spatial separation. Orange stands for R-SPIN, red stands for talker 1, blue for talker 2, and grey for babble.
Dialogue comprehension task. Each participant was asked to listen to six dialogues presented in babble noise in a sound-attenuating chamber. These dialogues were previously created and used by Murphy et al. (2006). Each dialogue was based on a published one-act play and had only two characters; in three of the six dialogues, the two characters were female, and in the other three, the two characters were male. At the end of each dialogue, which was 10-15 minutes long, the participant was presented with a series of 10 multiple-choice questions regarding the contents of that dialogue (for more information regarding the dialogues and the questions see Murphy et al., 2006). There were four conditions in the experiment: (a) Real spatial separation with the babble presented over the central loudspeaker, with the voice of one of the talkers presented over the left loudspeaker, and the voice of the other talker presented over the right loudspeaker; (b) Real, no spatial separation, with the babble, and the voices of the two talkers presented over the central loudspeaker only; (c) Virtual no spatial separation with the babble and one of two voices being played simultaneously over both lateral loudspeakers; (d) Virtual spatial separation with the babble presented simultaneously over both lateral loudspeakers, and the two voices also presented over both lateral loudspeakers with the perceived location of the two talkers manipulated using the precedence effect (the voice of talker one over the left loudspeaker playing 3 ms in advance over the same voice playing over the right loudspeaker, with the opposite timing arrangement for the second voice). For a graphic illustration of the four dialogue comprehension conditions, see Figure 2 rightmost column. Half of the participants were tested in conditions a and c (separation conditions), the other half in conditions b and d (no separation conditions). Hence, in this design, virtual versus real location was a within-subject factor and spatial location was a between-subjects factor. The participants in each group were randomly assigned to one of the two spatial location conditions. The dialogues were presented in babble at an SNR that was individually adjusted for each participant.
and condition based on his or her babble threshold and R-SPIN results according to the following calculations:

Dialogues were presented at babble threshold + 45 dB; Babble was presented at babble threshold + 45 dB – R-SPIN SNR-50 threshold + 21 dB.

To evaluate the average SPL for each of the six dialogues, first, the digitized voltage values were squared. Then, any period longer than 10 ms in length in which the root mean square (RMS) level was down by more than 50 dB from the full scale, was defined as a pause and was removed from the dialogue. Finally, the RMS was determined for the reminder of the dialogue.

At the end of each dialogue, participants were asked to answer a set of 10 multiple-choice questions with four alternatives, that were also constructed and previously used by Murphy et al. (2006). Each question referred to a specific item of information that was mentioned explicitly only once during the dialogue.

Language proficiency measures

Vocabulary knowledge. Participants were asked to complete the Mill Hill vocabulary test (Raven, 1965), which is a 20-item synonym test. In this test participants were required to match each test item with its closest synonym from six listed alternatives (for example, item number 10 in this test is “elevate”, and the six listed alternatives are: revolve, raise, waver, move, work and disperse). No time restraints were applied. The extent of a person’s vocabulary represents knowledge that can be used to facilitate word recognition. When listening becomes difficult, and the fidelity of the bottom-up information contributing to word identification becomes questionable, we might expect the role played by top-down knowledge (e.g., the extent of an individual’s vocabulary) to increase. Hence, we might expect the correlation between vocabulary
knowledge and the number of dialogue questions correctly answered to increase with listening difficulty.

**Reading comprehension skill.** The Nelson-Denny test (Brown et al., 1981) was used to assess the reading comprehension skills of each participant. In this test, the participants had to read through a series of eight independent passages and answer multiple-choice questions regarding the content of the passages. This test includes a total of 36 questions and was limited to 20 minutes. Participants were instructed to complete as many questions as possible within the time given.

### 2.2 Results

Table 1 presents the gender breakdown, mean age, educational level, Mill Hill test of vocabulary knowledge and Nelson-Denny test of reading comprehension results for each age group. One of the young EL2 listeners had an R-SPIN threshold of 22 dB SNR in the virtual no separation condition. Since this value was more than three standard deviations above the mean for that group, this value was identified as an outlier and was replaced by the average R-SPIN threshold of the young EL2 listeners group after excluding the outlier (6 dB SNR).¹

---

¹ Analyses were done for both the R-SPIN SNR-50 thresholds and the number of correct answers to the comprehension questions, in which the participant was excluded completely. All the results which were statistically significant in the analyses reported in this paper remained significant and all the results which were statistically non-significant remained as such. We have chosen to replace this one data point with the group average in order to minimize the loss of data.
Table 1

Demographic information (Mean Age and Years of Education, Gender Distribution, Mean Vocabulary and Reading Comprehension Scores) and Mean Babble and R-SPIN SNR-50 thresholds for the participants in the two separation conditions divided into the three groups tested. Statistically significant differences between groups are noted.

<table>
<thead>
<tr>
<th></th>
<th>Separation con.</th>
<th></th>
<th>No-Sepration con.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Younger natives</td>
<td>Older natives</td>
<td>Young nonnatives</td>
<td>Younger Natives</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Age</td>
<td>21.09</td>
<td>2.17</td>
<td>68.83</td>
<td>3.33</td>
</tr>
<tr>
<td>Education</td>
<td>15.58</td>
<td>2.27</td>
<td>15.85</td>
<td>2.44</td>
</tr>
<tr>
<td>Gender</td>
<td>10 F + 2 M</td>
<td>11 F + 1 M</td>
<td>9 F + 3 M</td>
<td></td>
</tr>
<tr>
<td>Vocabulary (max=20)</td>
<td>13.25</td>
<td>2.34</td>
<td>15.00</td>
<td>1.35</td>
</tr>
<tr>
<td>Reading comprehension (max=36)</td>
<td>26.33</td>
<td>7.13</td>
<td>19.92</td>
<td>3.90</td>
</tr>
<tr>
<td>Babble threshold from central loudspeaker (dB SPL)</td>
<td>5.23</td>
<td>1.47</td>
<td>11.68</td>
<td>4.02</td>
</tr>
<tr>
<td>Babble threshold from lateral loudspeakers (dB SPL)</td>
<td>-3.76</td>
<td>2.65</td>
<td>-1.52</td>
<td>2.29</td>
</tr>
<tr>
<td>R-SPIN threshold for real condition (dB)</td>
<td>-1.40</td>
<td>1.62</td>
<td>1.30</td>
<td>1.40</td>
</tr>
<tr>
<td>R-SPIN threshold for virtual condition (dB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**R-SPIN SNR-50 thresholds**

Figure 3 plots the average 50% correct R-SPIN SNR-50 thresholds (dB) as a function of separation condition and group for the younger EL1 listeners (dotted rectangles), the older EL1 listeners (lined rectangles), and the young EL2 listeners (solid rectangles).

The SNR levels required for 50% correct repetition of the last word in low-context sentences was highest when there was no spatial separation, versus when there was a separation between the target sentences and the babble background, and were on average lower for real than for virtual locations. The R-SPIN SNR-50 thresholds also appear to be higher for young EL2 listeners than for older EL1 listeners, who, in turn, had higher thresholds than the younger EL1 listeners. In addition, the advantage due to spatial separation is larger when spatial location is real than when it is virtual. The benefit of separation over no separation on average was larger for the young EL2 listeners (5.18 dB) than both the younger (4.2 dB) and older EL1 listeners (2.18 dB).
Figure 3. Average R-SPIN SNR-50 thresholds (dB) calculated under each of the four spatial conditions. Standard error bars are shown.
A repeated measures ANOVA with two separation conditions (yes/no) and three groups (younger and older EL1 listeners and younger EL2 listeners) as between-subjects factors, and with the two types of spatial location (real/virtual) as a within-subject factor confirmed this description, showing a significant main effect of Separation Condition ($F[1, 66]=97.613, p<.000$), Group ($F[2, 66]= 51.539, p<.000$) and Type of Location (real versus virtual; $F[1, 66]= 30.451, p<.000$) as well as a two-way Group by Separation interaction ($F[2, 66]=3.559, p=.034$) and a Type of Location by Separation interaction ($F[1, 66]=46.367, p<.000$). No other effects were significant. Post hoc tests with Sidak adjustment found that all three groups differed significantly from one another ($p < .005$ for all three pairwise comparisons).

To better illustrate the nature of the Separation by Type of Location interaction, Figure 4 shows the advantage of real versus virtual location cues (R-SPIN SNR-50 threshold virtual – R-SPIN SNR-50 threshold real) for the two types of separation (target sentence and babble separated versus co-located). This figure clearly indicates that the advantage of real over virtual location is larger when the target sentence and babble were perceived to be spatially separated, than when they were perceived to be co-located.
Figure 4. Average advantage of real versus virtual location cues (R-Spin threshold virtual – R-SPIN threshold real) for the two types of separation (target sentences and babble separated versus co-located). Standard error bars are shown.
Figure 5 suggests that the Group by Separation interaction is due to the fact that the benefit due to spatial separation (the difference in R-SPIN SNR-50 thresholds between spatially–separated vs spatially co-located conditions) is smaller for older EL1 listeners than either of the younger groups.

An examination of the group by spatial separation interaction for older versus younger EL1 listeners, found the benefit due to spatial separation to be significantly smaller for older EL1 than for younger EL1 listeners (Group × Separation interaction: $F[1, 44]=6.127, p=.017$). The Group × Separation interaction was also significant when the two groups were older EL1 and young EL2 listeners ($F[1, 44]=4.927, p=.032$), but not when younger EL1 listeners were compared with young EL2 listeners ($F[1,44] < 1$). Figure 5 also suggests that when there is no separation, the R-SPIN SNR-50 thresholds are equivalent for both younger and older EL1 listeners. A separate ANOVA on the no-separation condition revealed a significant effect of group ($F[2, 32] = 38.154, p < .001$) with post hoc tests with Sidak adjustment indicating that the EL2 listeners differed significantly from both EL1 groups ($p < .001$ in both cases) but that the younger and older EL1 listeners did not differ significantly from one another. A comparable analysis of the data from the Separation Condition revealed a significant Group effect ($F[2,33] = 20.842, p < .001$) with the post hoc tests with Sidak adjustment, showing that all three groups differed significantly from one another ($p < .03$ for all pairwise comparisons).
Figure 5. Average FR-SPIN SNR-50 thresholds (dB) calculated under each of the two spatial location conditions for each of the three groups. Standard error bars are shown.
Dialogue comprehension results

The main finding of interest was that once the SNR levels were adjusted based on the R-SPIN results, only small differences were found among the three groups tested in the amount of comprehension questions correctly answered (see Figure 6). In general the younger EL1 listeners seemed to perform slightly better than either the older EL1 listeners or the young EL2 listeners. However, a 2-within-subject (real versus virtual) × 2-between-subject (separation versus no-separation) by 3-between-subject (younger and older EL1 listeners, young EL2 listeners) ANOVA revealed only a significant main effect of Separation ($F[1, 66]=4.671, p=0.034$), with performance being better when the voices were separated rather than co-located.
Figure 6. Average percentage of correctly answered questions calculated under each of the four spatial conditions for each of the three groups. Standard error bars are shown.
The contribution of vocabulary knowledge and reading comprehension to dialogue comprehension

We explored the degree to which the different levels of the factors (Separation, Group and Type of Location) were differentially associated with individual differences in vocabulary knowledge and in reading comprehension skill. First, we examined whether individual differences in vocabulary knowledge were more predictive of the number of dialogue comprehension questions answered correctly when the cues to spatial location were real, as opposed to virtual. Figure 7 relates the percentage of questions correctly answered as a function of vocabulary score separately for the virtual location conditions (upper panel), and the real location conditions (lower panel).

In this figure, the vocabulary scores were first centered within each group of participants to normalize the vocabulary scores across the six groups of participants. Comprehension scores (percentage of questions correctly answered) were also normalized within each of the 12 conditions in the experiment to eliminate the contribution of any residual effects of conditions on performance. Before conducting these regression analyses, we first removed any effect that individual differences in reading comprehension had on performance (see Appendix A). This allowed us to evaluate the effects of individual differences in vocabulary once the effects of reading comprehension on performance had been removed.
Figure 7. Percentage of correctly answered dialogue questions plotted against the individual performance on the Mill Hill vocabulary test after the contribution of reading comprehension to performance had been removed. Both the adjusted number of questions answered and the Mill Hill scores are centered within each group. A least squares regressions line is presented for each of the two types of location.
Figure 7 shows that slope of the line relating percent correct to vocabulary knowledge is considerably higher for virtual location than real location. A regression analysis (see Appendix A), found vocabulary knowledge to be significantly related to dialogue comprehension for the virtual conditions but not for the real location conditions, with the difference in correlation between the two (the interaction between vocabulary and Type of Location) also being significant ($F[1,142] = 4.70, p = .03$). However, similar regression analyses failed to find any evidence that the relationship of vocabulary knowledge to percent correct differed between the no-separation conditions and the separation conditions, or among the three groups of participants (younger EL1 listeners, older EL1 listeners, and young EL2 listeners).

A similar analysis was conducted to examine the contribution of reading comprehension skill to dialogue comprehension. In particular, before evaluating the contribution of reading comprehension to performance, we first moved any effect that individual differences in vocabulary had on performance. The results of this analysis indicated that the contribution of reading comprehension skill to dialogue comprehension did not differ between the virtual and real location conditions, nor did the contribution differ between the no-separation and separation conditions. However, the contribution of reading comprehension skill to performance on the dialogue comprehension task did differ across the three groups. Figure 8 shows that reading comprehension was positively correlated with performance for the younger EL1 listeners only. A regression analysis (see Appendix A) indicated that the strength of the relationship differed among the three groups, with Bonferroni corrected pairwise comparisons confirming that the slopes differed significantly between younger EL1 and young EL2 listeners.
Figure 8. Percentage of correctly answered dialogue questions plotted against the individual performance on the Nelson-Denny reading comprehension test after the contribution of vocabulary to performance had been removed. Both the adjusted number of questions answered and the Nelson Denny scores are centered within each group. A least squares regressions line is presented for each of the three groups.
We also examined whether either the vocabulary scores or the reading comprehension scores were related to R-SPIN in each of the 12 conditions (3 Groups × 2 Spatial Locations × 2 Types of Location). Similar analyses to those conducted for the percentage of correctly answered dialogue questions failed to find any evidence that either vocabulary knowledge or reading comprehension skill could account for individual differences in the R-SPIN threshold values.

**Comparison of current data to previous data**

Two of the conditions in the current study replicate two similar conditions that were found in the Murphy et al. (2006) study. In both the Murphy et al. and the current study, younger and older EL1 listeners were volunteers recruited from the University Community and local neighborhood community, respectively. No EL2 listeners were tested in Murphy et al. The real separation condition in the present study is comparable to the high babble noise condition where the dialogues were presented from loudspeakers at 45° azimuth in Experiment 3 of Murphy et al. and the real no spatial separation condition replicates the condition in which the dialogues were presented once the spatial separation was removed in the last experiment reported (Experiment 4) by Murphy et al. The R-SPIN SNR-50 thresholds and the dialogue comprehension results of these two comparable conditions from both studies were analyzed separately using a Univariate Analysis of Variance with Age, Experiment (Murphy et al., (2006)/ current experiment), and Separation as between-subjects variables. The results of this analysis revealed a significant main effect of Age ($F[7,88]= 4.296, p=.041$) but not of Experiment ($F[1,88]=.963, p=.329$) or Separation ($F[1,88]=.210, p=.648$) on dialogue comprehension performance. In addition there were no significant two- or three-way interactions among these three factors. Hence there is no evidence that the participants in this experiment differed in performance from those in Murphy et al.
A similar analysis which compared the R-SPIN SNR-50 thresholds calculated under the two comparable conditions in the two studies showed a significant effect of Age ($F[1,88]=125.94, p<.001$) and Separation ($F[1,88]=504.45, p<.001$). As was the case for the dialogue comprehension performance, no significant effect of Experiment was found ($F[1,88]=0.089, p=0.89$); however a significant interaction was found between Age and Experiment ($F[1,88]=21.4, p=0.039$). The latter interaction is probably a result of the younger adults in the current study performing slightly worse than the younger adults who participated in Murphy et al. and the older adults in the current study performing slightly better than the older adults who participated in Murphy et al. when there was no spatial separation. Overall, the evidence suggests that the current study successfully replicated the study conducted by Murphy et al. (2006) in regard to the two comparable conditions, and that participants in this study did not differ significantly from the participants in Murphy et al. with respect to their performance in both the dialogue comprehension task and the R-SPIN word recognition task in those conditions, which were comparable across the two experiments.

2.3 Discussion

Using the R-SPIN results as an index of spoken word recognition difficulties

The R-SPIN results indicate that R-SPIN SNR-50 thresholds are lower for real than for virtual location (see Figure 3). This is consistent with the hypothesis that it is more difficult to parse the auditory scene when the location of the sources is virtual rather than real. There is also evidence (see Figure 4) that the advantage of real over virtual location of stimuli is considerably larger in the presence of spatial separation. The R-SPIN results also indicate that young EL2 listeners and the older EL1 listeners find it more difficult to recognize words in babble than
younger EL1 listeners (see Figure 3). Moreover, the results show that older adults benefit less from spatial separation than EL1 and EL2 younger listeners (see Figure 5).

The results also show that when the target sentence and babble appear to be co-located, younger and older EL1 listeners have similar R-SPIN SNR-50 thresholds that are significantly lower than those of the EL2 listeners. However when the target sentences and the babble are perceived to be separated, younger EL1 listeners have significantly lower thresholds than the older EL1 listeners, who, in turn, have lower thresholds than the young EL2 listeners. This suggests that when target and masker are co-located, normal-hearing older and younger EL1 listeners do not differ with respect to spoken word recognition in a background of babble, but that both groups have lower speech-recognition thresholds than do young EL2 listeners. However, spoken word recognition in the presence of spatial separation is better for younger than for older EL1 listeners.

