ABSTRACT

Title of Dissertation:	EFFECTS OF TALKER FAMILIARITY ON SPEECH UNDERSTANDING AND COGNITIVE EFFORT IN COMPLEX ENVIRONMENTS
	Julie Ilana Cohen, Doctor of Philosophy, 2020
Dissertation directed by:	Professor, Sandra Gordon-Salant Department of Hearing and Speech Sciences
	Douglas Brungart, Ph.D. Chief Scientist Walter Reed National Military Medical Center

The long-term goal of this project is to understand the cognitive mechanisms responsible for familiar voice (FV) benefit in real-world environments, and to develop means to exploit the FV benefit to increase saliency of attended speech for older adults with hearing loss. Older adults and those with hearing loss have greater difficulty in noisy environments than younger adults, due in part to a reduction in available cognitive resources. When older listeners are in a challenging environment, their reduced cognitive resources (i.e., working memory and inhibitory control) can result in increased listening effort to maintain speech understanding performance. Both younger and older listeners were tested in this study to determine if the familiar voice benefit varies with listener age under various listening conditions.

Study 1 examined whether a FV improves speech understanding and working memory during a dynamic speech understanding task in a real-world setting for couples of younger and older adults. Results showed that both younger and older adults exhibited a talker familiarity benefit to speech understanding performance, but performance on a test of working memory capacity did not vary as a function of talker familiarity. Study 2 examined if a FV improves speech understanding in a simulated cocktail-party environment in a lab setting by presenting multi-talker stimuli that were either monotic or dichotic. Both YNH and ONH groups exhibited a familiarity benefit in monotic and dichotic listening conditions. However, results also showed that talker familiarity benefit in the monotic conditions varied as a function of talker identification accuracy. When the talker identification was correct, speech understanding was similar when listening to a familiar masker or when both voices were unfamiliar. However, when talker identification was incorrect, listening to a familiar masker resulted in a decline in speech understanding. Study 3 examined if a FV improves performance on a measure of auditory working memory. ONH listeners with higher working memory capacity exhibited a benefit in performance when listening to a familiar vs. unfamiliar target voice. Additionally, performance on the 1-back test varied as a function of working memory capacity and inhibitory control.

Taken together, talker familiarity is a beneficial cue that both younger and older adults can utilize when listening in complex environments, such as a restaurant or a crowded gathering. Listening to a familiar voice can improve speech understanding in noise, particularly when the noise is composed of speech. However, this benefit did not impact performance on a high memory load task. Understanding the role that familiar voices may have on the allocation of cognitive resources could result in improved aural rehabilitation strategies and may ultimately facilitate improvements in partner communication in complex real-world environments.

EFFECTS OF TALKER FAMILIARITY ON SPEECH UNDERSTANDING AND COGNITIVE EFFORT IN COMPLEX ENVIRONMENTS

by

Julie Ilana Cohen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Sandra Gordon-Salant, Co-Chair
Adjunct Professor Douglas Brungart, Co-Chair
Associate Professor Samira Anderson
Professor Rochelle Newman
Professor Catherine Carr, Dean's Representative

© Copyright by
Julie Ilana Cohen
2020

Dedication

For Annie.

Acknowledgements

I would like to thank my dissertation committee for their guidance and encouragement throughout the many stages of this project. I am grateful to Sandy Gordon-Salant, my advisor and committee co-chair, for her 13+ years of dedication to my education and development as an independent researcher. Your incredibly high standards for conducting research has pushed me to sharpen my skills as a writer and as an analytical thinker. I would like to acknowledge my committee co-chair, Doug Brungart, whose expertise was invaluable in formulating the research questions and methodology. While their research philosophy and mentoring style were not always in agreement with one another, Sandy and Doug were both instrumental in my academic and professional development.

Thank you to the other members of my committee that helped to make this project a success. Samira Anderson, Catherine Carr, and Rochelle Newman have provided excellent feedback, and have always offered me support when it was most needed. Additionally, I would like to thank my "unofficial" committee member, Stefanie Kuchinsky, who provided advice on statistical analyses and interpretations.

I am beyond grateful to my partners in crime, Jaclyn Schurman and Maureen Shader. They have been my sounding board and lunch dates, and have supported me through the highs and lows of the Ph.D. program over the last five years. Your texts and video chats have made the past seven months bearable, and I know I would not have been able to make it to this point without you. I also need to acknowledge my work wife, Danielle Zion, who has been with me on this journey since Day 1.

I would like to thank my family for their support and motivation throughout this process. Mom, Dad, and Wendy, thank you for your love and for always believing in me. Thank you to Annie for believing I could earn a Ph.D. long before I believed it myself, and for constantly reminding me to finish my comps. I would also like to thank all of my extended family. The phone calls, meals, notes of support, and even study dates at the library, have been instrumental to my success.

Most importantly, to Cory and Nora: your love and understanding has helped me through some of the most challenging times of my life. You make me smile every day, and much of my success belongs to you.

This work was supported, in part, by the National Institute on Deafness and Other Communication Disorders (NIDCD) from an institutional training grant (T32DC000046E, Co-PIs: Sandra Gordon-Salant and Catherine Carr), a Student Investigator Research Grant from the American Academy of Audiology Foundation (PI: Julie Cohen; Mentor: Sandra Gordon-Salant), and the Ann G. Wylie Dissertation Fellowship.

Table of Contents

Dedication ii
Acknowledgements iii
Table of Contents v
List of Tables vii
List of Figures viii
List of Abbreviations x
Introduction 1
    Overview 1
    The Cocktail Party Problem 2
        Speech Segregation 2
        Energetic and Informational Masking 3
    Speech Understanding in Noise: Environmental Factors 7
        Real-world vs. Laboratory Environments 7
    Speech Understanding in Noise: Subject Factors 11
        Aging 11
        Aging and Working Memory Capacity 12
    Speech Understanding in Noise: Signal Factors 15
    Talker Familiarity 17
        Trained Familiarity 17
        Incidental Familiarity 19
        Personal Familiarity 20
        Clinical Relevance 23
    Summary and Hypotheses 24
Study 1: Effect of Talker Familiarity in a Real-World Environment 27
    Introduction 27
    Research Questions and Hypotheses 31
    Method 32
        Participants 32
        Stimuli 33
        Procedures 34
    Data Analysis 36
    Results 37
        Sound Level in the Real-world Environment 37
        Familiarity Benefit on the Cafeteria Test 38
        Performance on the N-Back Test 42
    Discussion 46
        Talker Familiarity in a Real-world Environment 46
        Talker Familiarity Benefit on a Test of Working Memory 49
    Conclusions 52
Study 2: Effect of Target Talker and Masker Familiarity on a Dichotic Speech Perception Task 54
    Introduction 54
    Research Questions and Hypotheses 60
    Method 62

Participants 62
Stimuli 63
Procedures 65
Data Analysis 68
Results 69
Monotic Conditions: Target Talker and One Competing Talker 69
Dichotic Conditions: Target Talker and Two Competing Talkers 81
Comparison of Monotic and Dichotic Conditions 85
Discussion 89
Talker Familiarity Benefit for Monotic Conditions 89
Talker Familiarity Benefit for Dichotic Conditions 93
Talker Familiarity Benefit for Monotic vs. Dichotic Configurations 95
Conclusions 96
Study 3: Effect of Talker Familiarity on Working Memory Capacity in a Competing Talker Task 98
Introduction 98
Research Questions and Hypotheses 104
Method 104
Participants 104
Stimuli 105
Procedures 105
Data Analysis 107
Results 108
Speech Recognition Thresholds in the 0-Back Task 108
Proportion Correct Word Recall in the 1-Back Task 112
Discussion 120
Performance on the SRT_{80} Tracking Procedure 121
Performance on the 1-back Test 124
Conclusions 128
General Discussion 131
Real-world Environments 134
Aging 137
Working Memory Capacity and Inhibitory Control 138
Conclusions 141
Bibliography 143

List of Tables

Table 2.1 39

Table 2.2 44

Table 3.1 70

Table 3.2 78

Table 3.3 82

Table 3.4 84

Table 3.5 86

Table 4.1 109

Table 4.2 113

List of Figures

Figure 2.1. 35

Figure 2.2. 38

Figure 2.3. 39

Figure 2.4. 41

Figure 2.5. 42

Figure 2.6. 43

Figure 2.7. 45

Figure 3.1. 64

Figure 3.2. 65

Figure 3.3. 66

Figure 3.4. 71

Figure 3.5. 72

Figure 3.6. 74

Figure 3.7. 75

Figure 3.8. 76

Figure 3.9. 79

Figure 3.10. 81

Figure 3.11. 83

Figure 3.12. 88

Figure 3.13. 89

Figure 4.1. 110

Figure 4.2. 112

Figure 4.3. 115

Figure 4.4. 116

Figure 4.5. 118

Figure 4.6. 120

List of Abbreviations

CRM	Coordinate Response Measure
DV	Dependent Variable
ELU	Ease of Language Understanding Model
FV	Familiar Voice
GLME	Generalized Linear Mixed Effects Model
KEMAR	Knowles Electronic Manikin for Acoustic Research
LE	Listening Effort
LSPAN	Listening Span
MRT	Modified Rhyme Test
NIH	National Institutes of Health
OE	Opposite Ear
ONH	Older Normal Hearing
RSPAN	Reading Span
RSPIN	Revised Speech Perception in Noise Test
RAU	Rationalized Arcsine Units
SNR	Signal-to-Noise Ratio
SSN	Speech Shaped Noise
TE	Target Ear
TMR	Talker-to-Masker Ratio
WMC	Working Memory Capacity
YNH	Young Normal Hearing

Introduction

Overview

Conversing with a partner in a noisy environment, such as over lunch at a busy restaurant or at a party hosted in a tightly packed apartment, requires the listener to use strategies to pick out the important information (i.e. speech) from the cacophony of background noise and distractions (e.g. neighboring conversations and background music). This type of scenario, where the listener must extract the target speech from the background noise by selectively attending to the target and filtering out the other competing sounds, is referred to as the “cocktail party problem” (Cherry, 1953). This process of segregating sounds is critical for speech understanding in noise (Bronkhorst, 2000). Speech understanding ability in noise is dependent on the environment in which an individual is listening (e.g. the type of competing signals, real-world vs. laboratory setting), individual subject factors of the listener (e.g. age, hearing ability, cognitive function), and characteristics of the talker(s) (e.g. gender, and familiarity of the voice). The ability and ease with which a listener can selectively attend to the target speech is modulated by the factors listed above. These factors will be discussed in greater detail in the sections below.

The Cocktail Party Problem

Speech Segregation

Auditory object formation is the process by which a listener perceptually groups a series of sounds (Shinn-Cunningham, 2008). For example, the voice of a man and the honk of a horn are perceived by the auditory system as separate objects. Listeners group and separate these objects based on characteristics such as their similarity in frequency, amplitude modulation, and spatial location (Bregman, 1990). A similar, yet higher level of processing is auditory object selection, defined as the process by which a listener chooses to attend to a particular object. Both the acoustic characteristics of the sound object, as well as attentional filtering, contribute to object selection (Shinn-Cunningham, 2008). Auditory grouping and selection can also occur with complex stimuli, such as a series of words spoken by a particular talker, which span over long stretches of time. Again, features such as the similarity in fundamental frequency, temporal characteristics, and spatial location contribute to a listener's perception that the series of sounds is from the same source, and is thus processed as an "auditory stream."

In a real-world environment, auditory object formation and object selection often occur simultaneously in the presence of competing sounds, such as background noise or competing voices (Shinn-Cunningham, 2008). The listener employs both bottom-up and top-down processes to selectively attend to the desired stream. Cues for object formation overlap with those of object selection, however, they are distinct functions. Stimulus factors such as talker gender, background noise level, and spatial separation of the target from the competing signal, contribute to a listener's ability to segregate speech (Brungart, 2001b; Darwin et al., 2003; Kidd et al., 2005).

One feature that can influence object selection is the familiarity of the voice or personal relevance of the message to the talker. The familiarity of the attended speech signal can enhance focus on the target speech, such as when the listener is trained to identify a particular voice (Nygaard & Pisoni, 1998), or the listener has a personal familiarity (e.g., spouse or close friend) to a voice (Johnsrude et al., 2013). In these instances, the listener is better able to understand the target speech signal. However, the presence of a personally familiar voice or relevant message can cause the listener to select the wrong auditory object. In a study by Moray (1959), the listener's attention was diverted when their name was spoken in the unattended or distractor channel in a dichotic shadowing task. These factors will be discussed in greater detail in the "Subject Factors" section of this chapter.

Energetic and Informational Masking

A listener's ability to attend to a particular speech stream is influenced by the presence of competing signals. These can include environmental noise or speech from a competing talker. The type, spectrum, and level of a competing background signal have strong effects on a listener's ability to understand speech in noise (ANSI, 1969). The presence of the competing signal, in general, reduces audibility of the target signal. According to the Articulation Index Theory, changes in intelligibility for a particular stimulus can be predicted for a group of listeners, given the signal level, noise level, frequency importance function for a particular signal, and hearing sensitivity of the listeners (ANSI, 1969, 2017). However, the masking effectiveness of noise also depends on the modulations in the noise as well as on the informational content of the noise.

Masking that results purely in the reduction of the audibility of the target signal is known as “energetic” masking (Freyman et al., 1999). The use of maskers such as narrow band noise (NBN) and speech shaped noise (SSN) can result in energetic masking effects (Kidd et al., 1998), because the spectrum of the masker overlaps the target in frequency and time. Thus, the effectiveness of an energetic masker is dependent on the amount and extent to which it activates similar regions along the cochlea as the target signal. The greater the overlap, the more effective the masker.

Informational masking has been described as masking that occurs when there is informational content in the masker that can be easily confused with the target signal, usually in addition to energetic masking (Kidd et al., 2008). In the case of speech, informational masking occurs when competing voices contain meaningful linguistic content that could be confused with a target spoken message, thereby causing effects beyond those attributed to energetic masking (e.g., Brungart, 2001b). This includes a single competing talker or multiple competing talkers that provide confusion regarding the target source location. Cues that contribute to the reduction of informational masking include differences in fundamental frequency and gender of the target and competing talkers (Brungart, 2001b; Darwin et al., 2003), the number of competing talkers

(Brungart, 2001b; Brungart et al., 2001), and the spatial separation between the target and competing talkers (Brungart & Simpson, 2002; Freyman et al., 2004).

Brungart and colleagues evaluated the effectiveness of a speech masker when the competing talker was the same or different sex than the target talker (Brungart, 2001b; Brungart et al., 2001). Normal-hearing adults reported the color and number from phrases of the Coordinate Response Measure (CRM) speech corpus, which is a closed-set task where the listener hears a phrase with a call sign followed by a color and number (Bolia et al., 2000). Listeners performed more poorly when the talkers were the same sex as the competing talkers, compared to when they were the opposite sex. This effect was attributed primarily to informational masking, because the error analysis revealed that primarily all number/color word errors were a result of the listener reporting the color or number that was spoken by one of the competing talkers. Thus, the incorrect stream was selected due to confusion of the similarity between the target and competing talkers.

The meaningfulness of the target and competing speech can modulate the amount of informational masking generated. By using competing speech that is time reversed, the frequency and amplitude are preserved while the intelligibility of the signal is removed (Cherry, 1953). Freyman et al. (2001) reported a significant difference in masking effectiveness between forward speech (unprocessed) and time-reversed speech maskers. Speech understanding performance with the time-reversed maskers was similar to performance in the forward speech masker when the target and maskers were perceived to be spatially separated, but significantly better (relative to the forward speech masker) when the target and maskers were co-located. This pattern of performance was also observed when the competing speech was spoken in a different language from the target talker. Previous studies have reported a significant reduction in the effectiveness of a masker when the target talker was a native English speaker and the competing talkers were non-native accented English speakers compared to when the competing talkers were native English speakers (Freyman et al., 2001; Gordon-Salant et al., 2013). Additionally, and most relevant for the proposed research, the familiarity of a talker can modulate the effectiveness of a speech masker. Johnsrude et al. (2013) showed an improvement in speech understanding performance in positive signal-to-noise ratios (SNR) when the competing talker was the voice of the listener's spouse than to a novel voice. This effect will be examined in greater detail in Chapter 3.