Overall, the results of the R-SPIN word recognition task show that when there are cues to spatially separate the target from the masker, word recognition in older EL1 listeners and young EL2 listeners is inferior to that of younger EL1 listeners. When listeners of these two groups attempt to communicate in real-life situations where sound sources are often spatially separated, they will experience greater difficulty with respect to word recognition. In older EL1 listeners, this increased difficulty most likely is due to the reduction in the quality of the bottom-up information leading to word recognition. In young EL2 listeners, difficulties in word recognition are likely due to the increased difficulty they experience in segregating language streams in their L2 (Cooke et al., 2008; Ezzatian et al., 2010).
**Dialogue comprehension results**

The dialogue comprehension results demonstrated that when R-SPIN SNR-50 thresholds were used to adjust for individual differences in the ability to recognize words without supportive context, no effects due to Type of Location (real versus virtual), or Group (younger EL1 listeners, older EL1 listeners, young EL2 listeners) were found. However, the main effect of Separation was significant even though the SNRs were adjusted based on the R-SPIN SNR-50 thresholds. Listeners on average performed significantly better under spatial separation conditions than when sources were co-located. This suggests that spatial separation (real or perceived) facilitates the comprehension and retention of information obtained from the dialogues even when lexical access is presumably equated across the co-located and spatially separated conditions using the R-SPIN adjustment procedure. For example, spatial separation between the talkers may facilitate the association of the incoming information with a specific talker as well as facilitate its retention. However, it is not possible to determine from the present data the precise mechanisms responsible for this spatial separation effect. One possible reason why the R-SPIN adjustment in the spatial separation condition did not eliminate the Separation effect in the dialogues might be that R-SPIN SNR-50 thresholds were determined for a single voice emanating from the right of the listener with the babble occupying a central location. In the separation condition for the dialogues, however, a voice could be on the left or right depending on who was speaking, with the babble emanating from the center. Hence, spatial separation in the dialogue condition could have facilitated switching attention back and forth from the right to left depending on who was speaking. The fact that the R-SPIN test did not require the listener to switch attention from one side to the other may explain why it was not successful in eliminating the Separation effect in the dialogue portion of this study.
The results from the younger and older EL1 adults in the real location conditions of the present study were found to be consistent with the results from the equivalent conditions in the Murphy et al. (2006) study. The age difference found when we combined the real location conditions of these two studies supports the hypothesis that older adults might not be as good as younger adult at using the full range of interaural cues to either obtain or maintain stream segregation when sources are separated in space. Older adults frequently need to communicate in multi-talker daily situations taking place in a noisy environment, and naturally they have to do so without any SNR adjustments to accommodate for any individual age-related changes in hearing. The results described here emphasize the notion that in addition to age related difficulties in word recognition, older adults might have a limited toolbox of acoustic cues to assist them when attempting to meet a speech comprehension challenge in real-life situations (e.g., listening to a movie in a surround sound environment). This reduction in the ability to benefit from the acoustic cues provided by physical separation among sound sources is likely to have even greater implications in the presence of a hearing impairment (Neher et al., 2009).

It is interesting to note that the results of the current study indicate that adjusting for spoken word recognition eliminates speech comprehension differences between young EL1s and EL2s. This suggests that the difficulties EL2s have in noisy situations are mainly due to inferior spoken word recognition rather than to a reduced ability to extract meaning from the spoken message in L2. Hence, once lexical access is achieved there appear to be minimal differences in speech comprehension.

**The relationship between age, linguistic competence, and speech comprehension**

Given that the reasons for the spoken word recognition difficulties are likely to differ across the three groups, we might expect different compensatory processes to be engaged by
each of the three groups. To examine this, we took the two measures of linguistic competence and looked to see the extent to which those measures were correlated with performance under each of the acoustic settings used in the current study. This examination, which was done for both the R-SPIN results as well as for the dialogue comprehension results separately, can be used to help identify the relative importance of bottom-up and top-down influences on speech comprehension, and how the pattern of interactions among these factors is modulated by the nature of the acoustic scene.

Specifically, we looked at data from both the real and virtual locations at each of the 6 groups (12 younger EL1 in the separation condition; 12 younger EL1 listeners in the no-separation condition; 12 young EL2 listeners in the separation condition; 12 young EL2 listeners in the no-separation condition; 12 older EL1 listeners in the separation condition; and 12 older EL1 listeners in the no-separation condition) to determine the contribution of the two linguistic measures to performance (number of dialogue comprehension questions correctly answered) within each group and condition (see Appendix A for further details). This analysis showed that both the vocabulary and the reading comprehension tests results were related to the average number of comprehension questions correctly answered ($r = .32, p = .007, \text{ and } r = .38, p < .001$, for reading comprehension skill and vocabulary knowledge, respectively), but that these slopes did not differ across groups. More interestingly, the results of the analysis indicated that there was a significant correlation between vocabulary knowledge and the number of dialogue questions correctly answered under the virtual location conditions, but not under the real spatial location conditions (see Figure 7), and that this interaction between the Type of Location and Vocabulary was significant (the slopes of the lines differed significantly from each other).

We would like to suggest a hypothesis that could explain this finding. An early stage in speech comprehension involves obtaining lexical access to the meaning of words. It has been
shown that both bottom-up and top-down processes are involved in word recognition. We can hypothesize that the degree to which top-down processes are engaged in lexical access is modulated by acoustic factors. When listening is relatively easy, we might expect successful lexical access with minimal assistance from top-down processes. However, when listening becomes difficult, we might expect that the top-down processes involved with lexical access to be more fully engaged. When sound sources are located virtually in space using the precedence effect, they give the impression that their location is diffuse, and there are fewer acoustic cues that can be used to segregate the different acoustic streams than when sound source location is real (Freyman et al., 1999; Li et al., 2004). Therefore it is reasonable to expect that under such conditions, obtaining and maintaining stream segregation will be more demanding, and it is possible that the bottom-up processes involved in lexical access will be less reliable because of occasional intrusions from the competing streams. The R-SPIN results indicate that word-recognition is indeed more difficult under virtual as opposed to real location conditions (on average, 50% thresholds are 1.56 dB higher for virtual than for real location conditions), and might require a greater engagement of top-down lexical processes in order to maintain word-recognition accuracy. Hence we would expect that the relative contribution of top-down processes to lexical access to be greater for virtual than for real spatial location conditions. When there is relatively little need to draw on top-down, knowledge-driven processes to achieve lexical access, we would expect a small or negligible contribution of individual differences in the efficacy of these processes to performance. However, when the draw on top-down processes is heavy, then we would expect that some of the variance in performance to be accounted for by individual differences in the efficacy of these processes. Hence we might expect that the contribution of vocabulary knowledge to dialogue comprehension performance would increase with level of listening difficulty, as it appears to do in this experiment.
It is also interesting to speculate on the reasons why the relationship between reading comprehension skill and dialogue comprehension performance did not differ between real and virtual location conditions (see Appendix A). One possibility is that the linguistic and cognitive skills tapped by the reading comprehension measure are separate from those involved in lexical access. Bottom-up lexical access in these experiments is obtained exclusively through the auditory channel. We can hypothesize that reading comprehension taps higher-order processes that are modality independent and are related to the integration of information, extraction of themes, etc., and therefore are unlikely to be as affected by parameters of the acoustic scene such as whether the location type is real or virtual. On the other hand, we did find group differences with respect to the relationship between reading comprehension and the number of dialogue questions answered correctly (Figure 8). Specifically reading comprehension was found to be positively and significantly correlated with performance only in younger El1 listeners.

The fact that, in young El1 listeners, individual differences in reading comprehension skills account for a significant portion of the variance in the number of dialogue questions correctly answered indicates that there are higher-order, modality-independent skills that contribute to both reading and listening comprehension in this population. The lack of correlation in older El1 listeners suggests that listening comprehension in difficult listening situations depends on a different set of modality-specific processes that are engaged to compensate for the loss of fidelity in the auditory processing system. In other words, in younger El1 listeners, listening comprehension is akin to reading comprehension once lexical access has been achieved. In older El1 listeners and in young El2 listeners, the same degree of listening comprehension (the number of dialogue questions correctly answered did not differ across groups) appears to be achieved in a different way. This is consistent with the notion that different neural circuitry
supports speech comprehension in different populations (e.g. Harris et al., 2009; Wong et al., 2009).

There may be different reasons why reading comprehension appears to contribute little to individual differences in performance in the older EL1 listeners and young EL2 listeners. The Nelson-Denny reading comprehension test was developed and standardized for younger EL1 individuals and might not be as valid when testing other populations such as older adults or EL2 speakers. Because the Nelson-Denny is a time-limited task, it might not adequately reflect individual differences in reading comprehension in older adults whose reading speed is substantially slower than that of younger adults (Rodríguez-Aranda, 2003). It is also unlikely to be a good measure of individual differences in reading competence in young EL2 adults either because they too are likely slower readers, or because this test does not adequately gauge the linguistic and cognitive skills used by EL2 speakers in comprehending written language. In addition, the draw on attentional resources in young EL2 speakers may be higher than in younger and older EL1 speakers because lexical access in L2 speakers most likely requires the activation and integration of information from both their L1 and L2 lexicons (Kroll & Steward, 1994). Moreover, the execution of the higher-order tasks involved in listening comprehension by L2 listeners, such as extracting themes, integrating information with past knowledge, and storing this information for future use, may partially be executed in their L1. With respect to older adults, a number of cognitive aging theorists hypothesize that they have a more limited pool of attentional resources than do younger adults (Craik & Byrd, 1982). Alternatively, age-related changes in hearing may place a greater demand on attentional resources in older than in younger adults. Either or both of these factors would result in a greater degree of attentional focus within the auditory domain in older L1 listeners compared with younger L1 listeners. As a result,
speech comprehension in older adults may depend more on processes that are specific to the auditory modality when listening becomes difficult.

To determine whether the failure to find a relationship between reading comprehension and performance in older EL1 listeners when listening becomes difficult reflects an increased dependence on modality-specific processes in listening comprehension tasks, we examined the contribution of the reading comprehension performance to the number of correctly answered questions when older EL1 listeners were asked to perform a similar task under less demanding perceptual conditions. As previously mentioned, in a study conducted by Murphy et al. (2006), both younger and older listeners were asked to listen to the same two-talker conversations under different acoustic settings, one of which was in quiet. We used Murphy et al.’s data to test the contribution of vocabulary and reading comprehension to the dialogue comprehension performance in quiet (see Appendix B for a detailed description of the analysis conducted) and found that the least squares regressions of the adjusted percentage correct scores against reading comprehension for both the younger and older participants were highly significant (see Figure 9). Hence, under perceptually easy listening conditions, reading comprehension is as strongly related to performance in older EL1 listeners as it is in younger EL1 listeners. The results of this analysis is consistent with the hypothesis that the lack of a significant contribution of reading comprehension to performance in older adults in noise reflects an increased dependence on modality specific processes when listening becomes difficult.

**Toward a general model of resource allocation in speech comprehension**

The differential contribution of vocabulary to dialogue performance under virtual versus real location conditions suggests that difficult listening conditions require that attentional resources be deployed in aid of scene analysis and word recognition. In addition, the relative
weight given to bottom-up and top-down processes contributing to lexical access may be shifted in favor of top-down influences when listening becomes difficult. Previous theories such as the Ease of Language Understanding Model (ELU) have proposed that lexical access is impeded or slowed when listening becomes difficult (Rönnberg et al., 2013). Hence, more attentional and working memory resources are required to support lexical access. The current results suggest that the demand on such resources is modulated by the nature of the acoustic scene, which, in turn, affects the engagement of the more central, modality independent cognitive resources involved in language comprehension. Let us assume that the virtual location conditions require additional attentional resources be deployed to locate the diffused sources in space, depleting the pool of the resources available for phoneme identification and bottom-up lexical access. The notion here is that with real spatial location, the task of locating the stimuli is easy whereas virtual localization requires a larger amount of attentional processing. Now consider the problem facing the executive. When full auditory attentional resources can be devoted to lexical access and the bottom-up acoustic information is reliable and sufficient, the executive will trust the output from bottom-up lexical processing, and give less weight to top-down, knowledge-driven factors such as vocabulary knowledge. However, when attentional resources are required to locate the sound sources, this additional burden reduces the reliability of the information produced through bottom-up lexical processes. In that case, the executive may place more weight on the top-down processes involved in lexical access to compensate for the missing or corrupted bottom-up information.

The hypothesis presented here suggests that, depending on the acoustic scene, the listening strategy may change the relative engagement of the different processes involved in speech comprehension. The idea that listeners may systematically downplay the contribution of acoustic detail and increase their reliance on lexical-semantic knowledge has been previously
suggested by Mattys et al. (2009, 2010), and Mattys and Wiget (2011). Mattys et al. (2009,2010) and Mattys and Wiget (2011) demonstrated a shift that they refer to as a cognitive-load-induced “lexical drift” in cases of high cognitive load due to an additional secondary task even when no actual energetic masking or additional auditory information is involved (e.g. a secondary visual task). For example, Mattys and Wiget (2011) measured the magnitude of lexical bias on phoneme identification under cognitive load (CL) and no CL by adding a secondary visual search task to increase CL. Their results suggested that the CL interferes with detailed phonetic analysis that leaves the listener with impoverished encoding of the auditory input and a greater need to rely on lexical knowledge in order to compensate for the missing information. The collaborative evidence provided by Mattys et al. (2009, 2010) as well as Mattys and Wiget (2011) and the current study support the existence of a dynamic rather than stationary processing strategy that changes depending on the listening situation, as well as the age and linguistic status of the listener.

Individual differences in top-down lexical knowledge, as indexed by the Mill-Hill vocabulary test, may be expected to account for a greater proportion of the variance in speech comprehension when the accuracy of the bottom-up processes involved in lexical access is compromised by listening difficulty or by age. Hence we would not expect to find Mill-Hill scores to be related to comprehension in good listening conditions with only a single talker. However, as the listening situation becomes more complex and harder to analyze (competing sound sources, diffused virtual locations rather than compact coherent ones, etc.), the more likely it is that top-down lexical processes will be engaged, and individual differences in speech comprehension to be related to measures of top-down lexical processing. Note also that the complexity of the listening situation need not affect processes subsequent to word recognition. Hence measures indexing the contribution of higher-order processes involved in language
comprehension (e.g., integration of information across talkers and with stored knowledge) might not be affected by the acoustic parameters of the auditory scene.

The behavioural evidence that the nature of the acoustic scene and the age and linguistic competence of the listener modify the engagement of different auditory and cognitive processes involved in speech comprehension is consistent with recent findings from brain-imaging studies. These studies demonstrate that the degree to which the different neural networks involved in speech comprehension are activated, is modulated by the degree of stimulus complexity, type of task, and age. Previous neuroimaging studies which attempted to map the brain areas involved in speech recognition and comprehension demonstrated a frontal-temporal network in which temporal regions subserve bottom-up processes, whereas frontal regions subserve top-down processes (Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006). This network seems to be differentially activated depending on the nature of the auditory stimuli and the complexity of the task (Benson et al., 2001; Zekveld et al., 2006). In addition, neural activation seems to not only be affected by the characteristics of the stimuli and task, but also by the characteristics of the listeners as well. Harris et al. (2009) examined the performance of both younger and older adults on word recognition task in which the intelligibility of the stimuli was manipulated using low-pass filtering. Their results showed no age differences in the auditory cortex but differences were found in the anterior cingulate cortex which is presumed to be associated with attention. Age-related differences were also found in the Wong et al. (2009) study in which younger and older adults were asked to identified single words in quiet and in two multi-talker babble noise conditions (SNR=20, -5). The fMRI results for older adults showed reduced activation in the auditory cortex but increased activation in the prefrontal and precuneus regions that are associated with working memory and attention. The increased cortical activities in the general cognitive regions were positively correlated with the behavioral results in the older adults. Wong
et al. interpreted this correlation, as well as a more diffused activation involving frontal and ventral brain found in the older adults, as an indication of a possible compensatory strategy used in older age. These studies provide evidence for possible age-related changes in the involvement of the different brain regions engaged in speech recognition in noise. As Scott and McGettigan (2013) note in their recent review of the neural processing of masked speech, “Further outstanding challenges will be to identify cortical signatures that are masker specific and that might be recruited for both energetic/modulation masking and informational masking, ...and address the ways that aging affects the perception of masked speech while controlling for intelligibility (p. 65).”

In general then, we would expect that the auditory and cognitive processes that are engaged in speech comprehension to be modulated by a number of factors including but not limited to: 1) the complexity of the auditory scene, 2) the nature of the speech material, 3) the task demands placed on the individual, and 4) individual differences in the auditory, linguistic, and cognitive skills and knowledge available to the listener. In Experiment 2, we examine the effects of spatial separation in a more complex listening task, involving listening to a conversation conducted by three talkers who were either co-located or spatially separated.