Spatial separation of a target and competing sounds can result in a "better ear advantage," and can result in improved auditory object formation (Bregman, 1990). The improvement in speech intelligibility with the spatial separation of the target and competing signals is referred to as spatial release from masking (SRM) (Plomp & Mimpen, 1981). Spatial separation can reduce the effects of informational masking when listening in the presence of competing speech (Freyman et al., 2001; Marrone et al., 2008). Brungart and Simpson (2002) modeled the cocktail party effect in a laboratory setting by separating the target and masker voices under headphones across the two ears. In this design, a target talker and a competing talker were presented to one ear, and a separate competing talker was presented to the ear opposite the target talker, referred to as the unattended ear. They found that a speech masker presented in the unattended ear resulted in poorer speech segregation compared to conditions where no sound was presented to the unattended ear (i.e. monotic presentation). These findings are consistent with those of Moray (1959), which showed that the presence of a competing signal in the unattended ear resulted in reduced stream segregation of the target signal.

Speech Understanding in Noise: Environmental Factors

Real-world vs. Laboratory Environments

Since Cherry's 1953 paper was published, many studies have quantified the effects of factors that can influence speech understanding in noise. The majority of these experiments were conducted in a laboratory setting, and the findings were then used to infer an individual's speech understanding ability in a real-world environment. However, performance in the real-world may not be adequately quantified in a laboratory study. Real-world environments are dynamic – the level of the target talker, the identity of the target talker, the amount of visual cues present, and the location of the target talker can vary from moment to moment. Similarly, the characteristics and locations of competing talkers vary dynamically (or are constantly changing) in a real-world environment. Moreover, the relative intensity of the target vs. background noise can vary dramatically, but signal intensity is typically well-controlled in lab studies.

The noise conditions used in laboratory experiments don't often correspond to what listeners experience in a real-world environment. In laboratory experiments, the SNR may be adjusted to a variety of levels to examine the effects of a particular manipulation on speech understanding. These SNR manipulations often include instances when the noise is significantly more intense than the target speech, which may not be characteristic of typical communication environments (Smeds et al., 2015). For example, in many laboratory studies, SNRs range from -15 to + 15 dB SNR (Bernstein & Grant, 2009; Brungart, 2001b; Freyman et al., 2004). A study by Pearsons et al. (1977) was one of the first to quantify noise levels in real-world environments. Background noise and target speech levels were recorded in homes, schools, hospitals, department stores, trains, and airplanes. They found that the speech level increased linearly with increasing background noise up until 55 dBA, where the background noise then became more dominant. More recently, Smeds et al. (2015) measured the SNR that hearing-impaired listeners experienced in various real-world environments, including several of the environment categories from the Pearsons et al. (1977) study. Smeds and colleagues (2015) found that the average SNR in a moderately loud environment (60-70 dBA SPL) was +5 dB, which was significantly higher than values reported previously by Pearsons et al. (1977). Thus, while laboratory measures include a wide range of SNRs, more recent studies have shown that realistic listening environments fall within a narrower and more positive SNR range.

A restaurant setting is composed of a combination of background speech, environmental noise, and reverberation that directly and indirectly impacts speech understanding (e.g. clattering of dishes, televisions in the background). Several studies have attempted to replicate a real-world restaurant sound scene using auditory spatial simulators. Compton-Conley et al. (2004) recorded the background noise at a lively restaurant using two methods. First, they placed in-the-ear directional hearing aids on a Knowles Electronic Manikin for Acoustic Research (KEMAR) and made recordings from KEMAR in an actual restaurant. Recordings in the restaurant were also simultaneously made using a spatialized 8-microphone array, and were played back and recorded from KEMAR in a sound booth (Revit et al., 2002). The directional recordings were presented via headphones to normal-hearing listeners, and the results showed no difference in performance between conditions where the noise was recorded through KEMAR in the restaurant vs. through the spatialized microphone array. A more recent study by Best et al. (2015) compared performance for hearing aid users on a speech-in-noise perception task where the stimuli were presented from several speakers in a sound booth or in a simulated cafeteria soundscape. Performance was poorer in the simulated cafeteria environment; however, the largest hearing aid benefit was measured in the same simulated environment.

Listening in a real-world environment involves the processing of both auditory and visual stimuli. Listeners obtain a significant benefit of a visual signal when listening in the presence of noise (Grant & Seitz, 2000; Walden et al., 1993), and more so when the competing signal is speech (Helfer & Freyman, 2005). In a real-world environment, the visual image is most often not just the target talker, but a dynamic scene that involves both near and far-field visual distractions. These distractions can potentially result in a drain on attention, which can negatively impact speech understanding, particularly for older adults (Cohen & Gordon-Salant, 2017). Additionally, the visual scene is dynamic and can frequently change. The stagnant nature of laboratory auditory-visual speech perception measurements does not adequately capture the dynamic aspects of real-world environments (e.g. constantly changing auditory and visual inputs). As a result, findings in the laboratory may under-estimate the speech understanding difficulties of individuals in everyday, complex environments.

Social factors that are present during a normal conversation are not easily replicated in a laboratory setting. These factors include motivation and social constraints. Listeners' motivation is not easily controlled in a laboratory setting – they are participating in a study and will typically receive some form of compensation. Methods to manipulate motivation, based on a changing reward system, is one alternative to this problem. However, in a real-world setting, the listener may be more or less motivated to attend to the target speech due to the information being transmitted or the partner with whom they are communicating. This motivation may result in increased attentional resources, which can benefit speech understanding. Social constraints of an environment can result in the modulation of the talker's voice level. In a laboratory setting, the level of the target is fixed or systematically adjusted to quantify speech understanding in noise at a target performance level. In a real-world environment, the target voice level changes dynamically based on the particular environment. This phenomenon of the talker self-adjusting the level of their voice in the presence of noise is known as the Lombard Effect (Lombard, 1911). However, at some point, an increase in voice level is not appropriate, either due to physical constraints on the talker or to saturation of sound in the environment. These constraints can negatively impact speech understanding when the target voice exceeds social norms prior to reaching a level of optimum speech intelligibly.

Speech Understanding in Noise: Subject Factors

Aging

Older adults have greater difficulty understanding speech in noisy environments compared to younger adults. Plomp and Mimpen (1979) showed that although older adults perform adequately in a quiet environment, their speech understanding ability declined with increasing background noise. This decrement in speech understanding was more pronounced for older adults than younger adults. Additionally, Dubno et al. (1984) administered an adaptive procedure to determine the SNR corresponding to 50% correct performance in babble for normal and hearing-impaired adults. The results indicated that despite comparable performance in quiet, older normal-hearing listeners' ability in noise was reduced compared to that of the younger normal-hearing adults. More recent studies have examined the effect of informational masking on older adults. In the presence of two competing talkers, older adults were less able than younger adults to ignore the competing speech when the target and competing talkers were presented from the same location (Helfer & Freyman, 2008).

Differences in younger and older adults' ability to understand speech in noise can be partly attributed to differences in hearing sensitivity. Adults show incremental changes in hearing sensitivity across each decade of life, with more hearing loss in the high-frequency regions (Cruickshanks et al., 1998; Pearson et al., 1995). Hearing loss significantly reduces a listener's ability to understand speech in the presence of background noise (Dirks et al., 1982; Dubno et al., 1984; Humes & Dubno, 2010). Older adults with hearing loss show reduced speech perception compared to older normal-hearing listeners. This has been demonstrated with manipulations that include competing speech (Dubno et al., 1984), reverberation (Gordon-Salant & Fitzgibbons, 1993), and time compression (Gordon-Salant et al., 2007). These measures relate to the listener's ability to extract frequency and temporal cues of speech, and their ability to ignore irrelevant speech information. However, differences in speech understanding observed between younger and older adults with comparable hearing sensitivity are still present, and are attributed to factors beyond the peripheral auditory system.

Aging and Working Memory Capacity

In complex auditory environments, speech understanding is affected by both auditory and cognitive factors (Arlinger et al., 2009; Humes et al., 2012). While a listener's hearing sensitivity is correlated with speech understanding in quiet, there is a poor correlation between hearing sensitivity and speech understanding in noise (Plomp, 1986). This suggests that other factors beyond the audiogram contribute to speech understanding. As discussed previously, auditory temporal processing is a critical function for speech understanding in noise, and can contribute to age differences on speech understanding (Dubno et al., 1984; Gordon-Salant et al., 2008). Listeners' cognitive processing capacity, particularly working memory and inhibition, also contributes to their ability to understand speech in noise (Ronnberg et al., 2010; Salthouse, 1996).

Working memory (WM) is a limited capacity system that allows for the storage and manipulation of information in short term memory (Baddeley, 2000; Baddeley & Hitch, 1974). This system is critical for speech understanding in complex environments. The Ease of Language Understanding (ELU) model (Ronnberg et al., 2019; Ronnberg et al., 2013; Ronnberg et al., 2008) suggests that WM is strongly dependent on top-down processes such as language processing, when listening in a noisy environment. The model suggests that when bottom-up constraints such as hearing loss, or the top-down function of reduced working memory capacity are present, speech recognition performance may be reduced (Rudner & Ronnberg, 2008).

Working memory capacity can be measured through a "span" test where listeners must make mental judgements about an increasing number of stimulus items and then recall those items later. The higher the number of items recalled correctly (i.e., span), the larger the working memory capacity. Daneman and Carpenter (1980) developed a test where the participants read sentences and had to make a semantic judgement about each sentence. At the end of a series of sentences, they were prompted to recall the final word from each of the previous sentences. The number of sentence-final words to be recalled increased following correct responses to derive a "span," defined as the number of single words recalled correctly. This reading span (R-SPAN) test was adapted by Pichora-Fuller et al. (1995) to present sentences in an auditory modality, thus making it a more domain-specific test for measuring auditory working memory capacity. Listeners were instructed to make a semantic judgment about each sentence presented, and then recalled the final word of each sentence they heard in a particular set of sentences. The number of sentences in the set was increased to determine a listener's maximum sentence set size and was calculated as their listening span (LSPAN) score. They found that older adults had lower listening spans than younger adults.

Working memory capacity has been shown to contribute to a listener's ability to understand speech in noise. This effect has been measured using a 1-back test paradigm. Schurman et al. (2014) implemented a 1-back procedure where an initial sentence was presented, followed by a second sentence, and then the participants were prompted to repeat the first sentence. Listeners were also tested on an immediate sentence recall test, in which the SNR was adapted to yield an 80% correct level of performance. This SNR was then used during the 1-back test. Older normal-hearing adults required a higher SNR to achieve 80% intelligibility than younger adults on the immediate recall task. On the 1-back test, older adults were poorer at recalling the target sentence than the younger adults when tested at SNRs that corresponded to their 80% intelligibility performance.

Additionally, listener performance on the 1-back test was highly correlated with performance on the LSPAN test. This finding suggests that performance in real-world speech understanding tasks that require a listener to hold information in memory are highly related to WM capacity. In a study by Gordon-Salant and Cole (2016), younger and older normal-hearing adults with higher WM capacity, as measured through their LSPAN score, had better sentence recognition performance in noise compared to adults with lower working memory capacity. These studies suggest that WM capacity is critical for speech understanding in noise, even among adults with normal-hearing. Moreover, the typical age-related decline in WM capacity appears to contribute to older adults' difficulty listening in background noise.

Speech Understanding in Noise: Signal Factors

Listeners use various cues to extract target speech from background noise. They use cues such as context (van Rooij et al., 1989), gender (Brungart, 2001b), spatial separation (Freyman et al., 1999), and talker familiarity. Fundamental frequency (F0) has been identified as a strong cue for segregation of different talkers (Darwin et al., 2003). Male and female voices have characteristically different F0s, which can positively impact speech segregation, depending on the combination of talkers in a scene. Darwin et al. (2003) examined fundamental frequency (F0), vocal tract length (VT), and the combination of these two features (F0+VT) on listener performance on the CRM. The results showed that speech intelligibility performance was sensitive to shifts in both F0 and F0+VT, with larger differences in F0+VT across the target and competing talkers resulting in better speech understanding performance.

The rate of speech of a particular talker can influence the available speech information in a signal, and thus impact overall speech understanding for younger and older normal-hearing adults. Speech signals can be manipulated through time compression and expansion, which change the timing characteristics (i.e., speech rate) of the signal to reduce or extend the time of the signal while still preserving the quality of the speech signal (Gordon-Salant & Fitzgibbons, 1993; Vaughan et al., 2002). Speech understanding ability for both younger and older adults declines with increased rate of speech, however older are more negatively impacted by time-compressed or rapid speech. Additionally, the use of clear speech has been shown to improve speech understanding in noise as it slows the rate of speech, increases the consonant-vowel ratio (CVR), and increases temporal modulations of the stimuli (Krause & Braida, 2002; Picheny et al., 1985).

The use of accented speech compromises speech recognition performance, especially in noise. Gordon-Salant et al. (2010) measured speech recognition ability for Spanish-accented English by younger and older normal-hearing listeners, as well as older hearing-impaired listeners. The results showed a reduced speech understanding for the Spanish-accented speech for all listeners, relative to unaccented speech. Subsequent studies also showed that all three listener groups exhibited poorer speech recognition in noise for accented English compared to unaccented English, and that listeners were less able to extract cues for speech stream segregation in noise when the talker had a foreign accent (Gordon-Salant et al., 2013). Measures of speech perception of accented talkers are highly relevant to listening in a real-world environment, as a growing proportion of the U.S. population is comprised of non-native speakers of English. In addition, listeners must accommodate differences in American dialects when understanding speech, particularly because the familiarity of a dialect can improve speech understanding (Clopper & Bradlow, 2008; Mason, 1946).

The familiarity of a talker or masker can also improve a listener's speech understanding performance in noise. A benefit of 10-15% in speech intelligibility in noise has been shown when listeners attended to a familiar talker in the presence of background noise (Johnsrude et al., 2013; Souza et al., 2013). This benefit is also observed when the listener is trained to detect and identify the target talker's voice (Nygaard & Pisoni, 1998).

Talker Familiarity

Trained Familiarity

Listeners show improved speech understanding when they are trained to listen to a particular talker's voice. Nygaard and Pisoni (1998) employed a training paradigm for young adults in which they learned to identify 10 separate talkers through a regimen over several days. Familiarization of each talker was confirmed through a voice identification test, and intelligibility was then measured over a range of background noise levels. Half of the participants completed the training paradigm, and the remaining participants only completed the final test. A positive correlation was measured between identification of the talker and speech intelligibility performance, suggesting that the trained listeners exhibited higher speech intelligibility scores than listeners who did not undergo the training.

Older adults show a familiarity benefit from listening to a trained talker, and exhibit improved intelligibility when they correctly identify who said the word or phrase than when they misidentify the talker's identity (Rossi-Katz & Arehart, 2009). Yonan and Sommers (2000) measured intelligibility of trained older and younger adult listeners using a training method similar to that described by Nygaard and Pisoni (1998). Results showed that both groups received a significant intelligibility benefit when listening to the trained talker compared to the untrained talker. This benefit was greatest in the more adverse noise conditions (0 and -5 dB SNR).

Talker familiarity benefit can also occur without explicitly training the listener to identify particular talkers. Kreitewolf et al. (2017) trained listeners on a particular talker using a speech-understanding-in-noise test. The listeners became familiar with the particular talker's voice without being explicitly told that they were undergoing a training regimen. The listeners completed four consecutive training sessions before returning for the test phase, which included listening to speech from the trained talker and three additional untrained voices. The results showed that listeners' speech understanding performance improved for the trained talker across the training days.