3 Experiment 2

To further explore how language competence, age, and listening difficulty might modify how comprehension is achieved, we had participants listen to three-person conversations in different levels of background babble, with and without spatially separating the talkers. When a conversation is restricted to two talkers, the two talkers will most likely engage in a simple
alternating, turn taking routine so that who will talk next is highly predictable. The presence of three talkers makes turn taking much less predictable. Having to switch attention from one talker to another in the presence of uncertainty regarding which of the three talkers will speak next requires a quicker and more efficient switching in order to avoid missing out the information at the beginning of each utterance. In other words, when listening to two talkers, the listener can start the attention switching process once one of the talkers is done talking. However, when more than two talkers are participating in the conversation, the listener has to either wait until the next talker starts talking to identify which of the talkers he or she needs to focus on next, or infer who will be talking next from context. In addition, when listening to three as opposed to two talkers, listeners have to identify and track an additional talker, allocate incoming information correctly to one of three talkers participating in the conversation, and separately store the information accordingly for later use. Successfully differentiating among three talkers and allocating information to its correct source, will require a greater degree of executive control than when only two talkers are present. A recent study suggests that older adults take longer to recover when attention is misdirected to the wrong source than do younger adults (Singh et al., 2013). Thus, the greater memory and attentional demands that trialogues present may either reveal group differences that were unnoticeable in previous, less challenging tasks, or further confirm that the ability to focus and switch attention, as well as remember the heard information, does not deteriorate with age, or with linguistic status, as long as all individuals are equated with respect to their ability to recognize individual words in babble.
3.1 Method

Participants

The participants were 12 younger normal hearing listeners whose first language was English (Young EL1; mean age: 20.98 years; \(SD: 1.78\)), 12 older listeners with normal hearing for their age, whose first language was English (Older EL1; mean age: 69.82 years; \(SD: 3.88\)), and 11 young normal hearing listeners for whom English is their second language\(^2\) (Young EL2; mean age: 22.53 years; \(SD: 1.92\)). EL1 listeners were all born and raised in a country in which the primary language was English and were not fluent in any other language at the time of participation. EL2 listeners were those who first became immersed in an English speaking environment after the age of 14. The EL2 listeners were from a range of language backgrounds, specifically, Hebrew (n=1), Korean (n = 1), Mandarin (n = 4), Cantonese (n=1), Iranian (n = 1), Russian (n = 1), Lithuanian (n = 1), Filipino (n = 1). The younger participants were volunteers recruited from the students and staff at the University of Toronto Mississauga. The older participants were volunteers from the local community. Those who participated in Experiment 1, were exploded from participation in the current experiment. All participants were asked to complete a questionnaire regarding their general health, hearing, vision, and cognitive status. Only participants who reported that they were in good health and had no history of serious pathology (e.g., stroke, head injury, neurological disease, seizures, and the like) were included. None of the participants had any history of hearing disorders, and none used hearing aids. All participants reported having either normal or corrected vision.

\(^2\) One young EL2 subject was removed from this analysis because his or her data for the high noise condition versus the quiet conditions differed by 5 standard deviation units from the mean performance of the remaining participants. See Method.
Materials, Apparatus and Procedure

Audiometric thresholds, Nelson-Denny reading comprehension skill, and the Mill Hill vocabulary knowledge were measured during each participant’s first session. The trialogues, along with the babble thresholds and the low-context Revised Speech Perception in Noise (R-SPIN) thresholds (Bilger et al., 1984) were administered over the next two experimental sessions (3 trialogues per session). Sessions were typically 1.5-2 hours in duration and were completed within 2 weeks. All participants were paid $10/hr for their participation.

Hearing Measures

Audiometric Testing. Pure-tone air-conduction thresholds were measured at nine frequencies (0.25-8kHz) for both ears using an Interacoustics Model AC5 audiometer (Interacoustic, Assens, Denmark), and participants were required to meet the same hearing requirements as in Experiment 1. The average audiograms for the three groups of participants are shown for the right and the left ears in Figure 9. The two groups of younger adults had equivalent hearing levels at all frequencies. Hearing levels for older adults were about 6.57 dB poorer than those of the younger adults at frequencies ≤ 3 kHz, with the younger-older difference increasing as a function of frequency for frequencies > 3 kHz.

Babble Threshold Test. Babble thresholds were determined following the method described in Experiment 1. In the current study babble thresholds were determined for the case where the babble was presented simultaneously over three loudspeakers (two positioned laterally 45° to the right and left of the listener a central loudspeaker located in front of the participant). The babble threshold was used to individually adjust the intensity of the speech signal to a sensation level, SL, of 39 dB (speech signal level = individual babble threshold + 39dB).
Figure 9. Average audiograms for the three groups of participants are shown for the right and the left ears.
The same R-SPIN test employed in Experiment 1 was also administered to the participants in this study. Three R-SPIN SNR-50 thresholds were estimated for each participant: one for the R-SPIN sentences played over the left loudspeaker, one for the central loudspeaker, and one for the right loudspeaker. The R-SPIN threshold obtained from the central loudspeaker was used to adjust the SNR when the three trialogue voices were co-located and presented over the central loudspeaker. The average of the left- and right- R-SPIN SNR-50 thresholds was used to adjust the SNR when the three trialogue voices were played over separate loudspeakers. The R-SPIN threshold estimates were rounded to units of 1 dB before being integrated into the calculation of the individually adjusted SNR that was used for presentation of the trialogues. If the estimated R-SPIN threshold was calculated to be exactly half way between two integer values, rounding was conducted in the direction of alleviating the SNR difficulty (e.g. 3.5 dB SNR was rounded to 4dB SNR and -3.5 dB SNR to -3 dB SNR).

**Language Proficiency Measures**

As in Experiment 1, participants were asked to complete both the Mill Hill vocabulary test (Raven, 1965), and the Nelson-Denny test of reading comprehension (Brown et al., 1981).

**Triologue Comprehension Task.**

Each participant was asked to listen to the six trialogues presented in babble noise in a sound-attenuating chamber. At the end of each trialogue, which was on average 11.58 min (SD = 3.2) long, the participant was presented with a series of 10 multiple-choice questions regarding the contents of that trialogue. There were two spatial conditions in the experiment: (a) No-spatial separation with all three talkers presented over the central loudspeaker only; (b) Spatial separation with the voice of one of the talkers presented over the loudspeaker placed 45° to the
left of the listener, another voice presented over the loudspeaker placed 45° to the right of the
listener, and another voice presented from the central loudspeaker placed in front of the listener.
The multi-talker babble was played over all three loudspeakers in both the no-separation and
separation conditions. In addition to the two spatial location conditions, there were three masking
level conditions: (a) A quiet condition in which no multi-talker babble was played; (b) Moderate
masking level, in which the trialogues were presented in babble at an SNR that was individually
adjusted for each participant and condition based on his or her babble threshold and R-SPIN
results according to the following calculations: Trialogues were presented at babble threshold +
39 dB; Babble was presented at babble threshold + 39 dB – R-SPIN threshold + 15 dB; (c) High
masking level, in which trialogues intensity was the same as in the previous masking condition,
but the babble thresholds were individually adjusted according to the following calculation:
babble threshold + 39 dB – R-SPIN threshold + 21 dB. In this design, both the spatial location
and the masking level were within-subject factors; each participant was tested on each of the six
different conditions (2 spatial conditions × 3 masking levels), using one trialogue per condition.
The six trialogues were administered over two sessions, three trialogues per session. At the end
of each trialogue, participants were asked to answer a set of 10 multiple-choice questions with
four alternatives. Each question referred to a specific item of information that was mentioned
explicitly only once during the trialogue (see trialogue comprehension materials for more
information).

Trialogue comprehension materials.

The trialogues were edited versions of six published one-act plays: (a) The Battle of the
Bull Run Always Makes Me Cry by C. Real (1997); (b) Heads by J. Jory (2001); (c) Fourteen by
A. Gerstenberg (1920); (d) The Worst Possible Time for Writer's Block by J. Shanahan (2008);
Three Sons by D. Trasler (2004); The Bond They Shared by D. Pinsof (2007). The plays were selected and edited to meet the following criteria: (a) each had only three characters; (b) all three characters contributed to the conversation throughout the play; (c) in three of the plays, all three characters were male, and in the other three, all three characters were female; (d) complete comprehension of the trialogues could be achieved from the spoken words and suprasegmental information (e.g. intonation) of the three characters alone; that is, comprehension did not depend on any visual or other non-verbal cues or actions. See Appendix C for an excerpt from Three Sons, a play in which three men who are waiting for a connecting flight, start a conversation and find they have something in common.

For each trialogue, a set of 10 multiple-choice questions was constructed. Each question had four alternatives and tested for specific information that been mentioned explicitly only once during the course of the trialogue. The set of 10 questions probed for information that was distributed equally over the course of the entire trialogue. The extent to which the three talkers' contributions were probed was roughly proportional to their relative contributions to the trialogue. See Appendix C for four of the questions asked regarding the content of Three Sons.

Digital recordings were made of three male actors reading the three trialogues with male parts and three female actors reading the trialogues with female parts. Actors were situated in adjacent double-walled sound attenuating chambers whose background noise was < 20 dBA. Each actor had the script presented to him or her on a screen. The actors were wearing headphones and using a microphone that allowed them to hear one another and speak to one another, but they could not see each other. The microphone was situated approximately 30 cm away from each actor’s mouth. Actors were instructed to speak in as natural way as possible, keeping pauses between turns to a minimum and overlapping or interrupting each other as appropriate in a
conversation. The voices of each actor were sent to separate channels of a Tucker-Davis analog-to-digital converter, where they were digitized at a rate of 20 kHz and stored in time-linked digital files. Later, the three channels of the recorded trialogues were edited to clean up any noise or problems in the actors' rendering of their parts. Mean playing time for the six trialogues was 11.58 min ($SD = 3.2$). The average SPL for each of the trialogues was evaluated in a similar procedure to that used for the dialogues. During the experimental session, the three files were converted back to analog form and played simultaneously over the central loudspeaker only (no-separation condition) or each file over one of three loudspeakers (separation condition). Programmable attenuators were used to set the SPL's of the actors' voices over each channel. The 12-talker babble stimulus was generated on a second computer, converted from digital to analog using a separate Tucker-Davis digital-to-analog converter, and then presented to the participant through all three loudspeakers (for both separation and no separation conditions).

### 3.2 Results

Table 2 presents the gender breakdown, mean age, educational level, Mill Hill test of vocabulary knowledge and Nelson-Denny test of reading comprehension results for each age group. The column on the right indicates which of the measures and values differ significantly among the groups.
Table 2

Demographic information (Mean Age and Years of Education, Gender Distribution, Mean Vocabulary and Reading Comprehension Scores) and Mean Babble and R-SPIN SNR-50 thresholds for the participants divided into the three groups tested. Statistically significant differences between groups are noted.

<table>
<thead>
<tr>
<th></th>
<th>Younger EL1</th>
<th>Younger EL2</th>
<th>Older EL1</th>
<th>p &lt; .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>20.98</td>
<td>22.52</td>
<td>69.82</td>
<td>A, C</td>
</tr>
<tr>
<td>Education</td>
<td>15.42</td>
<td>15.09</td>
<td>16.83</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>7 F + 5 M</td>
<td>7 F + 4 M</td>
<td>6 F + 6 M</td>
<td></td>
</tr>
<tr>
<td>Vocabulary (max=20)</td>
<td>13.17</td>
<td>10.45</td>
<td>15.50</td>
<td>A, B, C</td>
</tr>
<tr>
<td>Reading comprehension (max=36)</td>
<td>26.17</td>
<td>18.45</td>
<td>23.00</td>
<td>B</td>
</tr>
<tr>
<td>Babble threshold (dB SPL)</td>
<td>13.54</td>
<td>14.78</td>
<td>19.00</td>
<td>A, C</td>
</tr>
<tr>
<td>Central R-SPIN threshold (dB)</td>
<td>-1.51</td>
<td>1.02</td>
<td>0.53</td>
<td>B</td>
</tr>
<tr>
<td>Lateral R-SPIN threshold (dB)</td>
<td>-4.68</td>
<td>-1.05</td>
<td>-2.58</td>
<td>A, B, C</td>
</tr>
</tbody>
</table>

Note. Max=maximum; dB= decibel; SPL= sound pressure level; R-SPIN= Revised Speech Perception in Noise.
A= A statistically significant difference between Younger EL1 and Older EL1 listeners.
B= A statistically significant difference between Younger EL1 and Young EL2 listeners.
C= A statistically significant difference between Older EL1 and Young EL2 listeners.
As expected, the average age of the older EL1 listeners was significantly higher than that of the two younger groups, while the age of the two younger groups was similar. The number of years of education was similar among the three groups. The older EL1 listeners had significantly higher vocabulary scores than did the younger EL1 listeners, who in turn had significantly higher scores than their EL2 counterparts. The younger EL1 listeners had significantly higher reading comprehension scores than did the other two groups. The babble detection thresholds in the older EL1 group were significantly higher than those in the two younger groups.

**Thresholds for Recognizing Individual Words in Low-Context Sentences**

Figure 10 plots the average 50% correct R-SPIN SNR-50 thresholds (dB) as a function of Location condition and Group for the younger EL1 listeners (solid bars), the older EL1 listeners (dotted bars), and the young EL2 listeners (lined bars). This figure indicates that the SNR levels required for 50% correct repetition of the last word in low-context sentences were highest when the location of the target was central (average = 0.013 dB) versus when the location was lateral (average = -2.77 dB). When the location was either lateral or central, the overall R-SPIN SNR-50 thresholds were lower for younger EL1 listeners (average = -3.1 dB) than for older EL1 listeners (average = -1.0 dB), who, in turn, had lower thresholds than the young EL2 listeners (average = 0.02 dB).
Figure 10. Average R-SPIN SNR-50 thresholds (dB) calculated under each of the two location conditions for each of the three groups. Standard error bars are shown.
A repeated measures ANOVA with the three groups (younger and older EL1 listeners and young EL2 listeners) as a between-subjects factor, and location (central/lateral) as a within-subject factor confirmed this description, showing a significant main effect of both Location condition ($F[1, 32]=59.822, p<.001$) and Group ($F[2, 32]= 6.551, p=.004$). The interaction between Group and Location was not significant. Post hoc tests with LSD adjustment found that younger EL1 listeners significantly differ from both young EL2 listeners ($p=0.001$) and older EL1 listeners ($p=0.026$).

**Trialogue Comprehension Results.**

Figure 11 plots the average percentage of questions correctly answered as a function of Babble Level, Separation Condition, and Group. Aside from the fact that performance decreased as a function of babble level, the percentage of questions correctly answered does not appear to differ with respect to spatial separation or among the three groups of participants.

A three-factor ANOVA of the percentage of questions correctly answered with Group (younger and older EL1 listeners, young EL2 listeners) as a between-subjects factor, and Spatial Separation (separation versus no-separation) and Babble Level (quiet/ moderate/ high) as within-subject factors, found a significant main effect of Babble Level ($F[2, 64]=73.745, p<0.001$), but neither Group nor Separation condition were significant. In addition, none of the interactions among these three factors were significant.
Figure 11. Average percentage of correctly answered questions calculated under each of the three babble level conditions and the two separation conditions for each of the three groups. Standard error bars are shown.
Paired-samples t-tests were conducted to compare the number of correctly answered questions between the different Babble levels. After applying a Bonferroni correction, all paired comparisons showed significant differences: Quiet versus Moderate Babble: \( t(34)=6.468, p<0.001 \); Moderate versus High Babble level: \( t(34)=4.968, p<0.001 \); and Quiet versus High Babble: \( t(34)=14.496, p<0.001 \). Hence performance was best in quiet (77.6% correct), declined when a moderate amount of babble was introduced (57% correct), and was the lowest when the babble background was high (43.3% correct), but still significantly greater (\( t(35) = 9.406, p < 0.001 \)) than chance performance (25% correct).

**The Contribution of Vocabulary Knowledge and Reading Comprehension Skill to Trialogue Comprehension.**

We explored whether vocabulary and/or reading comprehension were related to the average percentage of trialogue questions correctly answered by each of the 35 participants in the experiment. First we centered the reading comprehension scores, the vocabulary scores, and the average percentage of questions correctly answered within each of the three groups to correct for level differences in performance for the different groups and conditions (12 younger EL1 listeners, 12 older EL1 listeners, and 11 young EL2 listeners). A regression analysis found significant correlations between vocabulary and listening comprehension (\( r = .39, p < .001 \)) as well as reading comprehension and listening comprehension (\( r = .45, p < .001 \)). In order to investigate whether the relationship between the two language measures and performance was altered by Babble Level, Separation Condition or Group, the percent correct scores within each of the 18 conditions (2-Separations \( \times \) 3 Groups \( \times \) 3 Noise Levels) were first centered and different slopes were fitted for each of the 18 combinations. The null hypothesis that the relationship of these two variables (vocabulary and reading) to listening comprehension was the
same for all 18 conditions was rejected ($F[34, 174] = 1.658, p = .019$). A further test in which slopes were allowed to differ across groups but were restricted to be the same for each combination of Separation and Babble Level could not be rejected ($F[30, 174] = 1.149, p = .284$). Hence, Figure 12 plots the relationships between listening comprehension and the two language proficiency measures separately for each group: (Young EL1s, Young EL2s, and Older EL1s).

In this figure, the vocabulary scores and the reading comprehension scores were first centered within each group of participants to normalize the scores across the three groups of participants. Before evaluating the relationship between listening comprehension and vocabulary, we first removed any effect that individual differences in reading comprehension had on performance. In evaluating the relationship between listening comprehension and reading comprehension, we first removed the contribution of vocabulary to performance (see Appendix D). This allowed us to evaluate the effects of individual differences in one of the two measures of English proficiency once the effects of the other on performance had been removed. Figure 12 indicates reading comprehension contributes significantly to performance in both young EL1s and young EL2s. However, vocabulary is not significantly related to performance in young EL1’s, but is marginally related to performance in young EL2s. Older EL1s show a reversed pattern to that found in young EL1s in that vocabulary contributes significantly to performance, whereas reading comprehension does not.
Figure 12. Percentage of correctly answered dialogue questions for each of the three groups is plotted against the individual performance on the Mill Hill vocabulary test after the contribution of reading comprehension to performance had been removed (Left panel) and against the individual performance on reading comprehension after the contribution of the Mill Hill vocabulary test to performance had been removed (Right panel). The adjusted number of questions answered as well as the Mill Hill and the reading comprehension scores are centered within each group. A least squares regressions line is presented for each of the three Groups.
A further test in which slopes were allowed to differ across Babble Levels but were restricted to be the same for each combination of Separation and Group could not be rejected ($F[30, 174] = 1.368, p = .110$). Hence, a similar analysis was conducted to evaluate the degree to which the relative contribution of reading and vocabulary to performance might differ within each noise level (see Appendix D). Figure 13 plots the relationships between listening comprehension and the two language proficiency measures separately for each Babble Level: (Quiet, Moderate, High). In this figure, the vocabulary scores and the reading comprehension scores were first centered within each group of participants to normalize the scores across the three groups of participants. Before evaluating the relationship between listening comprehension and vocabulary, any effect that individual differences in reading comprehension had on performance was removed. In evaluating the relationship between listening comprehension and reading comprehension, the contribution of vocabulary to performance was removed (see Appendix D).