Incidental Familiarity

Newman and Evers (2007) measured the effect of listening to a familiar voice in the presence of competing speech. Students from a psychology class were recruited for the study, and the familiar voice was their professor. Participants were separated into three groups: students who were directly told that the voice was their professor (explicit familiarity), students who were not told the talker was their professor (implicit familiarity), and students who were in a different class and had no exposure to the familiar voice (novel). Accuracy was measured for young adults on a speech shadowing task, which required the listener to repeat the exact words the target talker spoke while the talker continued to speak. The explicit familiarity group had fewer shadowing errors compared to the implicit familiarity and novel groups. However, no effect of familiarity was noted when the familiar voice was the competing talker. More recent studies have shown that this level of real-world talker familiarity (e.g. listening to a professor) may not elicit a large speech understanding or recall benefit, and that talkers with a more personal familiarity to the listener may result in a greater familiarity benefit (Barker & Elliott, 2019).

Speech understanding for famous voices in popular culture, such as actors or politicians, has been shown to produce a familiarity effect. This type of voice is potentially more ecologically valid as a familiar talker than a voice that was learned over the course of a training study. While the listener may not have a personal relationship with the famous individual, they likely have had repeated exposures to their voice in the real-world context (Maibauer et al., 2014; Schweinberger et al., 1997).

Personal Familiarity

More recent studies have evaluated the effect of talker familiarity with voices that are personally familiar to the listener. Johnsrude et al. (2013) measured the impact of listening to a spouse on a speech segregation task for older adults. Participants were only required to have normal hearing in their better hearing ear, and therefore it is unknown what percentage of listeners had normal hearing in both ears or only one. The spouse for each participant recorded stimuli from the CRM corpus (Bolia et al., 2000). Results for middle-aged and older adults showed a familiarity benefit when the spouse was the target talker, compared to an unfamiliar talker, in the presence of one competing talker. Listeners exhibited a familiarity benefit, although smaller and only at the most adverse target-to-masker ratios (TMR), when the target talker was unfamiliar and the masker was familiar. These findings suggest that listeners were able to use familiarity as a tool for speech segregation, even when the familiar talker was not the intended target. This finding is unique in the literature and has not been confirmed in subsequent studies.

Personal familiarity benefit has also been measured in older adults with hearing loss. Souza et al. (2013) recorded familiar participant pairs (spouses and friends) saying sentences from the IEEE sentence corpus (IEEE, 1969). Speech intelligibility was measured in quiet and at two signal-to-noise ratios (SNR; +2 and +6 dB SNR) in the presence of six-talker babble or speech-shaped noise.

Listeners showed a significant familiarity benefit across all listening conditions, with no difference in benefit across the two background noise conditions.

There are several factors that potentially contribute to talker familiarity benefit on a speech understanding in noise task. As discussed in the previous section, the difference in fundamental frequency is a strong cue for speech segregation (Darwin et al., 2003). In a recent study by Holmes et al. (2018), the F0 and VT length of familiar (spouse or close friend) and unfamiliar voices were manipulated to determine whether these features impacted a listener's ability to recognize a voice as being familiar. The authors also examined whether the acoustic modifications affected the listener's familiarity benefit on speech understanding in noise. Results showed that familiar talker identification remained constant when the F0 was manipulated, but was reduced when the VT length or the combination of F0+VT was manipulated. For speech intelligibility, the familiarity benefit was greatest when the speech signal was not manipulated; however, there was still a benefit across all F0 and VT manipulation conditions. These findings suggest that the speech intelligibility benefit of a personally familiar talker is resilient to manipulations to these acoustic features of the voice, and correct identification of a familiar talker does not necessarily correlate with intelligibility benefit.

In a recent study by Domingo and colleagues (2019), talker familiarity benefit was examined using a spatial release from masking paradigm. The target talker was presented at 0° azimuth and two competing maskers (same voice) were presented symmetrically from 0° to 90° degrees at SNRs ranging from -3 to

+6 dB SNR. Participants were tasked with responding with the words spoken by the target talker. Overall performance was significantly better when the target talker was familiar vs. unfamiliar. Additionally, participants exhibited a talker familiarity benefit that was comparable to a 15° spatial separation between target and maskers of an unfamiliar talker. However, performance on the spatial release task varied more so by SNR than by talker familiarity. The authors concluded that while talker familiarity was equivalent to a 15° spatial release from masking, this benefit was reduced when audibility was more favorable (higher SNR).

The presence of a familiar talker can contribute to improved performance on a short-term memory recall test. A study by Ingvalson and Stoimenoff (2015) assessed working memory ability of young adults on a visual digit recall task while they concurrently completed a speech-in-noise test. At the beginning of each set of trials, eight digits were presented visually on a computer screen and the participants were instructed to hold the numbers in memory. Participants then completed three trials where they heard a sentence presented in 4-talker babble and had to repeat back what they heard. The target talker in these conditions was either their familiar partner or an unfamiliar voice. Following the speech-in-noise trials, the listeners were prompted to recall the eight digits in reverse order. Results showed that younger adults exhibited a familiarity benefit when the cognitive load was increased (i.e., in the conditions where they concurrently completed the speech recognition test and the number recall test). Familiarity benefit was not significant in the conditions where the listeners only completed the speech recognition test. Older adults were not tested on this paradigm, therefore, it is not yet known whether or not talker familiarity has an impact on working memory performance for older people.

Clinical Relevance

The speech understanding benefit elicited by listening to a familiar voice is highly relevant for clinical populations. Clinically driven studies have examined the benefit of incorporating a familiar talker (i.e., spouse) into the rehabilitative process for hearing-impaired patients. Preminger (2003) measured subjective hearing handicap ratings for hearing-impaired participants enrolled in a multi-session aural rehabilitation training program. Participants who included their spouse in the training program had a larger reduction in subjective handicap than those who did not include their spouse, suggesting that communication is improved with a familiar talker. Tye-Murray et al. (2016) assessed whether improvement from a speech recognition training program would be greater with the listener's spouse as the talker, than with the voice of a trained actor. Performance improved following training for both the spouse and the unfamiliar voice. The authors suggested that use of a frequent talker may be a valid way to train and test adults with hearing impairment. It is clear from this research that there is a trend of improvement on clinical subjective and objective measures with the participation of a familiar talker.

Summary and Hypotheses

The long-term goal of this research is to understand the cognitive mechanisms responsible for a FV benefit in real-world environments, and to develop means to exploit the FV benefit to increase saliency of the attended speech for older adults with hearing loss. The objective of this research is to measure the effect of a FV on speech understanding in a cocktail-party scene, while also evaluating performance on a cognitive task. The central hypothesis is that the presence of a FV results in reduced cognitive demand and effort when the FV is the target talker and increased cognitive demand and effort when the FV is the masker. The approach is to measure speech understanding for familiar partners (spouses) in a cocktail-party scene, while simultaneously evaluating cognitive performance through measures of WM. While adults have shown a speech intelligibility benefit to listening to a familiar voice in a controlled laboratory setting, it is unknown whether this effect is as pronounced in a real-world environment. Speech intelligibility of familiar and unfamiliar talkers was examined in a real-world environment (Study 1). Speech segregation of a familiar voice was evaluated when the familiar voice served as either the target or competing talker in a laboratory environment, where the spatial separation of the target talker and maskers varied (Study 2). Lastly, the allocation of cognitive resources when listening to familiar speech was measured with a memory task that was designed to assess WM (Study 3).

Study 1 determined the effect of talker familiarity on speech perception abilities of younger and older adults in a real-world environment, and the impact of a FV on tasks of working memory in a real-world environment. Couples self-administered a speech understanding test at a noisy restaurant. Additionally, as a measure of working memory capacity, participants were asked to recall a target word from previous trials. It was hypothesized that listeners would have better speech intelligibility when listening to a familiar voice compared to an unfamiliar voice. Because talker familiarity benefit has been shown in a laboratory environment (Johnsrude et al., 2013; Souza et al., 2013), it was predicted that a similar effect would be observed in a real-world environment. Based on the ELU model of working memory (Ronnberg et al., 2008), it was hypothesized that listening to a familiar talker in a complex environment would be less cognitively taxing compared to when the signal was spoken by an unfamiliar talker, which would then result in higher performance on an auditory working memory task. Lastly, it was hypothesized that talker familiarity benefit would be greater for older adults than for younger adults, because older adults have reduced cognitive resources compared to younger adults (Pichora-Fuller et al., 1995), and would glean more of a benefit from this cue.

Study 2 examined the interaction between voice familiarity and attention for younger and older adults. Stimuli were recorded from each participant's spouse (familiar voice), and were presented monotically or dichotically under headphones to simulate a cocktail-party environment in a laboratory setting. Speech understanding performance was measured for conditions where the FV was the target talker, and conditions where the FV was a masker. It was hypothesized that a familiar target talker would improve speech segregation, a familiar competing talker would reduce speech segregation, and that these effects would be larger for older compared to younger adults. It was expected that the presence of the familiar voice in a dichotic signal, which was configured by a target and competing talker in one ear and a different competing talker in the opposite ear, would draw the listener's attention away from the target ear when the familiar voice was in the non-target ear (Conway et al., 2001; Moray, 1959). This would result in reduced speech understanding performance. Furthermore, the familiar target talker was expected to be beneficial for older adults who generally have a greater difficulty understanding speech in noise (Dubno et al., 1984) and who also have increased listening effort in noise versus younger adults (Gosselin & Gagne, 2011; Kuchinsky et al., 2013).

Study 3 examined if the presence of a familiar voice resulted in improved performance on a measure of auditory working memory. Similar to Study 2, stimuli were recorded from couples and were presented under headphones. Listeners completed a task of auditory recall that required them to hold a spoken sentence in memory and then recall it after hearing another sentence (i.e., 1-back task). It was hypothesized that listeners would perform better on the 1-back test when the FV was the target talker. This finding would suggest that the FV talker resulted in a reduced demand on cognitive resources, whereas an unfamiliar voice would have the opposite effect. It was also hypothesized that older adults would perform more poorly on the WM measure than younger adults. Poor performance on the 1-back test was expected to correlate with reduced working memory capacity (Gosselin & Gagné, 2011; Schurman et al., 2014).

Study 1: Effect of Talker Familiarity in a Real-World Environment

Introduction

Older adults, regardless of hearing ability, are known to be at a disadvantage when listening in complex environments (Dubno et al., 1984; Gordon-Salant & Fitzgibbons, 1999; Helfer & Freyman, 2008). Many contributing factors, including declines in peripheral, central, and cognitive abilities, are thought to be involved in this age-related speech-in-noise problem (CHABA, 1988). Previous studies have investigated whether the known identity of a talker could provide benefit when listening to speech in noise. These studies, predominantly conducted with older adult listeners, revealed that there was a speech intelligibility benefit to listening to a familiar talker, both for talkers who are familiarized through auditory training (e.g. Nygaard & Pisoni, 1998) and for talkers who are known in real life (e.g. Johnsrude et al., 2013). However, these studies were conducted in a laboratory environment, which may not capture the effects of listening to a familiar voice during a realistic listening situation, such as while at dinner in a noisy restaurant.

Unlike the controlled laboratory setting, talkers and listeners in real-world settings need to actively modulate their attention and vocal effort to communicate with their partner. An individual’s communication strategy may vary depending on whether their communication partner is someone who is familiar to them, such as a spouse, or is unfamiliar to them, such as a new acquaintance. This study evaluated listeners’ performance on a speech intelligibility test while in a real world environment, where the talker was either familiar or unfamiliar to the listener. Social factors that are present during a normal conversation are not easily replicated in a laboratory setting. These factors include a listener's motivation and general social etiquette. Methods to manipulate motivation that shift the reward system, such as allocating increases in payment based on the number of correct responses, are one alternative to this problem. However, in a real-world setting, the listener may be more or less motivated to attend to the target speech due to the information being transmitted or to the relationship to the communication partner.

There has been a recent emphasis in research to examine speech understanding performance in a more ecologically valid environment than the standard laboratory (Smeds et al., 2015). This has been studied most intensively in the hearing-impaired population, primarily with respect to the impact of hearing loss and use of amplification in a real-world noise setting. For hearing-impaired listeners, there are often significant discrepancies between the hearing aid benefit measured in a controlled laboratory environment versus that obtained in real-world listening environments. Best et al. (2015) compared speech intelligibility performance with and without the use of a hearing aid in two acoustic environments: a sound booth with spatially separated target and noise signals, and an acoustically simulated cafeteria with voices and noise of a typical cafeteria. Performance differences across environments were apparent, with a greater benefit of amplification noted in the simulated cafeteria environment.

The complexity of the communication environment can result in reduced speech understanding ability. Both visual and auditory distractions constantly change in a real-world environment, whereas laboratory based auditory-visual speech perception may not reflect those dynamic shifts in the environment. These competing signals can potentially result in a drain on attention for older adults, which can negatively impact speech understanding (Cohen & Gordon-Salant, 2017). As a result, findings in the laboratory may under-estimate the speech understanding difficulties of individuals in everyday, complex environments.

Working memory is a cognitive function that is critical for speech understanding in complex environments for both younger and older adults. Working memory is a limited capacity system that allows for the storage and manipulation of information in short term memory (Baddeley, 2000; Baddeley & Hitch, 1974). The ELU model suggests that when bottom-up constraints such as hearing loss, or the top-down process of reduced working memory capacity are present, speech recognition performance may be reduced (Rudner & Ronnberg, 2008). Working memory capacity is considered to be the amount of information a listener can manipulate and then recall from short term memory. The LSPAN was developed as a means to assess working memory capacity in the auditory domain (Daneman & Carpenter, 1980; Pichora-Fuller et al., 1995). Listeners are instructed to make a semantic judgment about each sentence presented, and then recall the final word of each sentence they hear in a particular set of sentences. The maximum number of sentences that the listener can recall is their

LSPAN score. Schurman et al. (2014) measured the effect of sentence context and type of competing noise on younger and older normal-hearing adults' ability to recall sentences in a 1-back paradigm. They found that older adults had poorer sentence recall on the memory task. The results on this test were highly correlated with the participant's LSPAN scores, and suggest that the difference in performance across age groups may have been due to the older adults' reduced working memory capacity. These studies suggest that WM capacity is critical for speech understanding in noise, and that older adults who have reduced WM capacity are more adversely affected than younger adults in noise because of this limitation.

Speech understanding in a real-world environment can involve attending to a voice that is familiar to the listener. Laboratory-based studies have shown that a familiar voice can improve speech understanding in noise, and may also contribute to improved performance on a concurrent recall task. Johnsrude et al. (2013) demonstrated that middle-aged and older adults exhibit a talker familiarity benefit when attending to their spouse, such that speech understanding in the presence of a single competing talker was better when the target talker was familiar versus unfamiliar.

Beyond providing a benefit for speech understanding in noise, a familiar talker can also contribute to improved performance on a delayed recall task. When listeners were asked to recall whether a word had been presented more than once within a continuous stream, identification performance was significantly better when the same talker spoke the repeated word, compared to when a different voice repeated the word (Palmeri et al., 1993). In a study by Ingvalson and Stoimenoff (2015), speech understanding performance was measured while listeners simultaneously completed a digit recall task. The results showed that younger adults only exhibited a familiarity benefit for the conditions that required the listener to concurrently complete the speech recognition test and the number recall test. These findings suggest that listening to a familiar voice may reduce cognitive demand (i.e., improve the availability of cognitive resources).

Research Questions and Hypotheses

The first aim of this study was to examine the effect of talker familiarity on a speech understanding test in a real-world environment. Talker familiarity benefit has been demonstrated in a laboratory environment (Johnsrude et al., 2013; Souza et al., 2013), and it was predicted that a similar effect would be shown in a real-world environment. It was hypothesized that talker familiarity benefit will be greater for older adults than for younger adults, because older adults have less access to cognitive resources than younger adults (Pichora-Fuller et al., 1995), and rely more heavily on this cue for improved speech understanding in the real-world environment.

The second aim of this study was to determine the impact of a familiar voice as the target talker on a measure of working memory in a real-world environment. Based on the ELU model of working memory (Ronnberg et al., 2008), the hypothesis was that listening to a familiar talker in a complex environment would be less cognitively taxing than when listening to an unfamiliar talker, resulting in higher performance on an auditory working memory task.

Method

Participants

Pairs of normal-hearing adults aged 18-80 years old who were highly familiar with each other were recruited. Eight spouse pairs, for a total of 16 participants (mean age = 47.9 yrs ± 19.9), were recruited for this study. Familiar partners were defined as spouses who have cohabitated for at least 1 year. Normal hearing sensitivity was defined as thresholds ≤ 25 dB HL from 250-4000 Hz (ANSI, 2018). Participants completed the Montreal Cognitive Assessment (Nasreddine et al., 2005), with a passing score of ≥ 26 . All participants had completed a high school level of education and were native speakers of English.

A priori power analyses were conducted to determine a sufficient sample size for a repeated measures (RM) analysis of variance (ANOVA) with 10 measurements with a small effect size (0.25), power of 0.8, and alpha level of 0.05. Based on these assumptions, as well as comparisons of effect size from previous relevant research (Johnsrude et al., 2013), the required sample size is 18 adults. However, additional planned statistical modeling, including the examination of the between-subject factor of age, was expected to benefit from a larger sample. Due to the COVID-19 pandemic (Wuhan Flu, 2019) data collection was halted after data were collected on 16 participants. The between-subjects grouping of age was removed, and age in years was used in the statistical models.

Stimuli

A hybrid corpus composed of components of the CRM and the Modified Rhyme Test (MRT; House et al., 1963) were used as the stimuli for this study. The original CRM sentences have a structure of “Ready <call sign> go to <color> <number> now,” where the call sign could be one of eight choices (arrow, baron, charlie, eagle, hopper, laker, ringo, tiger), four colors (blue, green, red, and white), and numbers between one and eight (Bolia et al., 2000; Brungart, 2001a). This test is a highly sensitive measure of speech intelligibility in the presence of competing speech (Brungart, 2001a), but is not as effective in the presence of non-speech maskers, such as speech-shaped noise (Brungart, 2001b). The MRT, however, is a sensitive measure of speech intelligibility in speech-shaped noise (SSN) (House et al., 1963). The MRT is composed of 300 unique consonant-vowel-consonant (CVC) words that are embedded in a carrier phrase. The target word is presented in a closed-set task where the listener selects the word spoken from six foils, which either vary based on the initial or final consonant (e.g. “You will mark <CVC> please.”). As an example, the target word “tap” could have the following foils that vary in their final consonants: tab, tap, tam, tan, tack, tang.

To implement a self-administered speech understanding paradigm in a real-world environment, the use of a closed-set sentence corpus was necessary. In this paradigm, the WM task uses the same trials as the speech intelligibility task, but in an n-back format, which only utilized the CRM call signs as the target. The CRM call signs were chosen for the WM task because they are known to be highly intelligible words, and thus recognition of these words was not expected to impact performance on the WM measure (Brungart, 2001a). The hybrid corpus therefore included both the CRM call sign and a mixture of the MRT word stimuli, with the new structure of “Ready <call sign> you will mark the <MRT word #1> and the <MRT word #2> again.”

Procedures

Groups of four participants, consisting of two pairs of familiar couples who are unfamiliar with one another, were tested in a casual-dining restaurant that was known to have a relatively high level of background noise. Testing was conducted during lunch (11am – 1pm) or dinner hours (5pm – 8pm), when background noise was expected to be at peak levels. The members of each group were seated around a rectangular table, with two participants seated on each of the long sides of the table. Seating position was randomized such that familiar pairs were not always seated next to or across from one another. The experimental protocol was implemented using a MATLAB script (Mathworks, 2017b) on a laptop. Each participant was given a touch-sensitive Nexus Tab E tablet that was paired via Bluetooth to a Dell Latitude laptop computer (Windows 10; 64 bit). For a given trial, one participant was randomly selected as the talker. As illustrated in Figure 2.1, a prompt appeared on the talker’s tablet and instructed that participant to read aloud the specified sentence. After the stimulus was read, the talker touched their tablet and a response grid (Figure 2.1, middle panel) appeared on the listeners' tablets. Each listener then identified the words they heard from a closed set of word choices for the call sign and the two MRT words. A new grid appeared for each trial that included the same call signs (left column) and new foils corresponding to the new target MRT words.

Figure 2.1. The left panel is a screenshot of the prompt that the target talker received. After the talker reads aloud the MRT phrase, they tap their screen and trigger the response grid (middle panel) to appear on the listeners' tablets. On the n-back trials, all participants respond on a grid of the possible call sign names for each of the four previous trials (right panel).

Listeners also completed an n-back style working memory task that appeared after four successive trials of the hybrid CRM-MRT speech perception task. All participants were prompted to recall the call signs that were spoken for the previous four trials and to indicate their responses on the tablet response grid (Figure 2.1, right panel). The response choices for the first trial in the set appeared in the far right column, and then progressed to the left with the response choices for Trial 4 in the left-most column. The same five call sign foils appeared in each column. The participants were prompted to respond with a selection for each column. The call sign for each trial was coded for each participant as being spoken by their familiar talker, by an unfamiliar talker, or as being spoken by themselves. Trials where the call sign was spoken by the listener were not included in the final analysis.

Participants completed 200 trials of the speech understanding task and 44 trials of the n-back recall task, with n-back trials occurring every 5th trial. Trials were blocked into sets of 55 trials. Participants were given a short break between each block of trials. Testing took approximately two hours over the course of a single session. All participants were paid for their participation.

In each trial of the study, the tablet microphones from the target talker and listener tablets measured the peak dB SPL. There were two main periods of time during a given trial, the "talker interval" and the "listener interval." The talker interval was the time period between the appearance of the sentence prompt on the target talker's tablet and the talker's screen tap signifying they finished saying the sentence out loud. Thus, this interval reflected the time period when the target talker read the sentence aloud. The listener interval was the time period immediately following the target interval until all listeners selected the target words from the response grid.

Data Analysis

Generalized linear mixed effects (GLME) models were conducted to evaluate fixed and random effects on speech intelligibility score (Aim 1) and n-back score (Aim 2). The GLME models were evaluated using the buildmer package v1.6 in R (Voeten, 2020), which implemented a forward feeding approach to determine the maximal random effect and fixed effects structures, and then used a backwards elimination approach to systematically reduce the fixed effects to obtain the best fitting model for the data. First, the order of the fixed effects was determined, such that the effects that account for the greatest amount of variance were maintained in the model to achieve model convergence. During the backwards elimination procedure, the deviance values for each model were compared using a Wald test, which has a χ^2 distribution (α = 0.05). The number of iterations was set to 2,000,000 to allow for model convergence. The best fitting model for each dependent variable was depicted in a table in the Results section. Each table consists of a list of main effects and interactions that were deemed significant or contributed to a higher order interaction effect in the model.

Results

Sound Level in the Real-world Environment

The left panel of Figure 2.2 shows the peak dB SPL during the talker and listener intervals, averaged across all listeners for a given trial. The top line represents the sound level during the talker interval and the bottom line depicts the sound level during the listener interval. From these measures, it can be inferred that the SNR during an individual trial was monotonically related to the difference between the talker interval, which contains the talker speech and the competing background noise, and the listener interval, which only contains the background noise. The second panel is a histogram of the computed SNRs recorded for each trial (n = 244) averaged across all listeners. The distribution shows that participants were primarily listening at a +5 dB SNR, which is consistent with the findings reported by Smeds et al. (2015).

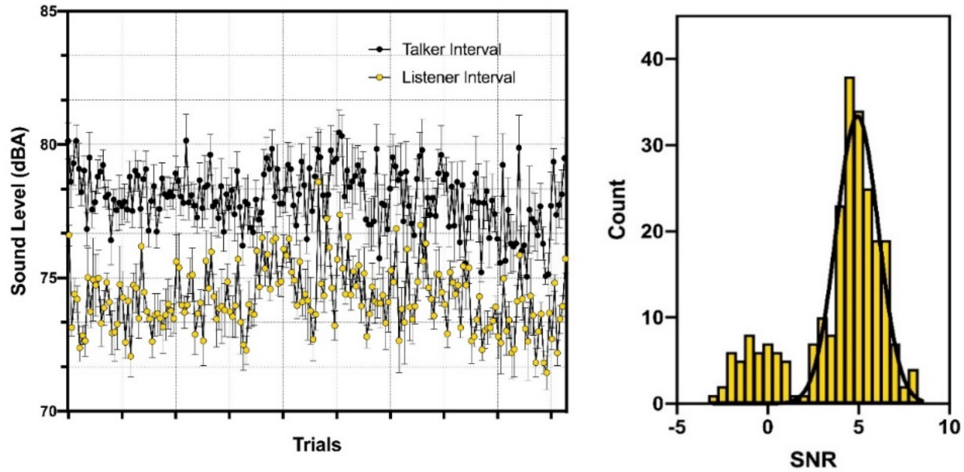


Figure 2.2. The left panel is a plot of the mean dB SPL measured by each tablet for each trial. The black line represents the sound level during the talker interval and the yellow line is the average sound level during the listening interval. The right panel is a histogram of the difference in dBA between the talker and listener intervals.

Familiarity Benefit on the Cafeteria Test

A GLME model was conducted to determine the contribution of talker familiarity, participant age, working memory capacity (LSPAN), and overall noise level to speech understanding scores. The values for the three continuous variables (age, LSPAN, and noise level) were grand-mean centered. The results from the final GLME model are shown in Table 2.1. All main effects were significant, and will be discussed with regard to their higher order interactions.

Table 2.1

GLME Model for the Speech Perception Task

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	2.060	0.085	24.139	<0.001	***
Target Talker: Unfamiliar (ref)					
Familiar	0.565	0.137	4.129	<0.001	***
Age (standardized)	-0.553	0.081	-6.871	<0.001	***
LSPAN (standardized)	0.206	0.090	2.293	0.022	*
Mean Overall Level dB A (standardized)	-0.129	0.065	-1.992	0.046	*
Interactions					
Familiar Target x Age	0.043	0.160	0.269	0.788	
Familiar Target x LSPAN	-0.008	0.173	-0.045	0.964	
Age x LSPAN	-0.216	0.073	-2.975	0.003	**
Age x Mean Overall Level	-0.191	0.071	-2.677	0.007	**
LSPAN x Mean Overall Level	-0.119	0.078	-1.527	0.127	
Familiar Target x Age x LSPAN	-0.139	0.060	-2.310	0.021	*
Mean Overall Level x Age x LSPAN	0.301	0.132	2.277	0.023	*

Random Effects	Variance	SD
Sentence Item Intercept	0.312	0.559

****p*<0.001. ***p*<0.01. **p*<0.05.

The primary question of interest in this study was whether or not there was an overall effect of talker familiarity on speech understanding in a real-world environment. The proportion of correct scores for when the target talker was familiar vs. unfamiliar is plotted in Figure 2.3. The individual points represent the speech understanding score for each participant, averaged across all test trials.

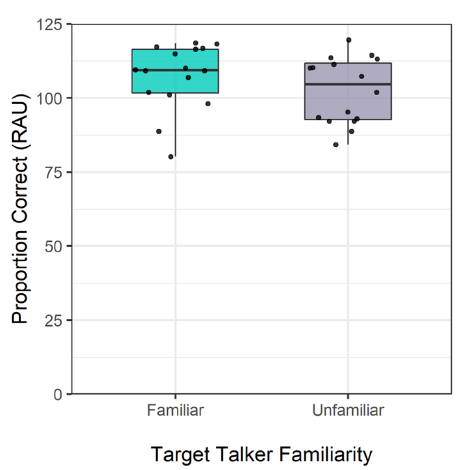


Figure 2.3. Boxplot of proportion correct speech understanding scores, plotted as a function of Target Talker conditions.

While the main effect of Target Talker was significant (see Table 2.1), there was a significant three-way interaction of Target Talker x Age x LSPAN ($p = 0.021$). The interaction is plotted in Figure 2.4, which shows the regression lines for the familiar and unfamiliar Target Talker conditions. To depict the influence of working memory, performance was plotted separately for listeners with relatively high LSPAN scores (left panel) vs. listeners with relatively low LSPAN scores (right panel). High LSPAN scores were values above the median (> 4 points), and low LSPAN scores were below the median (≤ 4) score. It appears that with increasing age, individuals with a High LSPAN score (Figure 2.4, left panel) exhibited better performance on the Familiar condition than the Unfamiliar condition ($p<0.01$). However, with increasing age, individuals with a Low LSPAN score (Figure 2.4, right panel) did not exhibit a difference in performance across the Familiar and Unfamiliar conditions ($p>0.05$). The younger participants with High LSPAN scores exhibited near-ceiling performance for both conditions, whereas younger participants with Low LSPAN scores showed a benefit of talker familiarity.

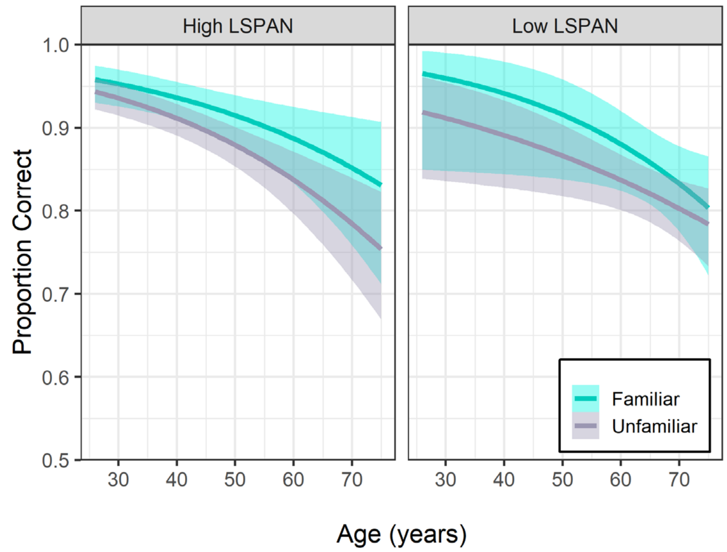


Figure 2.4. Proportion correct for speech understanding scores of High LSPAN and Low LSPAN (High: > 4; Low: ≤ 4), plotted as a function of Age for Familiar and Unfamiliar Target Talker conditions. Shaded areas around functions represent 95% confidence intervals.

In addition to talker familiarity, performance on the speech perception task was impacted by the mean overall noise level (dBA), which was measured for every trial from each participant's tablet. There was a significant three-way interaction of Overall (Noise) Level x Age x LSPAN (β = 0.301, SE = 0.132, Z = 2.277, p = 0.023). In Figure 2.5, separate plots are shown for younger vs. older listeners, with the older listeners representing those whose age was above the median value (Older >31 years). On each plot, the Overall Level is represented as high noise, with values above the median (> 73.95 dB SPL) and low noise, with values below the median (< 73.95 dB SPL). For each age group (younger and older), regression lines were plotted for each Overall Level x Age combination. When the noise level was high, the older adults showed fairly stable performance regardless of LSPAN score. However, when the noise level was lower (Low Noise), performance improved with increasing LSPAN score (*p*<0.01). Younger adults showed similar performance regardless of noise level (high vs. low). Those with low LSPAN scores showed a wide range in performance, as noted by the 95% confidence intervals for the younger adults with LSPAN scores below 3. However, younger adults with higher LSPAN scores consistently showed a high level of performance. These performance trends did not vary with the presence of a familiar vs. unfamiliar target talker, as talker familiarity was not involved in this interaction.