Figure 13 indicates that vocabulary contributes significantly to performance at moderate babble level only, whereas, reading comprehension contributes significantly to performance at all three babble levels. The null hypothesis that all three reading comprehension slopes are similar could not be rejected ($F[2,207] = 2.870, p = .059$). Hence the manner in which reading comprehension contributed to listening comprehension did not change with listening conditions. We also explored whether vocabulary and/or reading comprehension was related to the R-SPIN SNR-50 thresholds calculated for each of the 35 participants in the experiment; neither the central R-SPIN values nor the lateral R-SPIN values was significantly correlated with reading comprehension or vocabulary.
Figure 13. Percentage of correctly answered dialogue questions under each of three Babble Levels is plotted against the individual performance on the Mill Hill vocabulary test after the contribution of reading comprehension to performance had been removed (Left panel) and against the individual performance on reading comprehension after the contribution of the Mill Hill vocabulary test to performance had been removed (Right panel). The adjusted number of questions answered as well as the Mill Hill and the reading comprehension scores are centered within each group. A least squares regressions line is presented for each of the three babble levels.
3.3 Discussion

The R-SPIN Results as an Index of Spoken Word Recognition Difficulties

R-SPIN SNR-50 thresholds were found to be lower when the target sentences were presented from a lateral loudspeaker rather than the central loudspeaker (see Figure 10). This finding is likely a result of two acoustic cues which are available to the listeners in the lateral presentation only: 1) Interaural SNR differences due to the head-shadow effect that allow the listener to take advantage of the better SNR in the ear facing the loudspeaker playing the target sentences; 2) Spatial separation between the target sentences and the babble masker. When the R-SPIN sentences are presented by the central loudspeaker, the listener perceives both the diffused babble (which is played by all three loudspeakers) and the sentences as coming from the front. However, when the R-SPIN sentences are presented from one of the loudspeakers on either side of the listener, there is a spatial separation between the sentences, which are presented form either the right or the left loudspeaker, and the babble, which is presented form all three loudspeakers simultaneously and therefore is perceived as emanating from the front. This separation has been found to be an assisting cue that promotes release from masking by enhancing stream segregation (e.g., Neher et al, 2009).

The R-SPIN results also indicate that young EL2 listeners and older EL1 listeners find it more difficult to recognize words in babble than do young EL1 listeners (see Figure 10). This increased difficulty in young EL2 listeners could reflect the fact that it is more difficult for them to segregate language streams in their L2 (Cooke et al., 2008; Ezzatian et al., 2010). This increased difficulty in a babble background could also arise from the activation of additional lexical candidates due to less than optimal phonemic discrimination in their L2 (Weber & Cutler, 2004), and/or to the activation of lexical candidates from other languages that the listener
commands (Spivey & Martin, 1999; Chambers & Cooke, 2009). In older EL1 listeners, poorer spoken word recognition performance in babble likely arises because age-related changes in hearing impede stream segregation (Ezzatian, Li, Pichora-Fuller, & Schneider, under review), and/or reduce the quality of the bottom-up information supporting lexical access.

**Trialogue Comprehension Results**

The trialogue comprehension results demonstrated that even though the presence of a third talker is likely to add to the perceptual and cognitive demands placed on the listener, no significant effect due to Group was found once the R-SPIN SNR-50 thresholds were used to adjust for individual differences in the ability to recognize words without supportive context. This finding is congruent with the results of several previous studies that used a similar method to insure that younger and older L1 listeners (Schneider et al., 2000; Murphy et al., 2006) and young L2 listeners (Experiment 1), were equated with respect to lexical access. The consistency of this finding across three types of listening conditions (monologues, dialogues, and trialogues), further supports the hypothesis that the underlying reason why older L1 listeners and young L2 listeners have poorer speech comprehension than young L1 listeners, is that these two groups encounter greater difficulties when attempting to achieve lexical access in noisy situations.

In addition, unlike the previous dialogue results, spatially separating the three talkers did not improve comprehension once the R-SPIN SNR-50 thresholds were used to adjust for individual differences in word recognition. It was found in Experiment 1, as well as in Murphy et al. (2006), that when participants were listening to two-talker conversations (dialogues), separating the two talkers from the background babble and from each other did significantly improve performance. However, there were several differences in the separation conditions used in the current study from those used in the previous studies that could explain why the separation
effect was no longer significant here. First, in the current study, the babble was played from all three loudspeakers, which eliminated the clear spatial separation between the talkers and the babble that was present in the real separation condition used in previous studies. Second, in the current trialogue comprehension study, the voice of one of the three talkers was played over the central loudspeaker. Because the babble played from all three loudspeakers was perceived as coming from the center as well, only the voices of the other two talkers (presented on the left and right) could benefit from acoustic cues that signaled that their spatial position differed from that of the babble. In the previous dialogue studies, the voices of both talkers were perceived to be located to the left and right whereas the babble occupied a central location. Therefore, in the dialogue separation conditions, the two voices as well as the babble were perceived as originating from three separate spatial locations. Hence, the addition of spatial cues is likely to provide less of a benefit in the trialogue study than it did in the previous dialogue studies. If so, this could explain why we did not find an effect of separation in this experiment. Alternatively, because the addition of a third talker made turn-taking less predictable, there are likely to be times when the listener switches their attention to the wrong talker, and, consequently to the wrong spatial location. In such situations, spatial separation might prove harmful rather than beneficial. Because turn taking is more predictable in dialogue situations, it is less likely that the listeners would focus their attention to the wrong spatial position. Hence, it is reasonable to assume that spatial separation would be of greater benefit to listeners when turn taking is predictable.

It is, of course, not surprising that performance decreased as a function of babble level. An increase in the level of a masker would be expected to interfere with speech comprehension at a number of different levels. First, increases in babble level would increase the difficulty of segregating the voices from the babble background. Second, it would be expected to increase the
number of lexical candidates activated either because the background babble energetically masked the target stream, or because phonemes in the babble intruded into the target stream. Third, the presence of babble is likely to cause the listener to focus attention on stream segregation and lexical access, thereby reducing the amount of attentional resources left for integrating and memorizing the information. If so, we might expect to find the contribution of such higher-order cognitive resources to be modulated by the presence and intensity of the babble.

4 General Discussion of Experiments 1 and 2

By individually adjusting the SNR under which dialogues and trialogues were heard, we were able in Experiments 1 and 2 to equate listening comprehension in younger and older L1s, and young L2s under a variety of listening conditions. However, the manner in which listeners achieved these equivalent levels of comprehension may differ across these three groups, and differed from one listening situation to another. To further investigate the degree to which different processes might be engaged by individuals in these three groups, and to determine whether the degree to which these processes were engaged differed across listening situations, we examined the extent to which the two measures of linguistic competence were correlated with the dialogues and trialogues comprehension performance. We might expect the ways in which listening comprehension is achieved to be more varied as the perceptual and cognitive complexity of the listening task increases (e.g., increasing the number if participating talkers from two to three, eliminating spatial separation, increasing background babble level, etc.). Specifically, we looked at each of the three groups (younger EL1 listeners, older EL1 listeners, and young EL2 listeners) to determine the contribution of the two linguistic measures to
performance (number of comprehension questions correctly answered) within each group and condition tested in Experiments 1 and 2 (see Appendices A and D for further details).

A regression analysis conducted on the results of Experiment 1 showed that both the vocabulary and the reading comprehension test results were related to the average number of comprehension questions correctly answered ($r = .32, p = .007$, and $r = .38, p < .001$, for reading comprehension skill and vocabulary knowledge, respectively). Interestingly, the results of the analysis indicated that there was a significant correlation between vocabulary knowledge and the number of dialogue questions correctly answered under the virtual location conditions but not under the real spatial location conditions only (see Figure 7), and that this interaction between the Type of Location and Vocabulary was significant. When source location is virtual, there are fewer acoustic cues that can be used to segregate the different acoustic streams than when source location is real (Freyman et al., 1999; Li et al., 2004). Under such conditions, obtaining and maintaining stream segregation is likely to be more demanding, and the bottom-up processes involved in lexical access might be less reliable because of occasional intrusions from the competing streams. Hence, a well-developed and extensive lexicon could help disambiguate the impoverished acoustic input.

The regression analyses performed on the data of Experiment 1 also indicated that the contribution of reading comprehension to performance differed among the three groups, with reading comprehension being more strongly correlated with listening comprehension in young EL1s but not in young EL2s or older EL1s. One possible explanation for this group difference is that attentional resources in both young EL2s and older EL1s were being employed to enhance lexical access, but for different reasons. Older EL1s may be allocating a larger portion of their attentional resources to lexical access than do young EL1s to compensate for impoverished
auditory input to the lexicon. Young EL2s may be allocating a larger portion of their attentional resources to enhance lexical access because of either an insufficiently developed L2 lexicon, or because lexical access may be mediated, in part, through L1 lexical processes. Hence, both of these groups, because of this enhanced engagement with lexical access, may not be as able as young EL1s to engage the higher-order processes modality independent processes assessed by reading comprehension to facilitate listening comprehension.

A similar regression analysis conducted on the results of Experiment 2 showed that both the vocabulary knowledge and the reading comprehension skill tests results were positively related to the average number of questions correctly answered \((r = .45, p < .001, \text{ and } r = .39, p < .001, \text{ for reading comprehension skill and vocabulary knowledge, respectively})\). Interestingly, the evaluation of the relative contribution of vocabulary versus reading comprehension within each group showed that reading comprehension contributes significantly to performance in both young EL1s and young EL2s, whereas vocabulary does not in the young EL1s, and only contributes marginally in young EL2s. However, older EL1s show a reversed pattern to that found in young EL1s in which vocabulary contributes significantly to performance, whereas reading comprehension does not (see Figure 12).

Consistently across Experiment 1 and 2, the general pattern that emerges is that reading comprehension accounts for a larger portion of the variance in speech comprehension in young EL1s, for a smaller portion of the variance in young EL2s, and no significant portion in old EL1s. With respect to vocabulary, there is evidence from Experiment 1 that when the listening scene is difficult to parse (virtual sound source location), vocabulary accounts for a significant portion of the variance in dialogue comprehension in all three groups, with the results of
Experiment 2 suggesting that the relative contribution or reading versus vocabulary to speech comprehension performance differs across the three groups.

In older adults, age-related hearing deterioration may require a greater allocation of attentional resources to basic auditory processes to facilitate stream segregation and lexical access. In turn, this greater demand for the deployment of attentional resources to early auditory processes could reduce the pool of available resources that is often thought to be smaller in older adults than in younger adults (Craik & Byrd, 1982). In other words, difficult listening situations, such as a multi-talker conversation, require the engagement of attentional resources at an earlier lexical level in order to compensate for the loss of fidelity in the aging auditory processing system. In younger listeners, the finding that individual differences in reading comprehension skill account for a significant portion of the variance in trialogue comprehension suggests that both reading comprehension and listening comprehension depend on a common set of higher-order, modality-independent skills. In younger EL1 listeners, the greater weight given to higher-order processes could be a result of a lesser need for assistance at the lower lexical level compared to that required in older EL1 listeners. In young EL2 listeners, it might be the case that they lack confidence in their lexical knowledge of L2 and rather use other skills and knowledge to overcome the difficulties they experience when attempting to process spoken language. Overall, Experiment 2’s results show that although all three groups achieved a similar degree of listening comprehension, they did so by engaging the top-down processes involved in speech comprehension to different extents. These results are consistent with previous findings showing that different neural circuitry supports speech comprehension in younger versus older adults (e.g. Harris et al., 2009; Wong et al., 2009).
An additional analysis was conducted to evaluate the degree to which the relative contribution of reading comprehension and vocabulary to performance might differ within each noise level (see Appendix D). The results of this analysis indicated that vocabulary is the major contributor to performance at medium levels of babble whereas reading comprehension contributes to performance in quiet and in high levels of babble (see Figure 13). This description supports Experiment 1’s results which indicated that as listening becomes difficult, top-down processes involved with lexical access become more fully engaged. When the trialogues were played in quiet, the contribution of vocabulary was found to be non-significant, suggesting that lexical access was accomplished with minimal assistance from higher-order processes. However, when trialogues were played in a moderate babble level, the need for top-down assistance rose, and the contribution of vocabulary knowledge to speech comprehension increased significantly. The reduction in the contribution of vocabulary knowledge back to a non-significant level in high babble suggests that when the bottom-up acoustic information becomes substantially degraded, higher-level processes can no longer provide assistance at the lexical level.

The contribution of reading comprehension skill remains relatively consistent across the different babble levels tested in Experiment 2. However, there is a suggestion in the data that when the babble level was moderate, the contribution of reading comprehension was reduced. According to the cognitive-load hypothesis (Sweller, van Merrienboer, & Paas, 1998), this reduction could be a result of the relatively large attentional investment that was made at the lexical access level. However, no significant differences were found in the contribution of reading comprehension to performance across the three babble levels.

The findings of Experiments 1 and 2 contribute to our understanding of just how complex and dynamic speech comprehension is. The differential contribution of vocabulary to trialogue
comprehension performance under the three babble levels suggests that in difficult listening conditions greater attentional resources are deployed in aid of scene analysis and word recognition at the lexical level. The greater correlation between vocabulary and triologue comprehension performance found when the babble level is moderate compared with a quiet background, implies that the relative weight given to the top-down processes contributing to lexical access becomes greater as the incoming acoustic information becomes less reliable. The hypothesis supported by the results of experiments 1 and 2, that listeners under a heavy processing load tend to systematically downplay the contribution of acoustic detail and increase the reliance on lexical-semantic knowledge, has been previously suggested by Mattys et al. (2009, 2010), Mattys and Wiget (2011). In addition, the differences in the contribution of vocabulary knowledge and reading comprehension skill found among the three groups suggest that the demand on top-down resources is modulated not only by the nature of the acoustic scene, but by the age and linguistic status of the listener as well. In other words, the degree and manner in which the central cognitive resources involved in language comprehension are engaged, likely reflects the listeners’ different weaknesses and strengths with respect to basic auditory processes, the degree to which they have a good command of the language, and the degree to which their cognitive resources are limited.

In real life situations, when older native and young nonnative listeners are not compensated for their greater speech recognition difficulties, their attentional resources might be deployed to facilitate lexical access, depleting the pool of resources that are left available to other higher-order processes such as information integration and memory. Understanding the reasons behind the difficulties experienced by many older listeners and L2 listeners is not only an interesting theoretical question, but one that has important clinical implications. First, the identification of the real source/s causing these difficulties is important for the self-efficacy of
older adults and nonnative listeners. Auditory treatment and rehabilitation is time demanding and effortful. An individual will invest effort in order to succeed in a task according to his or her subjective assessment of whether he or she can perform the task successfully (Bandura, 1997; West, Bagwell, & Dark-Freudeman, 2008). For example, due to a cascading effect, older listeners could interpret the difficulties they experience when attempting to follow a conversation in noise as a result of irreversible higher cognitive declines rather than a perceptual difficulty which drains their attentional resources. As a result, they might be much less motivated to go through the long process of hearing-aid fitting, and of learning how to employ adaptive listening strategies. In the case of young nonnative listeners, experiencing unexplained fatigue after participating in social gatherings or difficulty recalling information that was said to them in a noisy auditory scene, could hamper their efforts to successfully integrate into their new environment. Providing individuals with the information that would allow them to better understand the sources of their communication difficulties will allow them to make a more educated decision regarding what they can and are willing to do, to improve their communication abilities. Second, investigating all levels and processes involved in speech recognition and comprehension may lead to the development of more comprehensive, multi-processes interventions and listening strategies.

5 Experiment 3

The variety and nature of the auditory scenes, in which daily communication takes place, have changed significantly over the years due to the growing use of electronic amplification and surround sound systems. As amplification becomes more common in both public spaces and in private homes, it is important to understand how the changes it creates in the auditory scene may
affect the ability to communicate efficiently in it. When amplification is used, sound sources often are presented over more than a single loudspeaker (e.g., surround sound systems), creating a broader and more diffused auditory image of the original sound source. In contrast, in most natural settings, sound sources typically have a compact image emanating from the actual source location. For example, imagine watching a play in which the actors are conversing in a marketplace. When no amplification is used, the voices that emanate from the actors and actresses on the stage in front of you will have a compact image. To increase the degree of realism, the director may have recorded activity in an actual marketplace, and plays this recording over loudspeakers placed on both sides of the stage, which adds to the dramatic atmosphere, but at times might mask to some degree what is being said. Now imagine watching the play with the voices amplified and played over the same loudspeakers, creating a diffused image of both the voices and the marketplace activity accompanying the play. Any sound that is played at the exact same time from both sides of the listener will be perceived as emanating from the front. Therefore, in each of these situations the perceived location of the sound sources remains in the front. However, in the latter case both the market background (which contains a babble of voices) and the actors’ voices are diffuse whereas in the former situation, the actors’ voices are compact while the background is diffuse. This raises the question as to whether it is easier to understand the actors when their voices are compact and the background diffuse, or when both the actors’ voices, as well as the background sounds are diffuse? The experiments reported here were motivated by a consideration of such situations.

At the present time there is very little information on how the perceived diffuseness or compactness of the various sound sources might affect a listener’s ability to comprehend what is being said. When examining the R-SPIN SNR-50 thresholds measured in Experiment 1 and 2, some differences are found between two of the spatial conditions tested; 1) the real no separation
condition used in Experiment 1, in which both the target voice and the babble masker were presented over the central loudspeaker only; and 2) the central condition in Experiment 2, in which the target voice was presented over the central loudspeaker only while the babble masker was playing from all three loudspeakers. While in both these conditions there was no perceived spatial separation and the target and masker stimuli used were similar, the thresholds obtained in Experiment 1 were somewhat higher than those obtained in Experiment 2 (1.86 vs. -1.51, 2.32 vs. 1.02, 6.54 vs. 0.53 for the younger EL1s, the young EL2s and the older EL1s respectively). The present experiment, we believe, is the first systematic attempt to investigate how a difference in diffuseness of target voices relative to background sounds affects spoken word recognition.