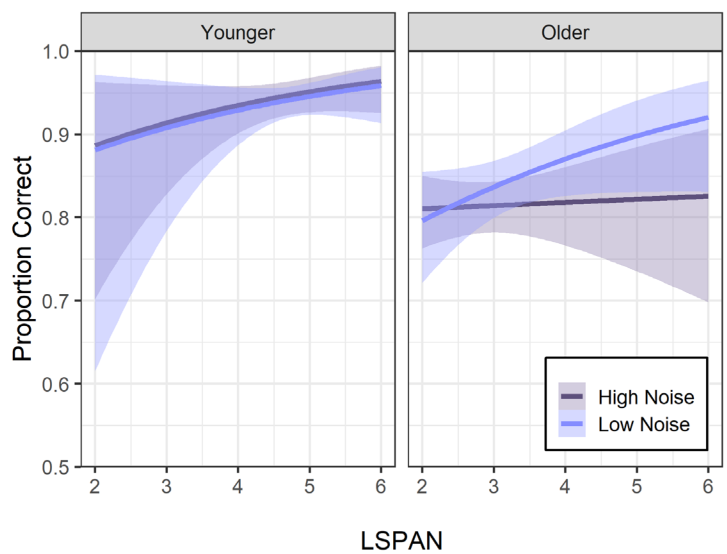


Figure 2.5. Proportion correct of speech understanding scores of Younger and Older listeners (Younger: ≤ 31 yrs; Older: > 31 yrs), plotted as a function of LSPAN for High Noise (>73.95 dBA) and Low Noise (≤ 73.95 dBA) trials. Shaded areas around functions represent 95% confidence intervals.

Performance on the N-Back Test

The second aim of this study was to evaluate whether a familiar voice improves word recall on a measure designed to assess working memory capacity. The n-back scores for when the target talker was familiar vs. unfamiliar are plotted in Figure 2.6 as a function of n-back word position. The individual points represent the n-back score for each participant, averaged across all test trials for a particular condition.

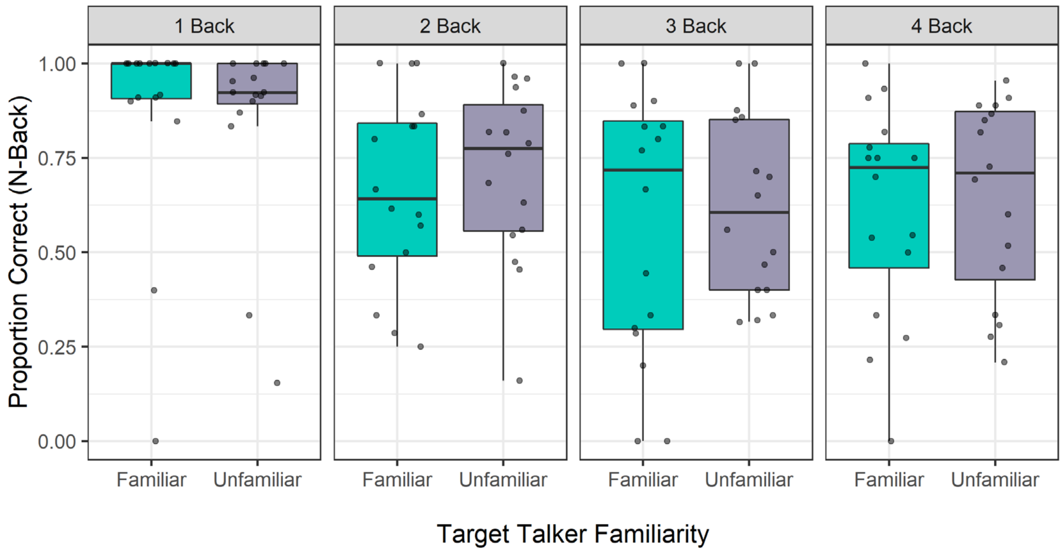


Figure 2.6. Boxplot of proportion correct n-back scores, plotted as a function of Target Talker conditions for each word position (1 = word spoken one trial prior, 4 = word spoken four trials prior.

A GLME model was conducted to determine the contribution of talker familiarity, word position, age, and working memory capacity (LSPAN) on recall scores in the n-back test. The values for the two continuous variables (age, and LSPAN) were grand-mean centered. The results from the final GLME model are shown in Table 2.2. For the n-back test, listeners were tasked with recalling the call sign (e.g. “Baron”) for the previous four trials.

Table 2.2

GLME Model for the N-Back Test

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	2.881	0.433	6.656	<0.001	***
Target Talker: Unfamiliar (ref)	-0.038	0.130	-0.295	0.768	
Familiar					
N-Back Position: 1-Back (ref)					
2-Back	-1.963	0.319	-6.154	<0.001	***
3-Back	-2.211	0.317	-6.970	<0.001	***
4-Back	-2.111	0.315	-6.695	<0.001	***
Age (standardized)	0.845	0.497	1.701	0.089	
LSPAN (standardized)	1.421	0.506	2.805	0.005	**
Interactions					
2-Back x Age	-0.427	0.380	-1.124	0.261	
3-Back x Age	-0.484	0.378	-1.281	0.200	
4-Back x Age	-0.651	0.374	-1.738	0.082	
2-Back x LSPAN	-0.063	0.405	-0.156	0.876	
3-Back x LSPAN	-0.022	0.398	-0.055	0.956	
4-Back x LSPAN	-0.142	0.395	-0.360	0.719	
Age x LSPAN	0.536	0.396	1.353	0.176	
2-Back x Age x LSPAN	-0.999	0.302	-3.312	<0.001	***
3-Back x Age x LSPAN	-0.733	0.299	-2.453	0.014	*
4-Back x Age x LSPAN	-0.668	0.298	-2.242	0.025	*

Random Effects	Variance	SD
Subject Intercept	0.802	0.896
Sentence Item Intercept	0.312	0.559

***p<0.001. **p<0.01. *p<0.05.

The main effect of interest was whether talker familiarity influenced performance on high memory load recall task. The factor of talker familiarity was not a significant predictor of performance on the n-back test (p>0.05). There were, however, significant main effects of n-back word position and LSPAN, as well as a significant three-way interaction of n-back Position x Age x LSPAN. Figure 2.7 depicts this three-way interaction by showing results separated by the position (1-back, 2-back, 3-back, 4-back). The continuous variable of LSPAN score was divided into values below the median (Low LSPAN: ≤ 4) and above the median (High LSPAN: > 4) categories, and regression lines were plotted for the Age x LSPAN interaction for each n-back word position. The effect of age on n-back performance for a given LSPAN category and n-back word position was only significant in the 1-back condition. The leftmost panel of Figure 2.7 shows the Age x LSPAN interaction for the 1-back condition. Across the span of ages, individuals with higher LSPAN scores performed near ceiling in the 1-back condition. However, performance declined with increasing age for individuals with lower LSPAN scores. The 2-, 3-, and 4- back conditions were not significantly different from each other ($p>0.05$), and recall performance on each condition was significantly poorer than on the 1-back condition ($p<0.05$). Additionally, it should be noted that there were a limited number of older adults with higher LSPAN scores (n = 3 for adults over 60 years). Thus, the 5-95% confidence intervals for the high LSPAN category were substantially wider at ages greater than 40 years than for younger adults (< 40 years).

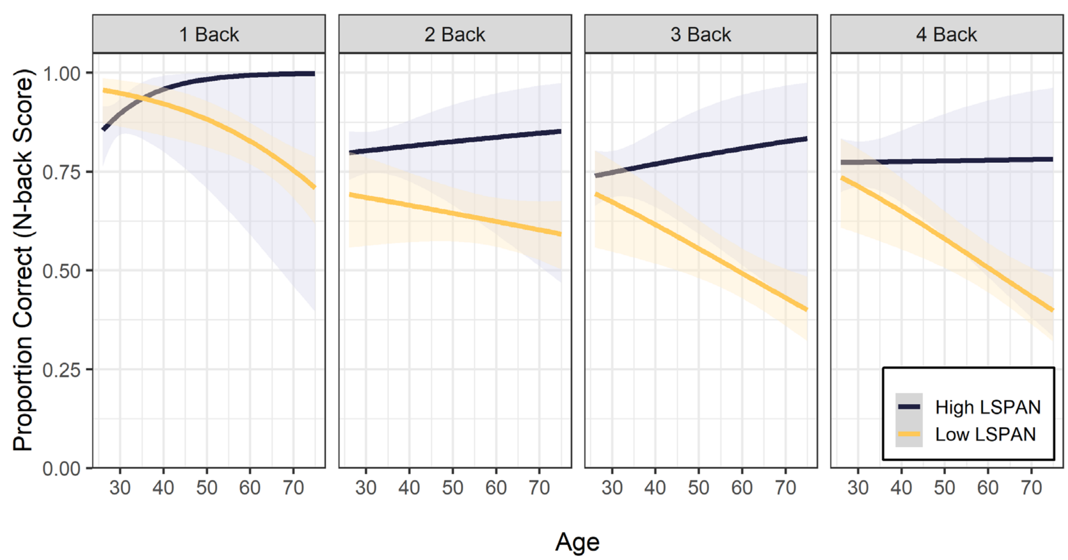


Figure 2.7. Probability of correct n-back score for each word position (1 = word spoken one trial prior, 4 = word spoken four trials prior), plotted as a function of Age and LSPAN (Higher: > 4; Lower: ≤ 4). Shaded areas around functions represent 95% confidence intervals.

Discussion

Talker Familiarity in a Real-world Environment

The first aim of this study was to examine the effect of talker familiarity on a speech understanding task in a real-world environment. Talker familiarity benefit has been demonstrated in a laboratory environment (Johnsrude et al., 2013; Souza et al., 2013), and it was predicted that a similar effect would be shown in a real-world environment. The results from the speech perception task (Table 2.1) revealed a significant effect of talker familiarity, confirming the hypothesis that listening to a familiar talker in a real-world environment can result in better speech understanding than listening to an unfamiliar voice. However, this familiarity benefit was involved in a 3-way interaction, indicating that this benefit was modulated by listener age and LSPAN score. Older adults with higher LSPAN scores exhibited a talker familiarity benefit, whereas older adults with lower LSPAN scores did not have a familiarity benefit. The younger participants with high LSPAN scores exhibited near-ceiling performance for both conditions, whereas younger participants with low LSPAN scores showed a benefit of talker familiarity. Previous studies have shown that both older and younger normal-hearing adults exhibit a talker familiarity benefit when listening in noise (Domingo, Holmes, & Johnsrude, 2019; Johnsrude et al., 2013). Johnsrude et al. (2013) found that speech understanding performance when listening to a familiar voice in the presence of a single competing unfamiliar voice did not vary with increasing age. Domingo, Holmes, and Johnsrude (2019), however, showed a significant decline in talker familiarity benefit with increasing age. In the current study, the effect of talker familiarity on speech understanding in a real-world environment was mediated by the listener's age and working memory capacity. The older adults with lower LSPAN scores did not show a talker familiarity benefit vs. older adults with higher LSPAN scores. When working memory capacity was reduced, older adults had increased difficulty understanding speech in a real-world environment and were less able to access familiarity cues that would potentially provide an improvement to their speech understanding ability. This suggests that talker familiarity benefit is somewhat dependent on working memory capacity.

It was also hypothesized that talker familiarity benefit would be greater for older adults than for younger adults, because older adults have reduced cognitive resources compared to younger adults (Pichora-Fuller et al., 1995), and therefore would rely more on the familiarity cue to derive some benefit in speech understanding. On the speech understanding task, there was a significant interaction between talker familiarity, working memory capacity (LSPAN score), and age, suggesting that the benefit of listening to a familiar voice varied by listener age and working memory capacity. With increasing age, individuals with higher working memory capacity exhibited a benefit from listening to a familiar talker vs. unfamiliar talker. This effect did not hold true for individuals with lower LSPAN scores. The younger adults with lower working memory capacity had better speech understanding performance with the familiar talker, whereas the older adults with lower working memory capacity did not benefit from listening to a familiar talker than an unfamiliar talker. This is consistent with other studies that have shown that older normal-hearing adults' performance on speech understanding in noise is correlated with their working memory capacity (Fullgrabe et al., 2015; Gordon-Salant & Cole, 2016; Schurman et al., 2014)

There was a high level of variability in noise levels during the experiment, which impacted listener performance across trials. As shown in Figure 2.2, the distribution of noise levels during the "talker interval" and "listener interval" approximated the signal-to-noise ratio across a trial. On average, the measured SNR was approximately +5 dB. This level is consistent with previous studies that have shown that average SNR in a moderately loud environment was approximately near this level (Smeds et al., 2015). However, there was significant variability in mean overall noise level across each trial, which ranged from 66 – 80 dBA. The trials were categorized as Low Noise if they had mean overall levels of less than 74 dBA (median overall noise level), and were categorized as High Noise if the mean overall levels were above that value.

In addition to talker familiarity benefit, working memory capacity also modulated performance on the speech understanding task as a function of background noise level. Performance across both noise conditions was shown to be better with higher LSPAN scores for the younger adults. The younger adults were more immune to variation in noise level on the speech understanding task, and performed similarly when the overall noise levels were above (High Noise) or below (Low Noise) the median noise level. However, the older adults showed a significant improvement in speech understanding when the mean overall noise level was low vs. high, and this improvement varied as a function of LSPAN score (Figure 2.5). When LSPAN scores were lower (i.e., low WM) for the older listeners, there was no significant difference in speech understanding scores across the two noise categories. For older adults with higher LSPAN scores, however, speech understanding scores improved in the low noise conditions. In the higher noise levels, speech understanding performance did not vary by LSPAN score. This effect could be due to the overall reduced audibility that all listeners experienced in the trials with higher noise levels.

These findings suggest that while both younger and older listeners exhibited improved speech understanding with a familiar target talker, benefit was modulated by working memory capacity as measured on the LSPAN test. The older adults with better working memory capacity exhibited the familiarity benefit, whereas older adults with lower working memory capacity did not show this same effect. These findings are consistent with those of Gordon-Salant and Cole (2016), who showed that for younger and older adults, individuals with higher working memory capacity had better speech recognition performance in noise than individuals with lower working memory capacity.

Talker Familiarity Benefit on a Test of Working Memory

The second aim of this study was to determine the impact of a familiar voice on a measure of working memory capacity in a real-world environment. Based on the ELU model of working memory (Ronnberg et al., 2008), it was hypothesized that listening to a familiar talker in a complex environment would be less cognitively taxing than when listening to an unfamiliar talker, resulting in higher performance on an auditory working memory task. The results from the n-back task (Table 2.2) revealed that there was no effect of talker familiarity on the n-back test. Thus, while performance on the speech understanding task benefited from the presence of a familiar talker, this benefit did not result in a reduction in cognitive resource consumption as measured by the n-back test. One reason for the absence of a talker familiarity benefit is that while speech understanding in noise has been shown to improve when the target talker is familiar (Domingo, Holmes, & Johnsrude, 2019; Holmes et al., 2018), the benefit to speech understanding possibly does not extend to the use of cognitive resources when listening in noise. However, the lack of a familiarity benefit on a task that increases memory load may depend on the nature of the increased cognitive demand. In this particular study, participants were tasked with recalling a single word from four prior sentences. It is possible that a more complex recall task that requires the listener to recall an entire sentence would elicit a different result. Other types of speech understanding tasks in which memory load is increased need to be conducted to verify this observation. Study 3 examines a similar issue using a different method of increasing cognitive demand.