The current study explores whether a contrast between a compact target and a diffused masker could provide listeners with acoustic information that would help them to better analyze the acoustic scene and minimize the interruption created by irrelevant sound sources. To better identify which level of auditory processing may be benefiting from this acoustic information, three different types of maskers were used; speech-spectrum noise, 12-talker babble, and two-taker competing speech. Based on the results of previous studies that examined the differences between energetic and informational masking, we would expect the benefit obtained from a contrast in diffuseness to be larger when the masker causes substantial informational masking rather than primarily energetic masking (e.g., Arbogast et al., 2002). In order to test this assumption we will conduct three pre-planned comparisons to examine if the degree of benefit due to a contrast in diffuseness between the target and masker differed across the three masker types. Because previous research has found striking differences between younger and older adults in their abilities to function in such complex listening situations, we included groups of younger and older adults who were native speakers of English. Because the degree of
competence in a language has also been shown to significantly affect one’s performance under such conditions, we also included a group of young listeners for whom English is a second language. In addition, we conducted three more pre-planned comparisons with respect to the degree of benefit obtained due to contrast in diffuseness between the target and masker differed across the three groups.

5.1 Method

Participants

The participants were 12 younger normal hearing listeners whose first language was English (Young EL1; mean age: 21.93 years; SD: 2.02), 12 older listeners with normal hearing for their age, whose first language was English (Older EL1; mean age: 72.76 years; SD: 4.54), and 12 young normal hearing listeners for whom English is their second language (Young EL2; mean age: 21.19 years; SD: 1.57). EL1 listeners were all born and raised in a country in which the primary language was English and were not fluent in any other language at the time of participation. EL2 listeners were those who first became immersed in an English speaking environment after the age of 11. The EL2 listeners were from a range of language backgrounds, specifically, Spanish (n=1), Hindi (n = 1), Mandarin (n = 6), Cantonese (n=1), Portuguese (n = 2), Malayalam (n = 1). The younger participants were students recruited from the University of Toronto Mississauga and the older participants were volunteers from the local community. All participants were asked to complete the same questionnaires regarding their general health, hearing, vision, and cognitive status that were administered in Experiments 1 and 2, with the same criteria applied for determining eligibility in the current study.

Materials, Apparatus and Procedure
Audiometric thresholds, Nelson-Denny reading comprehension skill and Mill Hill vocabulary knowledge were measured during each participant’s first session. The speech recognition task was administered over the next experimental session. Each of the two sessions was typically 1-1.5 hours in duration. All participants gave their written informed consent to participate in the experiments and were paid a modest stipend (10$/hour) for their participation.

**Hearing Measures**

*Audiometric Testing.* Pure-tone air-conduction thresholds were measured at nine frequencies (0.25-8kHz) for both ears using an Interacoustics Model AC5 audiometer (Interacoustic, Assens, Denmark). All participants were required to have a pure tone threshold 25 dB HL or lower from 0.25 to 3kHz and 35 dB HL or lower at 4 kHz. Older adults with hearing in the range described are considered to have normal hearing for their age (ISO 7029-2000). In addition, participants who demonstrated unbalanced hearing (more than a 15 dB difference between ears at any of the nine tested frequency) were excluded from participation. The average audiograms for the three groups of participants are shown for the right and the left ears in Figure 14. The two groups of younger adults had similar hearing levels at all frequencies. Hearing levels for older adults were about 7.41 dB poorer than those of the younger adults at frequencies ≤ 3 kHz, with the younger-older difference increasing as a function of frequency for frequencies > 3 kHz.
Figure 14. Average audiograms for the three groups of participants are shown for the right and the left ears.
Language Proficiency Measures

As in Experiments 1 and 2, participants were asked to complete the Mill Hill vocabulary test (Raven, 1965) and the Nelson-Denny reading comprehension test (Brown et al., 1981).

Semantically Anomalous Sentences-Recognition Task.

During test sessions, the listener was seated in a chair located in the center of an Industrial Acoustic Company (IAC) sound-attenuated chamber, whose internal dimensions were 283 cm in length, 274 cm in width, and 197 cm in height. Two loudspeakers were placed symmetrically in the frontal azimuthal plane at $45^\circ$ angles to the left and right of the listener, and a third loudspeaker was placed directly in front of the listener. The distance between the listener’s head and each one of the speakers was 169 cm. The height of the loudspeakers was adjusted to match the ear level of a seated listener of average body height. All the acoustic stimuli used for the current study were digitized at 20 kHz sampling rate using a 16-bit Tucker Davis Technologies (TDT, Gainesville, FL) System II and custom software. The digital signals were converted to analog forms using Tucker-Davis Technologies digital-to-analog converters under the control of a Dell computer with a Pentium 4 processor. The analog outputs were low-passed at 10 kHz, attenuated by two programmable attenuators, and then presented to the participant either through the central loudspeaker (compact target speech and compact maskers) or from all three speakers (diffuse maskers). Target sentences consisted of 312 syntactically-correct-but-semantically-anomalous sentences spoken by a female talker, which were developed by Helfer (1997) and previously used in experiments by Freyman et al. (1999), Li et al. (2004) and Ezzatian et al. (2010). Each of these sentences contained 3 target words in sentence frames such as “A spider will drain a fork”, or “A shop can frame a dog” (target word italicized). The sentences were divided into 24 lists containing 13 sentences each. The target sentences were presented over the front loudspeaker while the masker was either presented over all three
loudspeakers to create a diffused image, or over the central loudspeaker only to create a compact image of the sound source. Under both conditions, the target sentences as well as the masker were perceived as emanating from the center.

Target sentences were presented with either one of three types of masking stimuli: noise, babble or speech. The noise masker was a steady-state speech-spectrum noise recorded from an audiometer (Interacoustic [Assens, Denmark] model AC5), the babble was a 12-talker babble and the speech masker was a 315 second long track created using an additional set of semantically anomalous sentences uttered by two female talkers, and repeated in a continuous loop. The target sentences were presented at an average sound pressure of 55 dBA at the estimated center of a listener’s head, whether a single loudspeaker was playing the sentences (compact target) or all three (diffused target). The sound pressure was measured using a Brüel & Kjær (Copenhagen, Denmark) KEMAR dummy-head. Masker intensity was measured separately for the conditions in which the masking sentences were played only over central loudspeaker (compact masker), and when they were simultaneously played over all three loudspeakers. While the target’s sound pressure level remained constant at 55 dBA throughout the experiment, the sound pressure level of the masker was adjusted in order to produce 4 different SNRs depending on the masker type and the group tested. The different SNRs used were initially chosen based on previous studies that used similar stimuli in noise (e.g. Ezzatian et al., 2010) and then altered according to the results of preliminary pilot testing done under the present listening conditions. The SNRs used in the current study are presented in Table 3. A single list of 13 sentences was used for each of the SNR values which appear in the table.

Because the head was positioned facing the central loudspeaker, with the approximate center of a participant’s head being equidistant from each of the three loudspeakers, we would expect a zero interaural phase difference for both masker and signal, independent of whether or
not the masker was compact or diffuse. Hence any difference observed between the two types of
maskers is unlikely to be due to binaural unmasking. Binaural unmasking (a reduction in the
SNR required for word recognition) can occur whenever there is a large interaural phase
difference between the signal and the masker. For example, the SNR threshold for speech
recognition when the signal is in phase and the masker is 180 degrees out of phase at the two
ears is significantly lower than when the interaural phase difference between the two ears is the
same for both the signal and the masker (Licklider, 1948; Johansson & Arlinger, 2002). Because
we do not expect any interaural phase differences between the signal and the two types of
maskers (diffuse versus compact) when the observer’s head is properly positioned, it is unlikely
that any difference in threshold between a diffuse and compact masker is due to binaural
unmasking.
Table 3

The values of the four SNRs used under each condition (compact target and maskers ($T_C M_C$), compact target and diffuse maskers ($T_C M_D$), for each of the three masker types, presented separately for each of the three groups.

**Younger EL1**

<table>
<thead>
<tr>
<th>$T_C M_C$</th>
<th>$T_C M_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>-9</td>
<td>-8</td>
</tr>
<tr>
<td>-15</td>
<td>-13</td>
</tr>
</tbody>
</table>

**Older EL1**

<table>
<thead>
<tr>
<th>$T_C M_C$</th>
<th>$T_C M_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>-8</td>
<td>-7</td>
</tr>
</tbody>
</table>

**Young EL2**

<table>
<thead>
<tr>
<th>$T_C M_C$</th>
<th>$T_C M_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>-4</td>
</tr>
<tr>
<td>-7</td>
<td>-9</td>
</tr>
</tbody>
</table>
Each of the 24 target lists was presented at a constant SNR. Sentence lists and SNRs were counterbalanced across participants such that each list was presented at each of the 4 different SNRs an equal number of times in each group. Additionally, each sentence list was presented in each of the Diffuseness conditions (diffused masker, compact masker) and Masker (speech, babble, noise) combinations an equal number of times. In each group, six participants were first tested with a diffused masker (M_D) for the first 12 lists, and with a compact masker (M_C) for the remaining twelve. The other six participants were tested in the reversed order. Before beginning the experimental session, an explanation was given to familiarize the participant with the task. Participants were asked to repeat back the target semantically anomalous sentence after each presentation, and were scored for any keyword which was repeated correctly. Performance was assessed both online while the session was taking place, and later by a second research assistant who listened to the participants’ recorded responses. After the participant had responded, the researcher initiated the presentation of the next trial. Each trial started with the masker sound which was followed 1 s later by a target sentence. The masker remained on during the sentence, then the masker was gated off when the target sentence was turned off. After completing twelve lists, a short break was offered to the participants.

5.2 Results

Table 1 presents the gender breakdown, mean age, Mill Hill test of vocabulary knowledge and Nelson-Denny test of reading comprehension results for each of the three groups. The column on the right indicates which of the measures and values differ significantly among the groups. The average age of the older EL1 listeners was significantly higher than that of the two younger groups, while the age of the two younger groups was similar. The older and younger EL1
listeners had significantly higher vocabulary scores and reading comprehension scores than did the young EL2.

Table 4

Demographic information (Mean Age, Gender Distribution, Mean Vocabulary and Reading Comprehension Scores) for the participants divided into the three groups tested. Statistically significant differences between groups are noted.

<table>
<thead>
<tr>
<th></th>
<th>Younger natives</th>
<th>Young nonnatives</th>
<th>Older natives</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Age</td>
<td>21.93</td>
<td>2.02</td>
<td>21.19</td>
<td>1.57</td>
</tr>
<tr>
<td>Gender</td>
<td>8 F + 4 M</td>
<td>9 F + 3 M</td>
<td>11 F + 1 M</td>
<td></td>
</tr>
<tr>
<td>Vocabulary (max=20)</td>
<td>14.5</td>
<td>1.24</td>
<td>9.17</td>
<td>3.88</td>
</tr>
<tr>
<td>Reading comprehension (max=36)</td>
<td>28.58</td>
<td>3.92</td>
<td>18.08</td>
<td>6.29</td>
</tr>
</tbody>
</table>

A= A statistically significant difference between Younger EL1 and Older EL1 listeners.
B= A statistically significant difference between Younger EL1 and Young EL2 listeners.
C= A statistically significant difference between Older EL1 and Young EL2 listeners.

Figure 15 shows the percentage of correctly identified keywords, averaged over the twelve participants in each group, as a function of SNR when the masker was speech spectrum noise (left panels), two-talker speech (middle panels) or 12-talker babble (right panels) for compact maskers (circles) and for diffuse maskers (squares). Logistic psychometric functions of the form \( y = \frac{1}{1 + e^{-\sigma(x-\mu)}} \)
were fit to these data points. The parameter $\mu$ denotes the 50% point on the psychometric function (the threshold), and $\sigma$ controls the slope of the function (for a description of the fitting procedure see Yang et al., 2007). The estimated 50% points for the younger EL1s (top panels), older EL1s (middle panels), and young EL2s (bottom panels) are indicated by solid vertical lines when there is a contrast between a diffuse masker and compact target, and dashed vertical lines for the no-contrast (both target and masker compact) condition.

An examination of this figure suggests that indeed when there is a contrast in diffuseness between the target (compact sound source) and the masker (diffused sound source), this dissimilarity can be used as a cue to assist stream segregation and enhance speech recognition. In addition, it appears that younger EL1s are performing significantly better compared with the other two groups, and are possibly gaining a slightly greater benefit from a diffuseness contrast between the target (compact) and the masker (diffuse). When examining the differences in the estimated SNRs at which 50% correct repetition was achieved when there was a contrast in diffuseness compared with when there was not, the differences seem to be larger when the masker is either speech or babble than when it is noise. These slightly larger differences imply a greater release from masking due to the diffuseness contrast cue for informationally more complex maskers. The estimated slopes of the psychometric functions when the masker is noise appear to be steeper than those estimated when the masker is babble or speech.
Figure 15. Average percent correct word identification as a function of SNR in dB for three groups (from top to bottom: Young EL1s, Older EL1s, and Young EL2s) and for three masker types (from left to right: Noise, Speech and Babble). The circles represent the results for the condition in which both the target and masker were compact sound sources. The squares represent the results for the condition in which the target was a compact sound source while the masker was a diffuse sound source. Logistic functions were fit to the data. Thresholds (SNRs corresponding to 50% correct on the psychometric functions) are indicated by solid vertical lines when there is a contrast in diffuseness between target and masker (target compact and masker diffused) condition, and dashed vertical lines for the no-contrast (both target and masker compact) condition.
These visual impressions were mostly confirmed by statistical analyses performed on the parameters of the individual psychometric functions. Specifically, psychometric functions were fit to all individuals in order to obtain individual estimates of the threshold, $\mu$, and the slope, $\sigma$. These estimates were entered into a 2 Diffuseness Condition (masker compact, masker diffused) by 3 Masker Type (noise masker, babble masker, speech masker) by 3 Group (Younger EL1, Older EL1, Young EL2) analysis of variance (ANOVA) with Group as a between-subjects factor, and Diffuseness condition and Masker type as within-subject factors. The ANOVA for thresholds ($\mu$) revealed that all three factors have a significant main effect on thresholds (Group: $F[2, 33]=20.203$, $p<0.001$; Diffuseness condition: $F[1, 33]=40.927$, $p<0.001$; Type of masker: $F[2, 66]=1390.171$, $p<0.001$). The main effect of Diffuseness condition, which was found to be statistically significant, confirms that dissimilarity in diffuseness between a compact target sound source and a diffused masker sound source, can be used as an assisting cue to create a significant release from masking. However, the effect of Diffuseness condition did not significantly interact with Group ($F[2, 66]=1.229$, $p=0.306$), but the interaction between Diffuseness Condition and Masker approached significance ($F[2, 66]=2.669$, $p=0.077$). An examination of the three pre-planned comparisons with respect to the degree of benefit obtained from a contrast in diffuseness across the three masker types, showed that the benefit from the contrast was significantly larger when the masker was Speech compared with when it was Noise ($t(35)=2.501$, $p=0.017$). Neither of the other two planned comparisons (one for Speech versus Babble, the other for Babble versus Noise) was found to be significant. The lack of a statistically significant Diffuseness condition by Group two-way interaction does not support the idea that younger EL1 gain more from contrast in diffuseness. This conclusion was confirmed insofar as none of the three planned comparisons to test whether or not the benefit from diffuseness differed across the three groups approached significance. To further investigate the Group main
effect, post hoc tests with Sidak adjustment were conducted. The results indicated that the younger EL1 listeners differed significantly from both the older EL1 listeners ($p = .001$) and the young EL2 listeners ($p < .001$) but that the young EL2 and the older EL1 listeners did not differ significantly from one another ($p = .106$). In regard to the main effect of Masker, paired-samples t-tests were conducted to determine the degree to which the percentage of words correctly identified differed between pairs of maskers. After applying a Bonferroni correction, all paired comparisons showed significant differences: Noise versus Babble: $t(35)=37.735, p<0.001$; Babble versus Speech: $t(35)=-42.509, p<0.001$; and Speech versus Noise: $t(35)=-4.127, p<0.001$). In addition, a two-way Group by Masker type interaction was found to be significant ($F[4, 66]=6.214, p= < .001$). To describe this interaction, the SNR thresholds estimated for the Noise condition will be used as a baseline for comparison (see Figure 16 to identify the comparisons which show statistically significant differences). The steady-state noise was chosen as a reference condition because it is primarily an energetic masker, and as such is unlikely to interfere with speech recognition at higher levels of auditory processing. No significant difference in threshold was found in young EL1s between Speech and Noise maskers (0.1 dB difference). However, older EL1s, and young EL2, do show an increase in the thresholds when the masker is Speech versus Noise (Older EL1: 1.19 dB difference, Young EL2: 2.39 dB difference). A paired-sample t-test, which was conducted for each of the groups, confirmed this description showing no significant difference between Noise and Speech thresholds in Younger EL1s, a moderately significant difference in Older EL1s ($t(11)=2.183, p=0.052$), and a highly significant difference in Young EL2s ($t(11)=6.956, p<0.001$). When conducting a post hoc pairwise comparison with Sidak corrections to compare the Noise-to-Speech differences among the three Groups, the difference in Younger EL1s was found to be significantly smaller than that in Young EL2s ($p=0.003$), but not in Older EL1s ($p=0.268$).
Next we compared the SNR thresholds that were estimated for when the masker was Babble to those estimated for when the masker was Noise. The 12-talker Babble most likely creates a lesser degree of informational masking compare to the Speech masker, and because there are energy fluctuations in the babble masker, an opportunity to hear the target speech in the troughs of the babble masker. When comparing the Babble SNR thresholds to those estimated for Noise using a paired-sample t test, a significant decrease is found in each of the three groups ($t(11)=-70.19, p<0.001$; $t(11)=-22.47, p <0.001$; $t(11)= -21.22, p<0.001$ in Younger EL1, Older EL1 and Young EL2, respectively). When a post hoc pairwise comparison with Sidak corrections was conducted to compare the Noise-to-Speech differences among the three Groups, the difference was found to be significantly larger in the Younger EL1s than in both the Older EL1s ($p=0.001$) or the Young EL2s group ($p<0.001$).