Consistent with the speech understanding results, performance on the n-back task was modulated by working memory capacity as measured by the LSPAN test. Across all word positions, performance was better for individuals with higher LSPAN scores (Figure 2.7), irrespective of talker familiarity. This confirmed that the n-back test in this study was sensitive to differences in working memory capacity, such that overall performance was correlated with LSPAN scores. Schurman et al. (2014) used a sentence recall 1-back procedure to assess performance of younger and older adults while listening to speech in different types of noise configurations, including spatialized noise and competing speech. They found that older adults had poorer sentence recall in a 1-back paradigm, and that performance on the sentence recall test in noise correlated with older adults' working memory capacity.

The effect of age and LSPAN score on n-back performance was only significant in the 1-back condition. This effect suggests that in the 1-back condition only, increased age was associated with poorer percent correct with increasing LSPAN score. However, this effect is likely the result of a limited range in High LSPAN scores, in which adults past the age of 50 years did not achieve high LSPAN scores (> 4). It is clear from the data plotted in Figure 2.7 that the model estimates for High LPSAN score for older adults reflect the best fit predictions, but the range of the actual raw data is limited. This is also reflected by relatively large confidence intervals in the predictions for older adults. Thus, the three-way interaction of word position, age, and LSPAN was likely driven by the limited range in LSPAN scores (i.e., mostly lower scores), and may not reflect a true interaction between age and LPSAN score for this task. Taken together, the familiarity of the target talker did not modulate performance on the n-back test, which was used as a method to increase cognitive load on a working memory task. This may suggest that while talker familiarity improves speech understanding, this improvement does not also result in increased availability of cognitive resources.

Conclusions

This study examined the effect of talker familiarity on a speech understanding test in a real-world environment. It was hypothesized that talker familiarity benefit would be greater for older adults than for younger adults, because older adults have reduced cognitive resources compared to younger adults (Pichora-Fuller et al., 1995), and would have the potential to glean more benefit from this cue because of reduced speech intelligibility and reduced cognitive resources. The findings from this study indicate that while both younger and older listeners exhibited a talker familiarity benefit in a real-world environment, this benefit varies considerably more with working memory capacity than age, although the two are highly correlated. The older adults with lower working memory capacity did not show a significant talker familiarity benefit on speech understanding, whereas those with a higher working memory capacity exhibited this benefit. Conversely, younger adults with lower working memory capacity showed a benefit of talker familiarity, whereas performance was near-ceiling in the familiar and unfamiliar talker conditions for younger adults with higher working memory capacity.

This study also examined the effect of a familiar voice on tasks of working memory in a real-world environment. It was hypothesized that listening to a familiar talker in a complex environment would be less cognitively taxing than listening to an unfamiliar talker, and would thus result in higher performance on an auditory working memory task when the signal was spoken by a familiar vs. unfamiliar target talker. The familiarity of the target talker did not affect performance on the n-back test, which was used as a measure of auditory working memory capacity. This finding may suggest that while talker familiarity provides a speech understanding benefit in a real-world environment, this improvement does not translate to a change in available cognitive resources in that complex environment.

Study 2: Effect of Target Talker and Masker Familiarity on a Dichotic Speech Perception Task

Introduction

Conversing with a partner in a noisy environment, such as over lunch at a busy restaurant, requires the listener to use strategies to pick out the important information (i.e. speech) from the cacophony of background noise and distractions. This type of scenario, where the listener must extract the target speech from the background noise, is referred to as the “cocktail party problem” (Cherry, 1953). The process of segregating and grouping sounds is critical for speech understanding in noise (Bronkhorst, 2000) and is difficult for many listeners. The problem is complex, as speech understanding in noise is dependent on the environment in which the individual is listening (e.g. the type of competing signals, real-world vs. laboratory setting), individual subject factors (e.g. age, cognitive function), and characteristics of the talker(s) (e.g. gender, native language, familiarity of the voice).

Real-world environments are dynamic – the level difference between the target talker and the background noise, the characteristics of the talkers and noise, and their location can vary from moment to moment. Average signal-to-noise ratios (SNRs) in real-world environments, such as a school or restaurant, are about +5 dB SNR (Smeds et al., 2015). Beyond the SNR, listening in the presence of competing speech vs. steady-state noise can improve speech understanding by allowing the listener to glean portions of the target speech through dips in the speech masker (Bronkhorst, 2000).

As described in the classic cocktail party problem, listeners are able to improve speech understanding by ignoring irrelevant information in one ear and focusing their attention on the intended message in the target ear (Cherry, 1953). However, certain signals are known to divert attention towards the ear with the "irrelevant information." In a study by Moray (1959), listeners' attention was diverted from the primary listening task when their names were spoken in the unattended or distractor channel in a dichotic shadowing task. These results suggest that variations in context can result in a masker penalty. Brungart and Simpson (2002) developed a listening paradigm to attempt to replicate a cocktail party listening environment in a laboratory setting. The target ear contained the voice of the target talker and the voice of a competing talker. A separate competing talker voice was presented to the non-target talker ear, referred to as the unattended ear. They found that the presence of a speech masker in the unattended ear resulted in poorer speech segregation than when noise was presented to the unattended ear, as well as when no sound was presented in that ear (i.e. monotic presentation). These results are in agreement with Moray's findings, indicating that a relevant speech signal in the unattended ear interferes with attention to the target talker in the test ear. This effect was greater than the combination of energetic and informational masking that was present in the target ear. Iyer et al. (2010) used this same method, but varied the semantic content of the message presented by the maskers in both ears. They found a “penalty” when at least one masker was contextually similar to the target talker, especially in more adverse SNRs. These findings suggest that listeners perform some categorization of speech in the unattended (masked) ear, and that this processing can result in increased interference when the masker signal is highly relevant to the listener.

Voice characteristics of the target and competing talkers, such as the talker’s sex, can influence listeners’ attention to a particular speech stream. The similarity of vocal features across target and competing talkers can result in increased challenges in segregating the voices. Speech understanding in the presence of competing speech is better when there is a difference in sex (Brungart, 2001b; Brungart et al., 2001) or when there are differences in fundamental frequency and vocal tract length (Darwin et al., 2003) between the target and competing talkers. Both younger and older adults are able to take advantage of these segregation cues (Lee & Humes, 2012). However, older adults have greater difficulty listening in competing speech than younger adults.

Speech understanding in a complex environment can be especially difficult for older adults with normal hearing, who are less able to extract the target signal and inhibit the competing noise than younger adults with normal hearing (Dubno et al., 1984; Wingfield et al., 2006). These differences can be related not only to possible differences in signal audibility, but also to age-related decline in cognitive ability. Senescent changes in working memory capacity, inhibition, and speed of processing have been well described in the literature

(CHABA, 1988; Hasher & Zacks, 1988; Lipnicki et al., 2017; Park et al., 2002; Pichora-Fuller et al., 1995; Presacco et al., 2016a; Salthouse, 1996).

Working memory is the ability to analyze and temporarily store information during a processing task (Baddeley & Hitch, 1974). The Ease of Language Understanding model (Ronnberg, 2003; Ronnberg et al., 2019; Ronnberg et al., 2008) suggests that the WM system is engaged when the auditory signal is degraded, such as listening in background noise. Previous studies with normal-hearing adults have shown that WM capacity is critical for speech understanding in noise, and that age-related declines in WM capacity contribute to older adults' difficulty listening in background noise (Gordon-Salant & Cole, 2016; Pichora-Fuller et al., 1995).

One cue for speech segregation that improves speech understanding in noise for older adults is the familiarity of the talker's voice. Older adults with hearing loss exhibit a 10-15% improvement in speech understanding in the presence of competing noise and multi-talker babble when the target talker is familiar (Souza et al., 2013). Johnsrude et al. (2013) measured the impact of talker familiarity on a speech segregation task for older adults. Each participant recorded stimuli from the Coordinate Response Measure (CRM), which uses closed-set stimuli that follow the format "Ready <call sign> go to <color> < number> now" (Bolia et al., 2000). This measure has been used to evaluate the effects of informational masking and speech segregation (Brungart, 2001b; Brungart et al., 2001). Stimuli for the Johnsrude et al. (2013) were mixed to generate a target with one competing talker. Results from this study showed a talker familiarity benefit, such that speech understanding was best when the target talker was a familiar voice. This talker familiarity benefit can persist even when the familiar voice is acoustically modified to have a shifted fundamental frequency (Holmes et al., 2018). Additionally, listeners in the Johnsrude et al. (2013) study exhibited a familiarity benefit, although smaller and only at the most adverse target-to-masker ratios (TMRs), when the target talker was unfamiliar and the masker was familiar. This suggests that listeners were able to use the familiarity of a voice as a cue for speech segregation, even when the familiar voice was not the intended target; however, this effect of a familiar masker benefit has not been replicated in more recent studies (Domingo, Holmes, & Johnsrude, 2019; Holmes et al., 2018)

The talker familiarity benefit is effective in a spatial release from masking paradigm. For example, Domingo, Holmes, Macpherson, et al. (2019) showed that overall performance on a closed-set speech understanding test was significantly better when the target talker was familiar vs. unfamiliar. Participants exhibited a talker familiarity benefit that was comparable to a 15° spatial separation between targets and maskers composed of unfamiliar talkers. The talker familiarity benefit was dependent on the SNR, such that when the SNR was more adverse, there was a larger spatial release from masking. When the SNR was higher than 0 dB, however, listeners were not as dependent on the familiarity cue, and relied more heavily on the acoustic cues of spatial separation and SNR.

Talker familiarity benefit can also correspond with a listener's ability to identify the voice of the target talker (Best et al., 2018; Rossi-Katz & Arehart, 2009; Sheffert et al., 2002). Nygaard and Pisoni (1998) employed a training paradigm for young adults in which they had to identify 10 separate talkers through a regimen over several days. Talker familiarity was confirmed through a voice identification test, and speech intelligibility of each of the trained talkers was then measured over a range of background noise levels for each listener. They found that speech intelligibility was better for the trained vs. untrained talkers. Using a similar training paradigm, Yonan and Sommers (2000) showed that older adults performed more poorly than younger adults at identifying the trained "familiar" talker. However, both groups received a significant intelligibility benefit when listening to the trained talker compared to the untrained talker.

Listeners are faster and more accurate at identifying the voice of a famous person than the voice of a person who became familiar through training (Maibauer et al., 2014; Schweinberger et al., 1997). Fontaine et al. (2017) tested the hypothesis that listeners utilize learned acoustic information from a talker when exhibiting a familiarity benefit on a talker identification task. They measured talker identification accuracy when the listeners were trained to identify a particular voice and when the familiar voice was a famous voice. The results suggested that listeners utilized the acoustic information when identifying famous voices, but did not use these features when the familiar talker was a trained voice. When listening to the famous voices, accuracy was higher and response times were quicker on the talker identification task with increased amounts of acoustic information. The results from the trained voices, however, showed that greater amounts of acoustical information resulted in poorer talker identification accuracy. This difference in listener performance for trained vs. famous voices shows that listeners benefit from a more long-term acoustic representation of a voice for it to act as a personally familiar talker.

As discussed in the previous two chapters, the familiarity of a talker's voice can influence stream segregation. While listeners are better able to extract target speech information from competing speech when the target talker is familiar (Johnsrude et al., 2013), it is not clear how the familiar voice could potentially cause attentional resources to divert from the target if the familiar speech is in a competing stream. Johnsrude et al. (2013) found that when a competing talker was familiar, there was a benefit in speech understanding, compared to when both the target and maskers were novel. This study will examine whether or not the presence of a familiar voice serving as a masker will affect speech segregation in a simulated cocktail party listening task.

Research Questions and Hypotheses

The goals of this study were: 1) to determine if listeners experience a benefit on speech understanding in a complex auditory environment with one or two competing talkers when the target talker is familiar; 2) to determine if listeners' ability to correctly identify a talker as familiar impacts the benefit of a familiar talker in a complex environment; 3) to determine if these familiarity effects persist when the masker is familiar; and 4) to determine if these effects vary as a function of listener age.

Hypothesis 1: Speech understanding performance will improve when the FV is the target; and decline when the FV is the masker. Based on previous research, it was expected that listeners would have a speech understanding improvement when listening to a familiar target talker versus to an unfamiliar voice. It was predicted, however, that when the familiar voice is a masker, it will draw attention away from the unfamiliar target voice when the FV is a masking signal, resulting in a familiar talker "penalty" to speech understanding.

Hypothesis 2: Speech understanding performance will be poorer when listening in a dichotic vs. monotic configuration. It was expected that an increase in the number of competing talkers would result in poorer speech understanding (Brungart, 2001b; Brungart & Simpson, 2002). Furthermore, it was expected that the dichotic signal (i.e., the masker in the non-test ear) would attract the listener's attention when the familiar voice is the masker in the non-target ear (Conway et al., 2001; Moray, 1959).

Hypothesis 3: The benefit received from a familiar voice will not be dependent on whether the listener can knowingly identify the voice as familiar or unfamiliar. In other words, regardless of whether the listener could accurately identify the familiar talker as familiar, the listener would still exhibit a familiar voice benefit on speech understanding. Previous studies have shown that speech understanding performance of a familiar voice did not correlate with listeners' accuracy at explicitly identifying that particular voice (Kreitewolf et al., 2017; Yonan & Sommers, 2000).

Hypothesis 4: Older listeners were expected to benefit more from a familiar target voice and to experience a greater detrimental effect from a familiar masking voice, compared to younger adults. In other words, younger and older listeners would be differentially impacted by the presence of a familiar voice. This effect would manifest as either a greater improvement in speech understanding when the familiar voice was the target talker, or a greater detriment to speech understanding when the familiar voice was a competing talker. It was predicted that there would be an interaction between the effects of talker familiarity and the age of the listener, because older adults have reduced available cognitive resources (CHABA, 1988; Lipnicki et al., 2017) and poorer speech understanding in difficult listening environments (Dubno et al., 1984).

Method

Participants

Pairs of normal-hearing adults aged 18-80 years old who were highly familiar with each other were recruited. A total of 15 younger normal-hearing adults (YNH; $\bar{x} = 32.7$ yrs ± 3.8), and 15 older normal-hearing adults (ONH; $\bar{x} =$ 62.9 yrs ± 4.4) completed this study. Familiar partners were defined as spouses who have cohabitated for at least 1 year. Normal hearing sensitivity was defined as thresholds ≤ 25 dB HL from 250-4000 Hz (ANSI, 2018). Participants completed the Montreal Cognitive Assessment (Nasreddine et al., 2005), and had passing scores of ≥ 26. Participants completed a high school level of education, and were native speakers of English. If one of the two familiar partners did not qualify for the study, the participant who did not qualify was then asked to record stimuli for their partner to listen to in the study (see Stimuli section). The participants from Study 1 were eligible to participate in Study 2. A total of eight participants (Younger: n = 6; Older: n = 2) completed both Study 1 and Study 2.

An *a priori* power analysis was conducted to determine a sufficient sample size for a repeated measures (RM) analysis of variance (ANOVA) with 10 measurements and a small effect size (0.25), power of 0.8, and alpha level of 0.05. Effect size was based on data previously reported by Johnsrude et al. (2013). Although the required sample size for the proposed study is 18 adults, planned multilevel models of the data will benefit from a larger sample. Data collection was completed over the course of two sessions. In the first session, informed consent was obtained, followed by the hearing and cognitive screening to verify if the participant met the inclusion criteria. All other experimental measures were completed in the second session.

Stimuli

Sentences from the Boston University Gerald (BUG) speech corpus (Kidd et al., 2008) were used as the stimuli for this study. The sentences follow the form “<name> <verb> <number> <adjective> <object>”, where there are eight possible word foils for each of the five categories. The target stimuli were defined by the name category in each sentence. The participants were instructed to listen for the sentence with “Bob” as the name. Sentences with the seven remaining names were used as maskers. The stimuli for a given listener (familiar and unfamiliar voices) were the same sex as each other to avoid adding a voice pitch cue that could result in better than expected performance.