Finally there was no evidence indicating the presence of a three-way interaction of Diffuseness condition by Masker by Group ($F_{[4, 60]}=0.332, p =0.56$).
Figure 16. Average SNR at 50% correct speech recognition thresholds (dB) calculated under each of the three masker (S= speech, N=noise, B=babble) conditions for each of the test groups. Statistically significant differences are indicated by “**” as \( p<0.01 \). Standard error bars are shown.
The slopes, \( \sigma \), of the individual psychometric functions were also analyzed using a 2 Diffuseness Condition by 3 Masker Type by 3 Group ANOVA. This analysis revealed a significant main effect of Diffuseness Condition on slopes (\( F[1, 33]=14.219, p=0.001 \)), as well as significant main effects of Masker Type (\( F[2, 66]=53.321, p<0.001 \)) and Group ( \( F[2, 33]=20.802, p<0.001 \)). On average, slopes decreased by a value of 0.035 when there was a diffuseness contrast between a compact target and a diffuse masker, compared with when there was no contrast (both target and masker were compact). In order to investigate the Group main effect, Post hoc tests with Sidak adjustment were conducted. The results indicated that the younger and older EL1 listeners’ slopes differed significantly from those of the young EL2 listeners (for both comparisons \( p<.001 \)) but that the young EL1’s and the older EL1’s slopes did not differ significantly from one another (\( p=.527 \)). A paired-sample t test was conducted in order to compare the average slopes under the different Masker conditions. The results of the paired comparisons showed a highly significant difference among all three masker types, with the slope being significantly higher for Noise than for Speech (\( t(35)=-2.119, p=.041 \)) and for Speech compare to Babble (\( t(35)=6.762, p<.001 \)). In addition, the interaction of Diffuseness condition and Masker type was found significant (\( F[2, 66]=3.828, p=0.027 \)), indicating that the decrease in slope that occurred when maskers were diffused depended on the type of masker. As can be seen in Figure 17, the largest decrease occurred when the masker was noise (0.061), followed by when the masker was speech (0.0327), and the smallest decrease in slope occurred when the masker was babble (0.012). Paired-samples t-tests were conducted to compare the decrease in slopes between the different maskers. After applying a Bonferroni correction, the only paired comparison which showed a significant difference was Noise versus Babble: (\( t(35)=2.77, p=0.009 \)).
Figure 17. Average slope differences (slope when masker is a compact sound source – slope when masker is a diffused sound source) calculated under each of the three masker (S= speech, N=noise, B=babble) conditions. Standard error bars are shown.
Another statistically significant two-way interaction was found between Masker type and Group ($F[4, 66]= 4.821, p =0.002$). Figure 18 plots the slopes of the psychometric functions averaged across the two Diffuseness Conditions for each of the groups for the three masker types. This figure illustrates an interesting pattern of interaction in which the slopes of all three groups differ from each other when the masker is noise (younger EL1 vs. young EL2, $p<.001$; older EL1 vs. young El2, $p=.003$; younger EL1 vs. older EL1, $p=.002$), while for both Babble and Speech maskers the average estimated slopes for the younger and older EL1 were very similar to each other and higher than those estimated for the young EL2 (Babble: for younger EL1 vs. young EL2 $p=.003$ and for older EL1 vs. young El2 $p=.002$; Speech: for younger EL1 vs. young EL2 $p=.002$ and for older EL1 vs. young EL2 $p<.001$).

To determine whether individual differences in linguistic competence (vocabulary and reading comprehension skills) could account for a significant portion of the variance in the speech recognition task, the Mill Hill and Nelson–Denny scores were centered within each group and entered into multiple regression analyses to predict centered 50%-correct speech recognition thresholds. The results of this analysis showed that the cognitive measures account for a relatively large portion ($r^2=.495$) of the variance in average thresholds in young EL2s ($F[2,9] = 8.838, p = .009$), and a smaller ($r^2=.367$) moderately significant portion ($F[2,8]= 4.635, p = .046$) in older EL1s. In contrast, the two measures of linguistic competence did not account for a significant portion of variance of the variance in the spoken word recognition performance in younger EL1 ($r^2=.038, F[2,9]= 0.353$). The latter finding could be a result of the relatively small variance in the Mill Hill and Nelson-Denny scores in the younger EL1, which leaves less opportunity to observe a significant correlation.
Figure 18. Average estimated slopes calculated under each of the three masker (S= speech, N=noise, B=babble) conditions for each of the test groups. Standard error bars are shown.
5.3 Discussion

**SNR thresholds and performance**

The current study provides strong evidence that a contrast between the diffuseness of the target and that of the masker can provide a release from masking. Because both the target and the masker were perceived to emanate from the front of the listener regardless of whether the masker sound source was compact or diffuse, the benefit provided by the contrast between the diffuseness of the masker and the compactness of the target voice cannot be attributed to a perceived spatial separation between them. Interestingly, all three groups benefitted from a contrast in diffuseness. Hence older adults can take advantage of this contrast, and the fact that young EL2s could also benefit from this contrast suggests that this ability does not depend on the linguistic competence of the listener.

The fact that no statistically significant interaction was found among the groups tested and the benefit derived from a contrast in diffuseness implies that the ability to use this contrast as an assisting cue remains relatively preserved with age. The lack of a two-way interaction between Diffuseness Condition and Group supports the assumption that this cue is most likely not language-oriented and is as available and helpful to EL1 listeners as it is to EL2 listeners regardless of linguistic status.

While no evidence was found to support the effect of age or linguistic background on the size of the benefit due to the presence of a contrast in diffuseness, the data do provide evidence that this assisting cue may be more helpful when the listener is experiencing informational interference. Hence the diffuseness versus compactness of an image is an additional bottom-up cue that appears to provide some release from informational masking. The significant greater
difference in the benefit gained under speech masking conditions compared with that gained under the noise masking conditions further strengthens the hypothesis that there is a greater degree of informational masking when the competing masker is speech than when it is either noise or babble. It is therefore reasonable to attribute the greater benefit when the masker is speech to improved stream segregation in the presence of dissimilarity in the diffuseness of the target talker compared with that of the competing talkers. The better stream segregation achieved enables the listeners to focus better on the target stream, while more efficiently ignoring the irrelevant information in the speech of the competing talkers.

The masker condition under which the listeners were performing also had a significant main effect on the 50% correct speech recognition thresholds. Further analysis showed that thresholds were higher (when averaged across the three groups) under the speech conditions compare to those under the noise conditions, which in turn were higher than those under the babble conditions. This finding is probably a result of both the nature of interference that each of these types of maskers generates, as well as specific acoustic features of the maskers used in this study. The noise masker used in this study was a speech-spectrum noise that generated constant energy at the frequency range typical of the human voice, causing an energetic interference. There were no modulations or fluctuations in intensity in the noise to mimic the amplitude fluctuations that occur in speech; as a result, listeners cannot benefit from troughs in the amplitude envelope that are present in speech but not in steady-state noise. Therefore, the primary parameter affecting the extent of the interference was the SNR at which the noise was presented to the listener.

The speech masker used in this study consisted of two female (same gender) talkers who were speaking at a similar rate. Their utterances were short semantically anomalous sentences
which were recognizable, and as such they most likely create a substantial amount of informational masking. However, any differences found between speech recognition performances under the noise conditions versus the speech conditions cannot be attributed solely to a difference in the degree of informational masking between competing voices and noise. As mentioned previously, speech signals contain amplitude fluctuations that allow the listeners to take advantage of troughs in the amplitude envelope (Cooke, 2006). Therefore, these differences are likely to reflect a combination of greater informational masking as well as the ability to focus attention on the target speech in the troughs in the envelope of the speech masker. The significantly higher SNR thresholds (averaged across groups) obtained under the speech conditions, compared with those under the noise conditions, implies that the additional difficulty, which is added when competing information is presented, overcomes the benefit from the troughs in the amplitude.

When comparing the speech thresholds obtained in competing speech with those obtained in noise background in each group separately, the thresholds obtained in speech background were found to be significantly higher in the older EL1s and in the young EL2, while the younger EL1s had very similar SNR thresholds under both conditions. The origins of the greater difficulties older EL1s are experiencing in the presence of competing talkers compared with a background of noise, are likely to be different than those in the young EL2s. Older adults might be more susceptible to informational masking (Helfer & Freyman, 2008), which could be a result of age-related changes such as a reduced ability to inhibit irrelevant information (Hasher & Zacks, 1988), and/or a possible reduction in the ability to take advantage of rapid modulations and short troughs in the amplitude of the competing voices. There is similar evidence in previous studies that suggests older listeners experience a greater interference from meaningful maskers than younger listeners (Tun, O'Kane, & Wingfield, 2002; Rossi-Katz & Aerhardt, 2009).
Younger EL2s are not expected to experience similar reduction as their peripheral hearing as well as their speed of processing and attentional abilities should be similar to those of the younger EL1s. Therefore, in young EL2s it is more likely that the greater difficulty they are experiencing in the presence of a speech masker is a result of their L2 semantic and linguistic processes not being fully developed and/or not completely differentiated from the semantic and syntactic processes that occur in their L1 (Kroll & Steward, 1994). While the thresholds assessed for the older EL1s under the speech conditions were found to be only moderately higher than those under noise, the younger EL2s showed a much more significant difference between the two maskers. This implies that the interruption created by the competing semantic and linguistic information was such that, even with the presumably better ability to benefit from the modulation in the intensity, younger EL2s were experiencing greater difficulty compared with the older EL1s. This finding corresponds with previous findings showing poorer speech recognition and comprehension performance in young EL2s when compared with older EL1s, and even more so when compared with younger EL1s, that were found in Experiment 1 and 2.

When comparing the thresholds assessed under the babble conditions to those assessed under the noise conditions, a highly significant decrease in the SNR thresholds is found in all three groups. This decrease, although robust, is not surprising as a similar multi-talker babble masker has been found in previous studies to cause a relatively small amount of interference (Lewis, Benignus, Muller, Malott, & Barton, 1988; Brungart et al., 2001). In multi-talker babble, as used here, the individual voices are no longer intelligible. As a result the babble masker contains less linguistic information compared with the speech masker, and therefore produces less informational masking (e.g., Simpson & Cooke, 2005). In addition, as it is a mixture of speech signals, the babble masker contains amplitude fluctuations as does the speech masker. These temporal and spectral “dips”, although not as frequent and deep as in the speech masker,
allow the listeners to pick-up components of the target speech at more favorable SNRs (Peters, Moore, & Baer, 1998). Moreover, the 12-talker babble used in the current study was taken from the R-SPIN test. As such, its spectral composition matched that of the male target voice. The target voice in this study was female. As such, although its RMS average matched those of the two other maskers, its spectral composition was such that it had less energy in the low-frequency region and much more energy in the high frequency region (see Ben-David et al., 2012, for an example of how the spectral composition of a female voice differs from that of the babble masker). The increased energy in the high-frequency region most likely is responsible for the lower thresholds in babble compared with noise or competing speech.

Significant differences in speech recognition performance was also found among the groups, as the SNR thresholds of the younger EL1s were found to be significantly better than those of the older EL1s and the young EL2s. The young EL1s are yet to experience age-related declines in hearing or in higher-level processes involved in speech recognition, while equipped with the linguistic competence which seems to be reserved for natives. While young EL2s are not likely to differ from young EL1s in regard to their basic auditory and cognitive abilities, they might have inferior semantic and linguistic processes in L2. In addition, they are prone to additional interference from activity taking place in their L1, as the two languages might not be fully differentiated (Kroll & Steward, 1994). As a result, L2 listeners have been found to have lower performance in a number of speech recognition tasks compared with native listeners of a similar age both in the presence of stationary noise (e.g., Bradlow & Bent, 2002; Gracia Lecumberri & Cooke, 2006) and maskers composed of speech babble (e.g., Mayo et al., 1997; Ezzatian et al., 2010).
Slope differences and interaction patterns

In general, examining the slopes provides valuable information regarding how increases in SNR are translated into increases in speech recognition performance under the different conditions and for the different groups. The results show that once a contrast of diffuseness between a compact target sound source and a diffused masker sound source is present, the slopes decreased by an average value of 0.035. This finding corresponds with previous studies on spatial separation, which have shown that when a spatial cue is provided to the listeners, slopes become significantly less steep (Li et al., 2004; Ezzatian et al., 2010). In general, a release from informational masking will have its greatest effect at low SNRs. The reasons for this are as follows: Any acoustic cue that distinguishes the target voice from the masker is likely to promote stream segregation. This implies that, at high SNRs, the intensity of the target voice itself will function as a cue for stream segregation. Hence, we would not observe as much of an improvement due to a change in the diffuseness of the background relative to that of the target at high SNRs as we would at lower SNRs, since the difference in diffuseness at lower SNRs would provide a release from masking. The overall effect would be a flattening of the slope of the psychometric function.

The slopes of the psychometric functions also differed across Groups. Post hoc tests revealed that the younger and older EL1 had significantly steeper slopes compared with young EL2s. This implies that word recognition performance improved more slowly as SNR increased for EL2s than for EL1, regardless of age. Similar findings were found in previous studies that tested speech recognition in both EL1s and EL2s (Bradlow & Bent, 2002; Ezzatian et al., 2010; Mayo et al., 1997). There are several possible explanations as to why EL1 listeners seem to be
better able to make use of acoustic information which becomes available as the SNR increases to improve speech recognition. Certain aspects of human language effectively impose constraints on how listeners recognize speech (Garcia Lecumberri et al., 2010). One of the main constraints has to do with the limited number of phonemes available in a language, which is trivially small considering the size of the vocabulary it supports (Maddieson, 1984). The practical implication of this phoneme to word ratio is that it severely limits the dissimilarity between different members of the language’s lexicon. According to cohort models of lexical access (Taft & Hambly, 1986), all words that might be a potential match, given the unfolding auditory input available, will be activated. Following the notion that there is a critical period for phonological encoding during a person’s life (Florentine, 1985), we would expect EL1 listeners to be better at noticing fine phonological nuances than EL2 listeners. This superior ability could allow the EL1s to better eliminate competitors and reduce the lexical competition as the target word unfolds. Moreover, EL2 listeners might experience greater lexical competition not only due words in the target language which they haven’t been able to eliminate, but from additional competitors which might have been activated in their first language (L1). From a statistical point of view, in a state of uncertainty due to partial acoustic information, EL2 listeners might find themselves disadvantaged twice; once for having to choose a word out of a larger pool of remaining candidates, and second, for being less-familiar with the statistical properties of the target language (Saffran, 2003). The consequence of the second might be a larger number of possible branches that could lead to words. In addition, for EL2s, the terminus of many of these branches is a sequence of phonemes that lack meaning for them. This could potentially increase the difficulty when attempting to identify words based on partial acoustic input. For example, it might be harder for an EL2, compared with an EL1, to recognise that a branching that led to “flug” terminated in a non-word. Hence, even though an EL2’s vocabulary size might be smaller
than that of an EL1’s, the number of branches in their lexicon could be larger, making it more
difficult for them to access spoken words in the language.

In addition, an interesting pattern was revealed when the slopes of the psychometric
functions for the different maskers were compared across groups. When the masker was noise,
younger EL1 showed significantly steeper slopes compared with older EL1s, who in turn, had
steeper slopes compared with the young EL2s. However, when the masker was either speech or
babble, no significant difference was found between the younger and older EL1s, while both EL1
groups had significantly steeper slopes than the young EL2s. The steeper slopes found in the EL1
listeners compared with the EL2 under all masker conditions, implies, for the reasons discussed
earlier, that when listening to speech spoken in L2, the ability to translate increases in SNR to
better speech recognition performance is reduced.

The difference in slopes which was found between the younger and older EL1s when the
masker was noise, but not when it was either speech or babble, warrants further discussion. This
interaction implies that while older EL1s are capable of using the increase in SNR to improve
speech recognition to the same extent as younger EL1s in the presence of speech and babble
maskers, they fail to do so when the masker is noise. A noise masker is unlikely to elicit any
activation in the semantic or linguistic processes. As such, the interference it causes is essentially
energetic. Energetic masking is considered to be less subject to listener control compared with
informational masking (Mattys, Davis, Bradlow & Scott, 2013). When the masker is noise it is
reasonable to assume that a greater weight will be assigned to basic auditory processes, rather
than to high-order processes, in order to minimize the impact of the energetic masking.
Therefore, it is not surprising that older adults, who experience substantial age-related declines in
basic auditory processes, present different slopes than their younger counterparts when listening in noise.

In the current study, participants were also asked to complete two tests that are measures of language competence. The Mill Hill provides an estimate of the individuals’ vocabulary knowledge, while the Nelson-Denny provides reflects the processes and skills involved in reading and comprehending written prose. The individual scores were centered within each group and the individual differences were correlated with the speech recognition results in order to examine whether the processes and skills that these two cognitive tests tap, could account for individual differences in speech recognition performance. The results showed that the cognitive measures accounted for a relatively large portion of the variance in average thresholds in young EL2s, and a moderately significant portion in older EL1s. However, there was no indication that the two measures of linguistic competence were related to individual differences in speech recognition in younger EL1s. These results imply that the young EL2s and the older EL1s used top-down processes in order to complete the speech recognition task, while the young EL1s relied mostly on bottom-up acoustic information. The idea that a greater engagement of top-down processes takes place when the acoustic information available to a listener is insufficient or/and unreliable, has been previously presented in Experiment 1 and 2. While the reasons behind the greater difficulties young EL2s and older EL1s experience are likely to be quite different between the two groups, they both use top-down, knowledge-driven processes to compensate for the missing acoustic information. The young EL1s, however, do not seem to feel the need to engage higher-order processes to complete the recognition task given here.