Each participant recorded a unique subset of 240 sentences that were presented as the familiar voice for their spouse, and the recordings were potentially used as an unfamiliar voice for other participants. To guide all talkers to speak at the same rate, videos were created for each sentence to indicate the desired speed (Holmes, 2018) (Figure 3.1). The videos were compiled in Adobe Premiere Pro CC (Version 13), exported as a merged video, and played for the participant during the recording session.

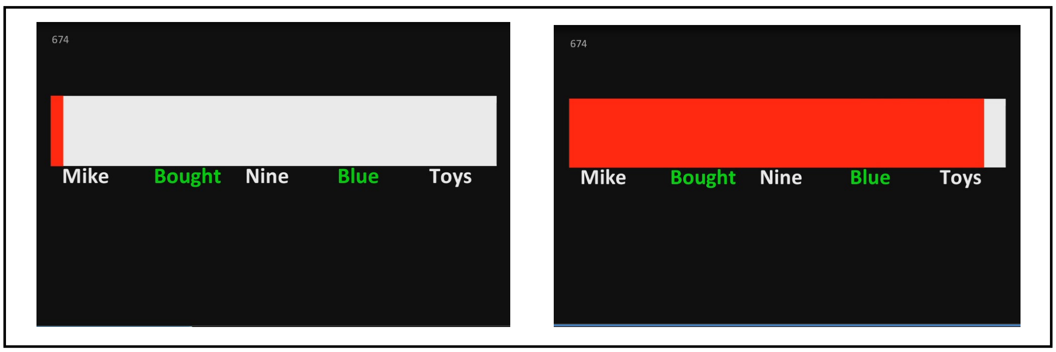
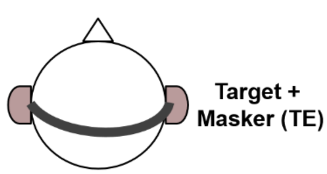


Figure 3.1. Two example frames from the video "Mike bought nine blue toys". The red bar progressed across the screen at a speed of 2.4 seconds per sentence.

The audio recordings were made in a sound-attenuating booth or in a quiet room using a Shure SM48 microphone and a Marantz PMD661 two-channel solid-state audio recorder. The pacing videos of each sentence were played while the talker recorded the sentences. Each sentence was recorded two times per talker and the sentence that was more intelligible was selected as the stimulus. The final sentences were equated in root-mean-square (RMS) level and a calibration tone was created to be equivalent to the RMS level of the sentence stimuli.

Procedures

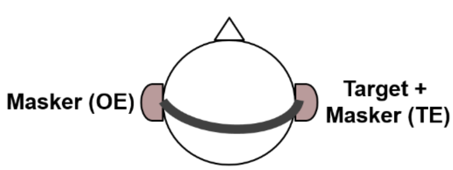
Speech Understanding for Monotic and Dichotic Conditions. There were two main listening configurations in this study: 1) monotic, and 2) dichotic. The three monotic conditions (Figure 3.2) include a target talker and one competing talker in the same ear. The conditions were as follows: a familiar target talker and an unfamiliar competing talker (Familiar Target), an unfamiliar target talker and a familiar competing talker (Familiar Masker), and an unfamiliar target talker and an unfamiliar competing talker (Both Unfamiliar). In a given trial, participants selected the words they heard from a closed set of eight possible foils for each word category using a touch screen monitor. Target sentences were presented at 65 dB SPL to the right ear, and the competing speech was presented at 0 dB and -5 dB SNR, also to the right ear. These levels were chosen to ensure audibility for the younger and older adults, and to avoid floor and ceiling performance. There were 20 trials for each condition at each SNR; these were presented in a randomized order such that each condition was presented twice in a block of 20 trials.



	TARGET EAR	
	Target	**Masker**
Familiar Target	**FAM**	Unf
Familiar Masker	Unf	**FAM**
Both Unfamiliar	Unf	Unf

Figure 3.2. Schematic of monotic condition configurations.

The four dichotic conditions always had a target voice and a single competing talker in the right ear, and a second competing talker in the opposite ear (Figure 3.3). The familiarity of the target and masker voices across each condition varied, and included the following conditions: a familiar target talker + two unfamiliar competing talkers (Familiar Target), an unfamiliar target talker + a familiar competing talker in the target ear with an unfamiliar talker in the opposite ear (Familiar Masker – TE), an unfamiliar target + unfamiliar competing talker in the target ear and a familiar competing talker in the opposite ear (Familiar Masker – OE), and an unfamiliar target talker + two unfamiliar competing talkers (All Unfamiliar). The target speech was presented at 65 dB SPL to the right ear, and the target ear masker was presented at 0 dB SNR; the opposite ear masker was presented at 0 dB SNR relative to the target ear (Brungart & Simpson, 2002). For both the monotic and dichotic conditions, sentences were scored for the number of target keywords correct. Prior to the start of the experiment, a practice block of 10 trials was administered to familiarize the listener with the task.



	TARGET EAR		OPPOSITE EAR
	Target	**Masker**	**Masker**
Familiar Target	**FAM**	Unf	Unf
Familiar Masker (TE)	Unf	**FAM**	Unf
Familiar Masker (OE)	Unf	Unf	**FAM**
All Unfamiliar	Unf	Unf	Unf

Figure 3.3. Schematic of dichotic condition configurations.

Talker Identification. A talker identification task was completed after each trial in the monotic and dichotic conditions to determine whether listener accuracy in identifying the familiarity of a talker is related to the effect (benefit or detriment) of familiar talkers as target or masker talkers. Immediately following each trial in the speech understanding experiment, the listener was prompted to identify whether the target talker was familiar (i.e. their spouse) or unfamiliar. There were 20 trials for each condition and at each SNR. Prior to the start of the experiment, a practice block of 10 trials was administered to familiarize the listener with the task.

Cognitive Measures. Participants completed several cognitive measurements from the domains of WM, selective attention and inhibition, and speed of processing. All subtests, with the exception of the LSPAN, were measured using the tablet-based NIH Toolbox (Gershon et al., 2013). The raw scores and response times from these measures are included in the planned analyses (see Data Analysis section).

Working memory capacity was measured using the List Sorting Working Memory Test (LSWM) from the NIH Toolbox (Tulsky et al., 2014). A series of items were presented visually and auditorily on the tablet. The participant was asked to repeat the items back in ascending size order (e.g. blueberry to elephant). The LSWM score is the number of items accurately reordered.

Selective attention and inhibition were measured using the Flanker test (Eriksen & Eriksen, 1974) administered through the NIH Toolbox. The Flanker test measures accuracy and response time on a task that requires the participant to select the orientation of an arrow (left vs. right) that is surrounded by arrows that are facing the same direction (congruent), or opposite directions (incongruent) from the target arrow. Accuracy scores and response times were measured.

Data Analysis

Speech understanding scores from each participant on each trial were analyzed with a generalized linear mixed effects (GLME) model, using the buildmer package (Voeten, 2020), as previously described in Study 1. Separate models were run for the monotic (3 levels: Familiar Target, Familiar Masker, Both Unfamiliar) and dichotic (4 levels: Familiar Target, Familiar Masker – TE, Familiar Masker – OE, and All Unfamiliar) configurations. The input parameters to each model included the dependent variable (speech understanding score) and fixed effects of talker familiarity condition, SNR (0 and -5 dB), talker identification score (Correct vs. Incorrect), scores from the cognitive measures (LSPAN and Flanker tests), their interactions, and the maximal number of random effects possible. The main effect of interest was familiarity condition, which would indicate whether speech understanding was modulated by the role (target talker vs. masker) and location (target ear vs. opposite ear) of the familiar voice. Additionally, word error patterns were evaluated using GLME models to determine if confusions were biased towards the FV by calculating the proportion of correct and incorrect responses for each condition. During the buildmer model building process, the cognitive measures (LSPAN and Flanker tests) were both removed from the final models.

Results

Monotic Conditions: Target Talker and One Competing Talker

Speech Understanding. A GLME model was conducted to determine the contribution of talker familiarity, SNR, talker ID performance, and age group to speech understanding scores on a monotic listening task. The three monotic listening conditions were: Familiar Target, Familiar Masker, and Both Unfamiliar (Figure 3.2). The results from the final GLME model are shown in Table 3.1.

Table 3.1

GLME Model for Monotic Speech Understanding Conditions

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	1.042	0.160	6.516	<0.001	***
Condition: All Unfamiliar (ref)					
Familiar Target	1.441	0.199	7.248	<0.001	***
Familiar Masker	0.051	0.156	0.328	0.743	
Signal-to-Noise Ratio: 0 dB (ref)					
-5 dB	-0.259	0.157	-1.656	0.098	
Talker ID: Correct (ref)					
Incorrect	-0.732	0.495	-1.478	0.139	
Age Group: YNH (ref)					
ONH	-0.979	0.224	-4.378	<0.001	***
Interactions					
Familiar Target x -5 dB SNR	0.074	0.185	0.399	0.690	
Familiar Masker x -5 dB SNR	0.092	0.157	0.588	0.557	
Familiar Target x Talker ID	-0.997	0.522	-1.910	0.056	
Familiar Masker x Talker ID	-1.919	0.480	-4.001	<0.001	***
Familiar Target x ONH	-0.165	0.270	-0.613	0.540	
Familiar Masker x ONH	0.355	0.219	1.622	0.105	
Talker ID x - 5 dB SNR	-1.631	0.836	-1.951	0.051	
Talker ID x ONH	-0.920	0.786	-1.171	0.242	
ONH x -5 dB SNR	-0.326	0.208	-1.564	0.118	
Familiar Target x -5 dB SNR x Talker ID	1.134	0.872	1.301	0.193	
Familiar Masker x -5 dB SNR x Talker ID	1.003	0.828	1.211	0.226	
Familiar Target x -5 dB SNR x ONH	-0.165	0.243	-0.679	0.497	
Familiar Masker x -5 dB SNR x ONH	-0.154	0.213	-0.720	0.471	
Familiar Target x Talker ID x ONH	0.807	0.794	1.016	0.310	
Familiar Masker x Talker ID x ONH	1.677	0.757	2.215	0.027	*
Talker ID x - 5 dB SNR x ONH	2.797	1.084	2.580	0.009	**
Familiar Target x Talker ID x -5 dB SNR x ONH	-2.263	1.115	-2.030	0.042	*
Familiar Masker x Talker ID x -5 dB SNR x ONH	-2.651	1.079	-2.458	0.014	*
Random Effects	**Variance**	**SD**			
Subject Intercept	0.312	0.559			
Subject Familiar Target Slope	0.306	0.554			
Subject Familiar Masker Slope	0.197	0.444			
Subject Talker ID Slope	0.675	0.822			
Subject -5 dB SNR Slope	0.160	0.400			
Subject Talker ID x -5 dB SNR Slope	0.835	0.914			

***p<0.001. **p<0.01. *p<0.05.

The effect of most interest was whether talker familiarity differentially influenced speech understanding on a monotic listening task for YNH and ONH listeners. The proportion of correct scores for the speech understanding task were transformed into rationalized arcsine units (RAU; Studebaker, 1985), and were plotted for YNH and ONH listeners as a function of monotic listening condition (Figure 3.4) Although performance is plotted for each group separately, there was no significant interaction of Age x Condition ($p>0.05$). There was a significant main effect of talker familiarity ($p<0.001$), with higher scores achieved when the target talker was familiar than when the target talker was unfamiliar (i.e. Familiar Masker and Both Unfamiliar conditions). Performance in the Familiar Masker condition was not significantly different from performance in the Both Unfamiliar condition ($p>0.05$). There was also a main effect of age group, with YNH listeners performing better than the ONH listeners ($p<0.001$). The main effects of age and familiarity should be interpreted with caution as they were involved in higher-order interactions.

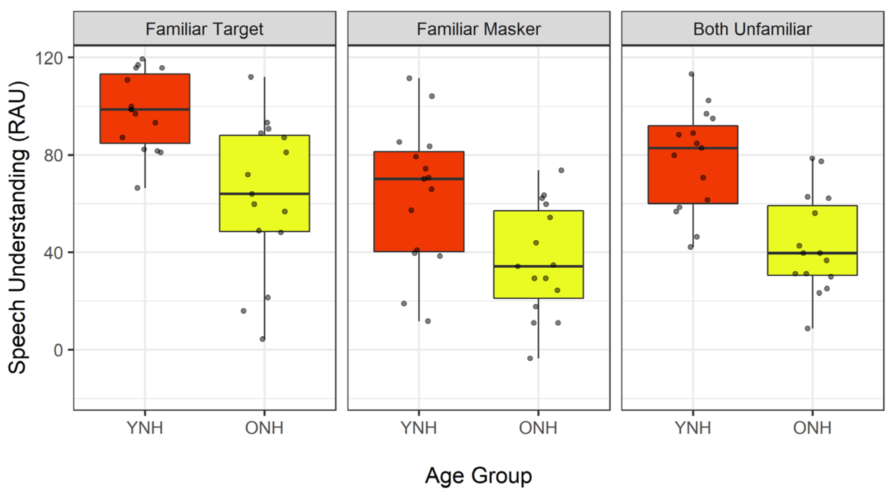


Figure 3.4. Speech understanding performance (RAU score) for each monotic condition, plotted separately for the YNH and ONH groups. The black dots represent individual subject scores averaged across all trials.

The primary effect of interest in this study was the effect of talker familiarity condition (Familiar Target, Familiar Masker, and Both Unfamiliar) across the two age groups (YNH vs. ONH). There were two significant three-way interactions, which in turn contributed to a significant four-way interaction between Condition, Target Talker ID, SNR, and Age. For reference, RAU scores for the target talker identification task were plotted with respect to monotic listening condition for both age groups (Figure 3.5). The individual points on both figures represent individual RAU scores for each participant, averaged across all test trials for a particular condition. While accuracy on the identification task was relatively high, performance on the Familiar Masker condition was significantly poorer than the Both Unfamiliar condition (*p*<0.001).

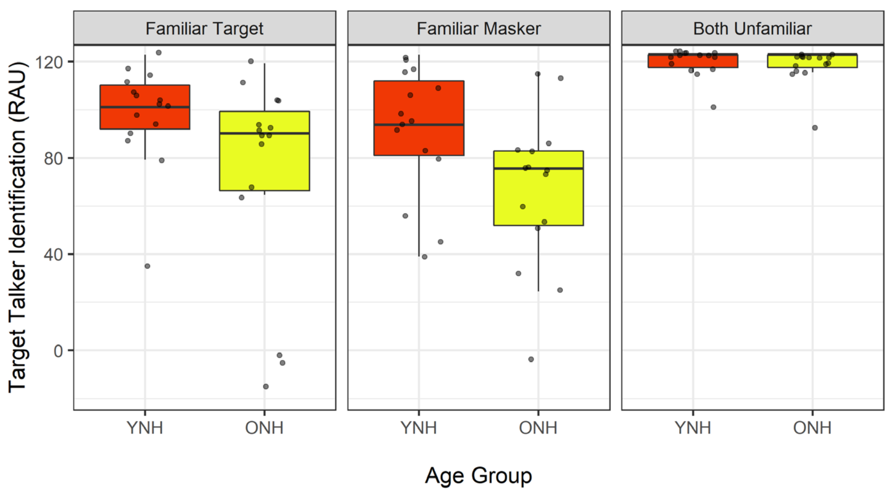


Figure 3.5. Target talker identification performance (RAU score) for each monotic condition, plotted separately for the YNH and ONH groups. The black dots represent individual subject scores averaged across all trials.