The results of the current study may have important practical implications as they call for a reassessment of how surround sound systems should be designed and sound tracks should be
mixed and assigned to channels in order to enhance speech recognition. For example, theatres could assist their audience, especially those experiencing difficulties, by amplifying the voices of the actors and actresses using a limited number of loudspeakers to maintain compact images for the voices while presenting the background sounds using loudspeakers placed all around the audience to create a contrast between the compactness of the target voices and the diffuseness of the background. On a similar note, television and movie sound technicians may want to mix the target voices into a limited number of channels so that they have a precise location in space.

6 Conclusions

Understanding multi-talker conversations in noisy situations is a common daily activity that challenges both the perceptual and cognitive capabilities of listeners. Due to possible age-related declines in the auditory, cognitive, and/or linguistic processes supporting speech comprehension, older listeners typically experience greater difficulties than younger listeners when attempting to follow multi-talker conversations in noise. Such situations are also quite challenging to young adults listening in their second language (L2). The aim of the present thesis was to identify the mechanisms supporting the comprehension of conversations in both groups in laboratory listening conditions that more closely resembled natural listening situations than is typical of speech recognition and comprehension studies. Specifically, we examined listeners' abilities to understand speech in challenging conditions by creating simulations of two and three-talker conversations in babble background. Each experiment included three groups of listeners who differed in age and/or linguistic status. Testing all three groups allowed greater insight into how age and linguistic status altered the ways in which listeners perceive and process speech, and changed the balance between the contributions of bottom-up versus top-down processes to
speech comprehension. More specifically, we investigated the extent to which individual differences in vocabulary knowledge and reading comprehension could account for individual variation in speech comprehension under different listening conditions in three populations: younger EL1 listeners, older EL1 listeners who were presumably experiencing age-related declines in auditory processes (and possibly cognitive processes); and young EL2 listeners who presumably had intact auditory and cognitive systems but who had not attained the same degree of mastery of English as those for whom English was a first language. In the following section, I will summarize my main findings and I will then discuss the implications of these findings to the understanding of how speech comprehension is achieved. Finally, I will suggest future research directions.

In Experiment 1, younger and older EL1s as well as young EL2s were asked to listen to several conversations played against a babble background, with or without either real or virtual spatial separation between the talkers and masker. Speech comprehension was assessed by the number of questions correctly answered at the end of each conversation. Individual hearing differences were compensated for by creating the same degree of difficulty in identifying individual words without supporting context when these words were embedded in sentences presented in babble. Applying such compensation equated all three groups with respect to spoken word recognition. If the greater degree of difficulty that older EL1s and young EL2s experience in everyday situations were solely due to spoken word recognition problems in these two groups, we would expect group differences in performance to be minimized or even disappear, which was precisely what was found. The results showed that all listeners performed better when there was a spatial separation between the two voices and the babble masker, implying that all three participants’ groups were able to use this spatial cue to achieve release from masking. The only difference that we found among these three groups was an indication that older adults benefited
less than younger adults when there was a real spatial separation between the target voices and the babble background. In all other conditions there was no evidence to suggest that the groups differed with respect to their comprehension of the dialogues.

The equivalence among the three groups with respect to comprehension allowed us to examine, using a correlational approach, the degree to which linguistic competence could account for individual differences in performance, as well as the degree to which the contributions of vocabulary knowledge and reading comprehension to performance differed as a function of listening conditions and/or group. This correlational approach revealed that the contribution of vocabulary knowledge to dialogue comprehension was larger when spatial location was virtual rather than real, whereas the contribution of reading comprehension skill differed as a function of age and language proficiency rather than as a function of the listening environment. The results indicate that the acoustic scene as well as the cognitive and linguistic competencies of listeners may alter how and when top-down resources are engaged in aid of speech comprehension. Difficult listening conditions may require that attentional resources be deployed in aid of scene analysis and word recognition, causing a shift in the relative weight given to the bottom-up and top-down processes contributing to lexical access in favor of top-down influences. More specifically, a listening scene in which the sound sources are diffused might be harder to parse and provides fewer acoustic cues that could assist stream segregation than when source location is real. As a result it is possible that the bottom-up processes will be sabotaged by intrusions from the competing streams, resulting in a greater engagement of top-down lexical processes in order to maintain word-recognition accuracy. Moreover, the need to engage top-down processes at the level of lexical access may reduce the available pool of attentional resources that could otherwise be applied at later stages of speech recognition and comprehension.
In Experiment 2, we increased the complexity of the comprehension task by adding a third talker to the conversation and varied the degree of listening difficulty by assessing comprehension under three different SNRs (quiet, moderate noise and high noise) in order to examine how the contribution of top-down processes may change as masking increases. Using the same three test groups as in Experiment 1, we asked participants to listen to a conversation between three talkers, with or without real spatial separation between the talkers. As in Experiment 1, the results revealed no significant differences in conversation comprehension among the groups after individual differences in spoken word recognition were compensated for.

As expected, comprehension performance decreased as babble level increased. Analyzing the results using a correlational approach showed that the contribution of individual differences in vocabulary and reading comprehension skills differed among the three groups, and were also modified by the level of the background babble. Vocabulary was found to be the major contributor to performance at medium level of babble whereas reading comprehension contributed to performance in quiet and in high levels of babble. This indicates that as listening becomes difficult, top-down processes involved with lexical access become more fully engaged, until the babble level becomes overwhelming. The relative contribution of vocabulary versus reading comprehension within each group showed that reading comprehension contributes significantly to performance in both young EL1’s and young EL2’s, whereas vocabulary does not in the Young EL1s, and only contributes marginally in young EL2s. However, older EL1’s show a reversed pattern to that found in Young EL1s. In older adults, age-related hearing deterioration may require a greater allocation of attentional resources to basic auditory processes to facilitate stream segregation and lexical access. In both studies, reading comprehension accounted for a significant portion of the variance in speech comprehension in younger EL1s but not in older EL1s. In general, reading comprehension accounted for less of the variance in
speech comprehension in young EL2 than in young EL1s (the correlation between reading comprehension and speech comprehension for EL2s was not significant in Experiment 1 whereas it was significant in Experiment 2). These findings support the idea, suggested by Experiment 1’s results, that speech comprehension is a dynamic process in which the pattern of top-down versus bottom-up contribution is modulated by the nature of the auditory scene, and the listeners’ linguistic status.

In Experiment 3, we investigated how the perceived compactness of sound sources may affect speech recognition. This experiment allowed us to further investigate how the auditory scene affects comprehension of information presented in conversations, and more specifically to examine whether a contrast between a compact target voice and diffuse speech, babble, and noise maskers could provide release from masking. This investigation was prompted by the observation that modern amplification systems routinely alter the degree to which a sound source appears to be diffuse or compact. As in Experiment 1 and 2, three groups of listeners were asked to repeat short semantically anomalous sentences played in either speech spectrum noise, 12-talker babble or competing speech, when there was either a contrast between the diffuseness of the target and masker sound sources (compact target versus diffused masker), or not (compact target and compact masker). Results showed a significant release of masking in each of the three groups when there was a contrast in diffuseness between the target sound source and the masker sound source, under all masking conditions. The release from masking was significantly greater when the masker was speech compared with when it was noise. In addition, individual differences in speech recognition were found to be significantly related to individual differences in vocabulary and reading comprehension in young EL2s, moderately significant in older EL1s, but not significant at all in young EL1s. This implies that young EL2s and older EL1s both use top-down processes to compensate for the distortion of the acoustic information in the target
voice, while young EL1s do not seem to engage higher-order processes to achieve speech recognition.

These three experiments provide evidence that supports the thesis that the manner in which speech comprehension is achieved is modulated by both the characteristics of the acoustic scene, the age of the listener, and his or her linguistic competence. Moreover, while all three experiments support the claim that older EL1s and young EL2s experience greater difficulties when listening in the same settings as the younger EL1, Experiments 1 and 2 demonstrate that these difficulties can be minimized, if not completely taken care of, by adjusting the SNR alone.

Remaining questions and future directions

Although the experiments conducted as part of this dissertation teach us several things about speech recognition and comprehension in naturalistic settings, there is much more that needs to be done. We would like to suggest several future studies that could provide additional information regarding the processes involved in the complex task of comprehending speech in less than ideal auditory scenes.

First, although the tasks in Experiments 1 and 2 are more similar to what listeners are often required to do when communicating in real life situations than are the commonly used speech recognition tasks, there are a few additional steps that can be taken when designing future experimental tasks to better reflect the complexity of daily tasks. For example, in Experiments 1 and 2, listeners did not need to take part in the conversation. All they had to do was to listen to what was being said so that they could answer questions about what they had heard. It would be interesting to design a study in which the listener is required to take an active part in the conversation or at least generate an opinion based on the information provided in the conversation. For example, when using a play such as “Three sons” in which each of the three
character tells about his relationship with his father, it could be interesting to implant questions directed to the listener, such as “what about you and your dad?” A couple of alternative continuances could be recorded in advance, and selected based on the participant’s response. If, in the current example, the participant describes an overall positive relationship, he or she will receive a feedback appropriate to their response such as “that’s nice” versus “father-son relationships are often complex” if the respond was overall not positive. These additional demands would better reflect processes such as integrating the new information with the listener’s own knowledge and perspective on the topic. Moreover, recording the participants’ responses would allow us to linguistically analyze their production, which could provide additional information regarding the depth of processing and the level of engagement.

In order to assess the contribution of higher cognitive and linguistic processes to speech comprehension, participants were asked to complete the Mill Hill and Nelson-Denny tests. This allowed us to conduct a regression analysis to evaluate the effects of individual differences in vocabulary and reading comprehension on speech comprehension performance. However, there are several tests and tools (e.g., Stroop test, digit span) that tap other skills and processes involved in speech comprehension (e.g., working memory, inhibition, lexical retrieval and more). Including other tests and tasks in future studies could potentially shed more light on the contribution of these skills and processes they tap to conversation comprehension. Moreover, in light of the significant main effect of the masker, and the interactions between masker and group found in Experiment 3, it would be interesting to repeat experiments 1 and 2 using maskers other than babble. Doing so could provide additional information regarding how different levels of informational and energetic masking, affect the complex and dynamic balance of bottom-up processes versus top-down processes contribution.
In addition, in Experiments 1 and 2, the location of each of the talkers remained stationary throughout the play. It is possible that this spatial consistency assisted the listeners in following, correctly identifying, and allocating the information to each of the talkers. Since in real-life situations it is quite common that both the target talkers and/or other competing sound sources (such as other irrelevant talkers) will change their position during the time of the conversation, it could be of value to conduct a multi-talker speech comprehension study using a more spatially dynamic auditory scene. Another change which could be applied to increase the auditory scene complexity has to do with the consistency of the background noise presented. Prior evidence implies that older adults experience a decline in the ability to quickly establish stream segregation (Ben-David et al., 2012). This important finding requires further investigation in order assesses the role it might be playing in the difficulties older adults experience in real-life communication. It could be of value to repeat Experiment 1 or 2 with the background noise suddenly appearing and disappearing, rather than present it constantly throughout the speech comprehension task as we have done in the experiments reported.
References


Ezzatian, P., Avivi, M., & Schneider, B. A. (2010). Do nonnative listeners benefit as much as native listeners from spatial cues that release from speech masking? *Speech Communication, 5*, 919-929.


Jory, J. (2001). Heads. In M. B. Dixon, A. Wegener & S. Moulds (Eds.), 30 ten-minute plays: 3 actors; From actors theatre of Louisville's national ten-minute play contest (pp.82-91). Hanover, NH: Smith and Kraus.


doi:10.3389/fnsys.2013.00031


Appendices

Appendix A: Evaluating the contribution of vocabulary knowledge and reading comprehension to R-SPIN SNR-50 thresholds and average percentage of questions correctly answered in Experiment 1

We first checked to see whether vocabulary and/or reading comprehension was related to the average percentage of questions correctly answered by the 72 participants in the experiment. In this and the other tests conducted here we centered the vocabulary and reading comprehension scores within each of the six groups by subtracting the mean score of a group from each of the measures in the group. The six groups were: the 12 younger EL1 listeners in the separation condition; 12 younger EL1 listeners in the no-separation condition; 12 young EL2 listeners in the separation condition; 12 young EL2 listeners in the no-separation condition; 12 older EL1 listeners in the separation condition; and 12 older EL1 listeners in the separation condition. We also centered the average percentage of questions answered correctly in each of these six groups. Centering in this fashion allowed us to evaluate the relative contribution of vocabulary to performance within each group of participants. A regression analysis found a significant correlation between both vocabulary and performance ($r = .38, p < .001$) and reading comprehension and performance ($r = .32, p = .007$).

After finding that both vocabulary and reading comprehension were related to average percentage of questions correctly answered, we then investigated whether the relationship between the two language measures and the dependent variable interacted with one or more of the three factors. Specifically, we tested three hypotheses concerning the slopes of the lines relating these two measures to percent correct. Before conducting this analysis, we centered the percent correct scores within each of the 12 conditions (2-Separations $\times$ 3 Groups $\times$ 2 Types of Spatial Location) in order to be able to compare the relative contribution of vocabulary and reading comprehension across these 12 conditions. Because both of the measures of language ability and the dependent variable were centered in all conditions and groups, the linear relationship between the individual measures and the dependent variable in a condition was
specified by a single parameter, namely the slope of the line relating the measure to performance. Hence, the full model is specified by

\[ Y_{i,j,k,m} = A_{i,j,k} V_{i,j,m} + B_{i,j,k} R_{i,j,m} + e_{i,j,k,m} \quad (A1) \]

where \( Y_{i,j,k,m} \) is the centered percent correct score in the \( ith \) level of Separation (voices separate versus co-located), \( jth \) Group (younger EL1 listeners, older EL1 listeners, young EL2 listeners), \( kth \) Type of Location (real versus virtual), of the \( mth \) individual in Group \( j \) and Separation Condition \( i \). \( V_{i,j,m} \) and \( R_{i,j,m} \) are the centered vocabulary and reading scores for the \( mth \) individual from Group \( j \) and Separation Condition \( i \), respectively, and the \( e_{i,j,k,m} \) are assumed to be random normal deviates with mean = 0, and standard deviation = \( \sigma \).

To test whether the relationships between the vocabulary and reading comprehension measures and the dependent variable were independent of the Type of Location (real versus virtual), we defined and fit a model in which

\[ Y_{i,j,k,m} = A_k V_{i,j,m} + B_k R_{i,j,m} + e_{i,j,k,m} \quad (A2). \]

This model allows for the slopes relating the two language measures to the dependent variable to differ only between the real and virtual spatial location conditions. The best-fitting least-squares parameters of this model are the \( a_k \) and the \( b_k \). We then tested the null hypothesis that \( A_{k=1} = A_{k=2} \) after adjusting the dependent measure for the estimated contribution of reading comprehension, \( R \), to performance. In this
adjusted model the dependent variable is $Y_{R,i,j,k,m} = Y_{i,j,k,m} - b_k R_{i,j,m}$ This null hypothesis was rejected ($F[1,142] = 4.70, p = .03$), indicating that the relationship between the centered vocabulary scores and the dependent variable differed between the real and virtual spatial location conditions. However, when the dependent variable was adjusted for the contribution of vocabulary to performance ($Y'_{V,i,j,k,m} = Y_{i,j,k,m} - a_k V_{i,j,m}$), the null hypothesis that $B_{k=1} = B_{k=2}$ could not be rejected ($F[1,142] < 1$).

To test whether the relationship of both vocabulary and reading comprehension to performance differed between the no-separation and separation conditions, we first averaged over the Within-Subject factor to arrive at a full Between-Subjects model

$$\bar{Y}_{i,j,m} = A_{i,j} Y_{i,j,m} + B_{i,j} R_{i,j,m} + \bar{\varepsilon}_{i,j,m} \quad (A3).$$

We then defined a model in which

$$\bar{Y}_{i,j,m} = A_{i} V_{i,j,m} + B_{i} R_{i,j,m} + \bar{\varepsilon}_{i,j,m} \quad (A4).$$

This model allows for the slope relating the language measures to the dependent variable to differ only between the situations in which the two voices were separated versus when they were co-located. We then tested the null hypotheses that $A_{i=1} = A_{i=2}$ after correcting for the contribution of reading comprehension to performance in the same fashion as described above. We also tested the null hypothesis that $B_{i=1} = B_{i=2}$ after correcting for contribution of vocabulary to performance. Neither null hypothesis could be rejected ($F[1,70] < 1$ for $A_{i=1} = A_{i=2}$, and $F[1,70]= 1.42, p = .24$ for $B_{i=1} = B_{i=2}$.
To test whether the contribution of vocabulary and reading comprehension to the dependent variable differed among the three groups, we first averaged over the Within-Subject factor and defined a model in which

\[ \bar{Y}_{i,j,m} = A_j V_{i,j,m} + B_j R_{i,j,m} + \varepsilon_{i,j,m} \]  

(A5).

This model allows for the slopes relating the two language measures to the dependent variable to differ among the three groups (younger EL1 listeners, older EL1 listeners, young EL2 listeners). We then tested the null hypotheses that \( A_{j=1} = A_{j=2} = A_{j=3} \) after correcting for reading comprehension, and

\[ B_{j=1} = B_{j=2} = B_{j=3} \]  

after correcting for vocabulary. The null hypothesis that \( A_{j=1} = A_{j=2} = A_{j=3} \) could not be rejected \((F[2,69] < 1)\), but the null hypothesis that \( B_{j=1} = B_{j=2} = B_{j=3} \) was rejected \((F[2,69] = 6.23, p < .01)\). Because the relationship between the reading comprehension score and percent correct (adjusted for the contribution of vocabulary) differed across the three groups, we tested 3 sub-hypotheses: 1) younger EL1 listeners’ slope = young EL2 listeners’ slope; 2) younger EL1 listeners’ slope = older EL1 listeners’ slope; and 3) young EL2 listeners’ slope = older EL1 listeners’ slope. Only the difference between younger EL1 listeners’ and young EL2 listeners’ slopes was significant \((F[1,69] = 12.37, p < .01)\) at the .05 level after applying a Bonferroni correction.

It should be pointed out that we obtained exactly the same pattern of results when we independently examined the contributions of reading comprehension and vocabulary to performance, that is, without correcting for the effect of the other language variable on performance.

Recall that R-SPIN SNR-50 thresholds were determined in two conditions: (1) when the precedence effect was used to determine the locations of the voices and babble (virtual location), and (2) when the voices and babble were played over individual loudspeakers (real location). We examined whether either the centered vocabulary measures or the centered reading comprehension scores were
related to both sets of centered R-SPIN SNR-50 thresholds. In conducting these tests, we eliminated the young EL2 individual in the no-separation group whose R-SPIN threshold for the virtual location condition was more than 3 standard deviations above the mean of that individual’s group. None of these four correlations approached significance (p > .2 in all four cases).
Appendix B: Evaluating the contribution of vocabulary knowledge and reading comprehension to performance in the quiet conditions of the Murphy et al. (2006) study

The same vocabulary and reading comprehension measures taken on the participants in this study were also taken on the 96 participants (48 young, 48 old) in the Murphy et al. (2006) study, which used the same two-person plays. Four experiments (12 younger and 12 older participants in each experiment) were conducted in both quiet and babble. In this analysis, the dependent variable was the percentage of questions answered in the quiet part of all four experiments. The vocabulary, reading comprehension, and number of questions answered correctly were first centered in each of the four experiments for the two age groups. The full model in the analysis of these data was

\[ Y_{i,j,m} = A_{i,j} V_{i,j,m} + B_{i,j} R_{i,j,m} + e_{i,j,m} \quad (A6) \]

where \( Y_{i,j,m} \) is the centered average percentage of questions correctly answered by the \( m \) participants in the quiet conditions for Age Group \( j \) of Experiment \( i \), and \( V, R, A, B \) and \( e \) are defined as above. We then defined and fit a model in which the \( A \) and \( B \) coefficients were the same for all four experiments and differed only with respect to the Age Group to which the participants belonged. Specifically, we defined and fit the model

\[ Y_{i,j,m} = A_j V_{i,j,m} + B_j R_{i,j,m} + e_{i,j,m} \quad (A7) \]

We then tested and failed to reject the null hypothesis that \( B_{j=1} = B_{j=2} \) after adjusting for the contribution of vocabulary to performance as described above (\( F[1,94] < 1 \)). Least Squares regressions of the adjusted percentage correct scores against reading comprehension for both the younger and older
participants were highly significant

\( (slope_{young} = 0.90, r_{young} = .47, p = .0006; \ slope_{old} = 0.98, \ r_{old} = .62, p < .0001) \)

After adjusting the centered average percentage of questions correctly answered for the contribution of reading comprehension, there was no evidence of any relationship of the dependent variable to vocabulary

\( (slope_{young} = 0.04, r_{young} = .01, p > .5; \ slope_{old} = 0.72, \ r_{old} = .21, p > .15) \).
Appendix C: Sample of Trialogue and Comprehension Questions

Characters:
Michael
Steve
Dave

Three men sit on chairs in the airport. Steve is reading the paper. Dave is dozing. Michael is watching Steve.

Michael: Do you mind if I borrow that newspaper?
Steve: Huh?
Michael: The paper - Can I have a look?
Steve: If you want it, here you go.
Dave: Wah! What? Did I miss the flight? I was only taking a quick nap.
Michael: Easy! Sorry, I was just borrowing the paper here. The flight’s still delayed sir.
Dave: Hey, could I see the sports pages? I’m going out of my mind with boredom just waiting.
Michael: Here you go, but let me know how the Laker’s did.
Dave: You a fan?
Michael: Nope, but my dad is just about their biggest. First question he’ll ask when I step off the plane, “See the game son?” It’ll be nice to at least know the score in advance for once.
Dave: So you’re going home to visit your folks?
Michael: Just my dad. Mom passed away two years ago. How about you?
Dave: What?
Michael: Visiting folks or off on business?
Dave: A little of both. I’m going to my father’s funeral.

[Pause.]

Michael: I’m sorry. Bit tactless of me, going on about my dad.
Dave: You weren’t to know. Besides, I’m not that affected by the actual funeral. Dad’s been ill for years, kept telling us he should’ve died years ago. Said this was his time off for good behavior.

[Michael and Dave laugh together.]

Michael: You’ve got brothers or sisters then? You said “Dad kept telling us”.

Dave: Yeah, I’m the youngest of three. I’m the one who got away, left the old home town and moved to the big smoke. John, my eldest brother, went to work for Dad in the family business and Zeke became apprenticed with the local carpenter and married the minister’s daughter.

Steve: Zeke? [Sarcastic.]

Dave: Short for Ezekiel. He spent his whole childhood being jealous of me and John for having normal names, then he calls his first two kids Zephyr and Keanu.

Steve: They’re both names for types of wind.

Dave: Oh, hey no offence intended. You’re not a Keanu, are you?

Steve: No, I’m… [relenting his anger] I’m Steve.

Dave: I’m Dave.

Michael: I’m Michael. It’s nice to meet you fellows.

Dave: So do you know why they haven’t boarded us on the plane yet?

Steve: You didn’t hear the announcement? There’s going to be a 40 minute delay because of a shortage of cabin crew on all Mid-western flights.

Michael: The wonders of modern technology. We no longer have to crawl over the surface of the earth in slow moving, badly polluting vehicles, arriving at our destinations cramped and irritable.

Dave: No, we can get cramped and irritable before we even leave.

[Dave and Michael laugh again.]

Dave: I actually really like the idea of a long road trip. Dad did it in his teens and he raved about seeing the real America.

Michael: Don’t tell me - He went after reading “On the Road”?

Dave: OK, I won’t tell you. I’ll tell you this though, I read that book over three days a few years ago and when I put it down…. God, I thought, “If I spend my life looking, I’ll never find a bigger pile of crap than that.”

[Michael laughs again.]

Questions

1. What team’s score is Michael interested in?
   a. Raptors
   b. Bulls
   c. Lakers
d. Knicks

2. What is Dave’s brother’s name?
   a. Ezekial.
   b. Keanu.
   c. Zephyr.
   d. Kerouac.

3. What book do the men discuss?
   a. The Real America.
   b. Road Through America.
   c. Wide Open Road.
   d. On the road.

4. Why is the plane delayed?
   a. Bad weather.
   b. A shortage of cabin crew.
   c. A shortage of ground crew.
   d. A mechanical problem.

Appendix D: Evaluating the contribution of vocabulary knowledge and reading comprehension to the average percentage of questions correctly answered in Experiment 2

We first checked to see whether vocabulary and/or reading comprehension was related to the average percentage of questions correctly answered by the 35 participants in the experiment. In this and the other tests conducted here we first centered the vocabulary and reading comprehension scores within each of the three groups: the 12 younger EL1 listeners; 11 young EL2 listeners; and 12 older EL1 listeners. We also centered the average percentage of questions answered correctly in each of these three groups. Centering in this fashion allowed us to evaluate the relative contribution of vocabulary to performance within each group of participants. A regression analysis found a significant correlation between both vocabulary and performance ($r = .39, p < .001$) and reading comprehension and performance ($r = .45, p < .001$).

After finding that both vocabulary and reading comprehension were related to average performance, we then investigated whether the relationship between the two language measures and the dependent variable was altered by the levels of a factor, the degree of separation among the talkers (co-located versus spatially separated) and by the group to which a participant belonged. In conducting this analysis, we first centered the percent correct scores within each of the 18 conditions (2-Separations $\times$ 3 Groups $\times$ 3 Noise Levels) in order to be able to compare the relative contribution of vocabulary and reading comprehension to listening comprehension. Because both of the measures of language ability and the dependent variable were centered in all conditions and groups, the linear

---

3 One young EL2 subject was removed from this analysis because his or her data for the high noise condition versus the quiet conditions differed by 5 standard deviation units from the mean performance of the remaining participants. See Method.
relationship between the individual measures and the dependent variable in a condition was specified by a single parameter, namely the slope of the line relating the measure to performance. Hence, the full model is specified by

\[ Y_{i,j,k,m} = A_{i,j,k} V_{j,m} + B_{i,j,k} R_{j,m} + e_{i,j,k,m} \quad \text{(A1)} \]

where \( Y_{i,j,k,m} \) is the centered percent correct score in the \( i \)th level of Spatial Separation (voices separate versus co-located), \( j \)th Group (younger EL1s, older EL1s, young EL2s), \( k \)th Noise Level (Quiet, Medium Babble, High Babble), of the \( m \)th individual in Group \( j \). \( V_{j,m} \) and \( R_{j,m} \) are the centered vocabulary and reading scores for the \( m \)th individual from Group \( j \), and the \( e_{i,j,k,m} \) are assumed to be random normal deviates with mean = 0, and standard deviation = \( \sigma \). \( A_{i,j,k} \) is the slope of the line relating the centered dependent variable, \( Y_{i,j,k,m} \), to the Vocabulary score of the \( m \)th individual in group \( j \), separation condition \( i \), and background level \( k \), and \( B_{i,j,k} \) is the slope of the line relating the centered dependent variable, \( Y_{i,j,k,m} \), to the Reading Comprehension score of the \( m \)th individual in the same condition. Note that in this full model different slopes are fit for each of the 18 combinations of separation, noise level, and participant group.

To test whether the relationships between the vocabulary and reading comprehension measures and the dependent variable were independent of the particular condition in which a
participant was tested, we tested the null hypothesis that all the $A_{i,j,k} = A$, and $B_{i,j,k} = B$, for all values of $i, j, k$. This null hypothesis was rejected ($F[34, 174] = 1.658, p = .019$). Hence we concluded that the relationship of these two variables (vocabulary and reading) to listening comprehension depends on some combination of the three factors (separation versus colocation, noise level, and/or participant group).

To identify which combination of factors affected the relationship between comprehension and reading and vocabulary, we then tested the null hypothesis that the slopes of the lines relating the dependent variable to performance were independent of the group and noise level in which participants were tested, i.e., we tested the null hypothesis that $A_{i,j,k} = A_i$ and $B_{i,j,k} = B_i$ for all $j, k$. This hypothesis was also rejected ($F[32, 174] = 1.628, p = .026$). This indicates that the relationship between the two measures of linguistic competence and listening comprehension differs among some of the combinations of noise level and group of participants.

We also tested whether the relationship between the dependent variable and the vocabulary and reading comprehension measures differed when we assumed that the slopes were independent of noise level and separation within each group of participants but were allowed to vary across groups. Specifically we tested and failed to reject the null hypothesis that $A_{i,j,k} = A_j$ and $B_{i,j,k} = B_j$ for all $i, k$ ($F[30, 174] = 1.149, p = .284$). To investigate the ways in which the relationships between the dependent variable and the vocabulary and reading comprehension measures differed within each group of participants we fit the following model to the data.
To evaluate the relative contribution of vocabulary versus reading comprehension within the young EL1s, we tested two null hypotheses: 1) $A_{j=1} = 0$ and 2) $B_{j=1} = 0$, where the subscript 1 identifies the young EL1s. We failed to reject the null hypothesis that $A_{j=1} = 0$ ($F[1,70] = 1.187, p = .280$), indicating that the contribution of vocabulary to performance was negligible for young EL1s.

However, the null hypothesis that $B_{j=1} = 0$ was rejected ($F[1,70] = 32.1805, p < .001$). Hence, we can conclude that reading comprehension contributes significantly to performance for young EL1s.

Similar tests of the same two null hypotheses for the young EL2 listeners found a similar pattern of results. We failed to reject that null hypothesis that $A_{j=2} = 0$ ($F[1,64] = 2.485, p = .120$), but did reject the null hypothesis that $B_{j=2} = 0$ ($F[1,64] = 6.626, p = .012$). Hence for young adults (both EL1s and EL2s), reading comprehension contributes significantly to performance whereas vocabulary does not. However the pattern is reversed for older EL1s. For older listeners, the hypothesis that $A_{j=3} = 0$ was rejected ($F[1,70] = 4.513, p = .037$), whereas the hypothesis that $B_{j=3} = 0$ was not ($F[1,70] = .006, p = .939$). Hence for older adults, vocabulary is significantly correlated with performance whereas reading comprehension is not. Figure 12 indicates the extent to which these two individual measures contribute to performance after the effect of the other measure had been taken into account in each of the three groups of participants. The ordinate on the
left-hand side of the figure is defined as

\[ y_{V,i,j,k,m} = y_{i,j,k,m} - b_j R_{j,m} \]

where \( b_j \) is the least square estimate of \( B_j \). The ordinate in the right panel is

\[ y_{R,i,j,k,m} = y_{i,j,k,m} - a_j V_{j,m} \]

where \( a_j \) is the least square estimate of \( A_j \).

An examination of relationship between vocabulary scores and listening comprehension (after adjustment has been made for reading comprehension) suggest that the slope of the line varies across participant groups. The model when the listening comprehension scores are adjusted for the contribution of reading comprehension is

\[ y_{V,i,j,k,m} = A_j V_{j,m} + e_{i,j,k,m} \]

The null hypothesis that \( A_1 = A_2 = A_3 \) was rejected \((F[2,207] = 5.5104, p = .005)\). Posthoc tests indicated that the slopes for the two young groups did differ significantly from one another \((F[1,207] = 5.956, p = .016)\); the slope of the young EL1 participants differed from that of the older EL1 participants \((F[1,207] = 10.891, p = .001)\), but that the slopes of the young EL2 participants did not differ significantly from those of the older EL1 participants \((F[1,207] = 1.270, p = .261)\).

When the listening comprehension scores were adjusted for the contribution of vocabulary in
the model

\[ y_{R,i,j,k,m} = B_j R_{j,m} + e_{i,j,k,m} \]

the null hypothesis that \( B_1 = B_2 = B_3 \) was rejected (\( F[2,207] = 20.808, p < .001 \)). Posthoc tests indicated that 1) the slopes for the young EL1s differed from those of the young EL2s (\( F[1,207] = 17.151, p < .001 \)); 2) the slopes for the young EL1s differed from those of the older EL1s (\( F[1,207] = 41.486, p < .001 \)); and 3) the slopes for the young EL2s differed from those of the older EL1s (\( F[1,207] = 7.328, p = .007 \)). Hence the degree to which reading comprehension contributed to listening comprehension differed significantly among the three groups of participants.

We also tested the null hypothesis that the relationship between the dependent variable and the vocabulary and reading comprehension measures did not differ as a function of separation or participant group within each noise level, i.e., we tested and failed to reject the null hypothesis that

\[ A_{i,j,k} = A_k \text{ and } B_{i,j,k} = B_k \text{ for all } i, j \ (F[30, 174] = 1.368, p = .110). \]

To investigate the ways in which the relationships between the dependent variable and the vocabulary and reading comprehension measures differed within each noise level we fit the following model to the data.

\[ y_{i,j,k,m} = A_k V_{j,m} + B_k R_{j,m} + e_{i,j,k,m} \]

To evaluate the degree to which the relative contribution of reading and vocabulary to performance might differ within each noise level, we tested two null hypotheses at each noise level. Specifically, we tested both \( A_{k=1} = 0 \), and \( B_{k=1} = 0 \), where the subscript 1 stands for the quiet
condition. The null hypothesis that $A_{k=1} = 0$ was not rejected ($F[1,68] = .170, p = .682$, but $B_{k=1} = 0$ was rejected ($F[1,68] = 14.767, p < .001$). Hence the contribution of reading comprehension to performance was significant whereas the contribution of vocabulary to performance was not when participants were listening in quiet. However when the same tests were performed for the participants listening in a medium level of noise, the null hypothesis that $A_{k=2} = 0$ was rejected ($F[1,68] = 9.634, p = .003$), but the null hypothesis that reading comprehension was unrelated to performance ($B_{k=2} = 0$) could not be rejected ($F[1,68] = 1.526, p = .221$). Finally, for participants listening in the high level of babble, the contribution of reading comprehension to performance was significant ($F[1,68] = 5.906, p = .018$) whereas the contribution of vocabulary to performance was not ($F[1,68] = .192, p = .663$). Hence vocabulary is the major contributor to performance at medium levels of babble whereas reading comprehension contributes to performance in quiet and in high levels of babble. Figure 13 indicates the extent to which these two individual measures contribute to performance after the effect of the other measure had been taken into account. The ordinate on the left-hand side of the figure is defined as

$$y_{V,i,j,k,m} = y_{i,j,k,m} - b_k R_{j,m}$$

where $b_k$ is the least square estimate of $B_k$. The ordinate in the right-hand panel is

$$y_{R,i,j,k,m} = y_{i,j,k,m} - a_k V_{j,m}$$
where $\alpha_k$ is the least square estimate of $A_k$.

An examination of the relationship between vocabulary scores and listening comprehension (after adjustment has been made for reading comprehension) suggests that the slope of the line varies across participant groups. The model when the listening comprehension scores are adjusted for the contribution of reading comprehension is

$$y_{V,i,j,k,m} = A_k V_{j,m} + e_{i,j,k,m}$$

The null hypothesis that $A_{k=1} = A_{k=2} = A_{k=3}$ was rejected ($F[2,207] = 10.337, p < .001$). Posthoc tests indicated that the slopes for the quiet and medium noises differ significantly from one another ($F[1,207] = 15.545, p < .001$). Also significant was the difference in slopes between the medium and high noise conditions ($F[1,207] = 15.468, p < .001$). The slope difference between the quiet and high babble conditions was not significant ($F[1,207] < 1$).

When the listening comprehension scores were adjusted for the contribution of vocabulary in the model

$$y_{R,i,j,k,m} = B_k R_{j,m} + e_{i,j,k,m}$$

the null hypothesis that $B_{k=1} = B_{k=2} = B_{k=3}$ could not be rejected ($F[2,207] = 2.870, p = .059$). Hence the manner in which reading comprehension contributed to listening comprehension did not change with listening conditions. However, the contribution of vocabulary did.