The first three-way interaction, shown in Figure 3.6, is between Condition, Target Talker ID, and Age. When the target talker identification was correct, the YNH group performed better in the Familiar Target condition than in the Familiar Masker and Both Unfamiliar condtions (*p*<0.001), but their performance did not differ between the Familiar Masker and Both Unfamiliar conditions. The ONH group, however, had significantly different scores across each of the three conditions when the target talker identification was correct. The ONH listeners performed best when the familiar voice was the target talker and poorest when both voices were unfamiliar. Speech understanding performance in the Familiar Masker condition was better than the Both Unfamiliar condition (β = 0.406, SE = 0.153, Z = 2.652, *p* = 0.008).

When the talker identification was incorrect, the YNH group's scores did not vary between the familiar target and both unfamiliar conditions (β = 0.444, SE = 0.531, = 0.836, *p* = 0.403), suggesting that talker identification accuracy may be related to talker familiarity benefit. Additionally, YNH listeners performed more poorly in the Familiar Masker condition than in the Familiar Target condition (p<0.001). The ONH listeners also demonstrated poorer performance on the Familiar Masker condition than the Familiar Target and Both Unfamiliar conditions when talker identification was incorrect. Comparing across groups, the YNH listeners performed better than the ONH listeners in the Familiar Target condition when the target identification was either correct (p<0.01) and incorrect (p<0.05). A similar pattern of performance was observed in the Unfamiliar Condition when the target talker identification score was correct. For all other contrasts, performance between groups was not significantly different (p>0.05).

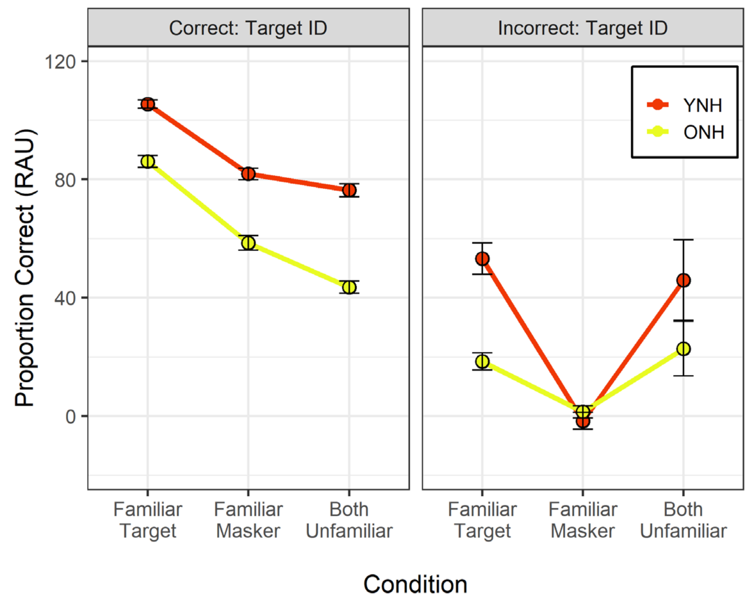


Figure 3.6. Speech understanding performance plotted with respect to talker identification accuracy as a function of listening condition. Scores were plotted separately for YNH and ONH listener groups. Error bars represent ± 1 SE.

The effect of talker identification was also impacted by age and SNR, and this three-way interaction is shown in Figure 3.7. Each of the three listening conditions was tested at 0 dB and 5 dB SNR. When talker identication scores were correct, YNH listeners did not perform differently in the two SNR settings, whereas the ONH listeners performed better when the SNR was 0 dB than -5 dB. Additionally, the YNH listeners performed better than the ONH listeners for both SNRs when talker identification was correct. A different pattern emerged when the talker identification scores were incorrect. In this circumstance, YNH listeners performed better in the 0 dB SNR condition than in the -5 dB SNR condition, and the ONH listeners performed equally poorly in both SNR conditions (β = -0.580, SE = 0.717, Z = -0.909, p = 0.418). The YNH listeners performed better than the ONH listeners at the 0 dB SNR condition, but did not perform significantly differently in the – 5 dB SNR condition (β = -0.571, SE = 0.845, Z = -.676, p = 0.499).

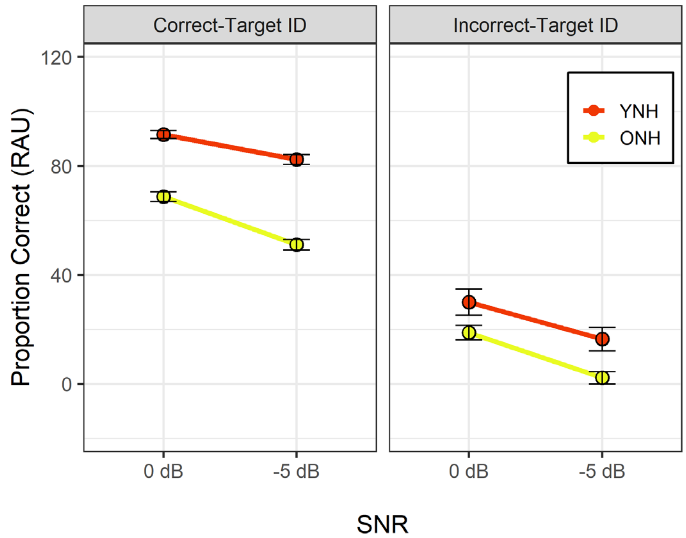


Figure 3.7. Speech understanding performance grouped across the monotic conditions, plotted as a function of SNR for Correct and Incorrect talker identification scores. RAU scores were plotted separately for YNH and ONH listener groups. Error bars represent ± 1 SE.

The pattern of performance of the two listener groups was influenced by all three of the other variables, as reflected in a four-way interaction between age group, condition, SNR, and target talker identification. As shown in Figure 3.8, the effect contributing to this interaction is the difference in performance of the YNH and ONH listeners, particularly in the Both Unfamiliar condition (right column of Figure 3.8). When the stimuli were presented at 0 dB SNR in the Both Unfamiliar condition, the YNH and ONH listeners performed better when they correctly identified the target talker than when the talker identification was incorrect, although the performance of the YNH listeners was better than that of the ONH listeners. This performance pattern changed when the Both Unfamiliar condition was presented at -5 dB SNR. In this case, the YNH listeners continued to show a decline in performance when the target talker identification was incorrect, but the ONH listeners showed no change in performance as a function of talker identification accuracy (β = 0.487, SE = 0.482, Z = 1.010, p = 0.313). The age-related differences (or lack thereof) were similar at the 0 and -5 dB SNRs for the Familiar Target and Familiar Masker conditions. Both the YNH and ONH listeners across these two conditions performed more poorly when the talker ID was incorrect vs. correct.

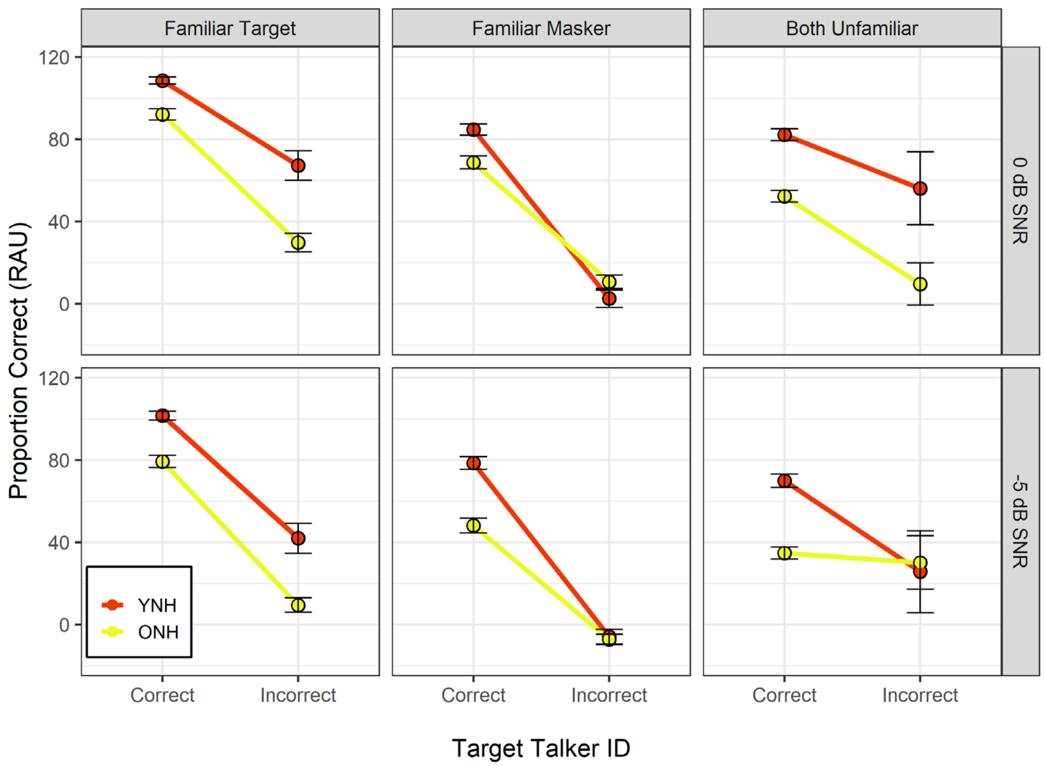


Figure 3.8. Speech understanding performance for the monotic conditions, plotted as a function of target talker identification score. Scores were plotted separately for YNH and ONH Groups, and for each of the listening conditions. Error bars represent ± 1 SE.

Error Analysis. An analysis was conducted to identify the types of errors participants made in each of the monotic conditions. The response errors were classified into two categories: masker and non-masker. A masker error occurred when the listener responded with a word spoken by the masker, and a non-masker error occurred when the listener responded with a word that was not present in that particular trial. The proportion of masker errors within a trial was calculated for each participant for all trials that had at least one incorrect target word. The masker error was of main importance to this analysis, thus all words with non-masker errors were excluded from the analysis. A GLME model was conducted to determine the contribution of talker familiarity, SNR, talker ID performance, and age group to the proportion of masker errors per trial in the monotic conditions. The results from the final GLME model are shown in Table 3.2.

Table 3.2

GLME Model for Monotic Errors

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	0.100	0.090	1.106	0.269	
Condition: All Unfamiliar (ref)					
Familiar Target	-0.756	0.146	-5.186	<0.001	***
Familiar Masker	-0.601	0.116	-5.188	<0.001	***
Signal-to-Noise Ratio: 0 dB (ref)					
-5 dB	0.276	0.078	3.524	<0.001	***
Target Talker ID: Correct (ref)					
Incorrect	-0.186	0.443	-0.421	0.674	
Age Group: YNH (ref)					
ONH	0.082	0.108	0.766	0.444	
Interactions					
Familiar Target x -5 dB SNR	-0.400	0.151	-2.639	0.008	**
Familiar Masker x -5 dB SNR	-0.075	0.127	-0.594	0.552	
Familiar Target x Target Talker ID	1.060	0.463	2.289	0.022	*
Familiar Masker x Target Talker ID	2.069	0.440	4.701	<0.001	***
Familiar Target x ONH	0.168	0.157	1.069	0.285	
Familiar Masker x ONH	0.218	0.130	1.683	0.092	
Target Talker ID x - 5 dB SNR	-0.873	0.460	-1.899	0.058	
Target Talker ID x ONH	1.109	0.548	2.022	0.043	*
Familiar Target x Target Talker ID x -5 dB SNR	1.427	0.497	2.869	0.004	**
Familiar Masker x Target Talker ID x -5 dB SNR	1.556	0.491	3.168	0.002	**
Familiar Target x Target Talker ID x ONH	-1.110	0.547	-2.029	0.042	*
Familiar Masker x Target Talker ID x ONH	-1.792	0.529	-3.386	<0.001	***

Random Effects	Variance	SD
Subject Intercept	0.036	0.189
Subject Target Talker ID Slope	0.358	0.598

****p*<0.001. ***p*<0.01. **p*<0.05.

The GLME analysis revealed significant main effects of condition and SNR. There also significant two- and three-way interactions involving the factors of Condition, SNR, Target Talker ID, and Age. The three-way interaction between Condition, Target Talker ID, and Age reflects that error proportions for age group were modulated by listening condition, as well as by whether the listener was able to correctly identify the familiarity of the target talker (Figure 3.9). Error proportion within a sentence was plotted for YNH and ONH listeners for each listening condition, and was plotted when talker ID scores were correct and incorrect. Both groups exhibited a greater proportion of masker errors when the talker identification was incorrect vs. correct for the Familiar Talker and Familiar Masker conditions (p<0.001). In the Familiar Talker condition, both groups showed a similar increase in error proportion from correct to incorrect talker identification scores (p>0.05). However, in the Familiar Masker condition, the YNH listeners had a greater increase in errors from correct to incorrect talker ID than the ONH listeners (β = -0.683, SE = 0.298, Z = -2.294, p = 0.02). In the Both Unfamiliar condition, the ONH listeners had a greater proportion of errors when the talker ID score was incorrect than the YNH listeners (β = 1.191, SE = 0.528, Z = 2.256, p = 0.024).The YNH listeners' errors did not vary across talker ID score in the Both Unfamiliar condition. Overall, these results suggest that with the exception of the Both Unfamiliar condition, both YNH and ONH have equally high error rates when they are unable to correctly identify the familiarity of the target talker.

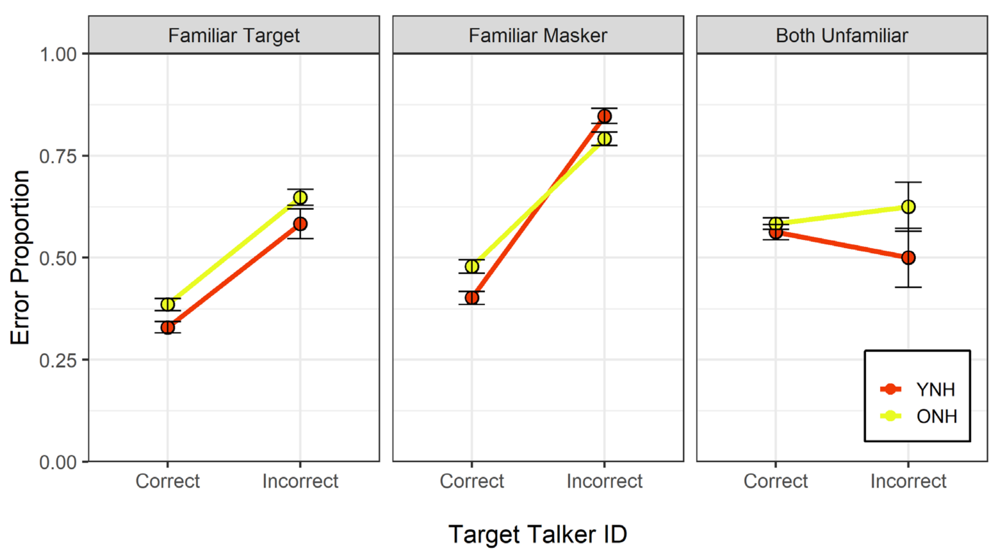


Figure 3.9. Proportion of masker errors for the monotic conditions, plotted as a function of when the talker identification score was “correct” or “incorrect”. Scores were plotted separately for YNH and ONH Groups. Error bars represent ± 1 SE.

Figure 3.10 shows that the error proportion for monotic conditions varied as a function of SNR and talker identification accuracy for each condition, reflecting the significant three-way interaction between Condition, Target Talker ID, SNR. Error proportions were collapsed across listener group for this analysis. For both the Familiar Target and Familiar Masker conditions, error proportions increased when the talker identification score was incorrect (relative to correct talker identification) for both the 0 dB and -5 dB SNRs ($p<0.001$). However, the magnitude of increase in error proportion differed for the two SNRs. There was a steeper increase in error proportion from correct to incorrect talker identification scores when the stimuli were presented at -5 vs. 0 dB for the Familiar Talker (β = -0.683, SE = 0.298, Z = -2.294, p = 0.02) and Familiar Masker (β = 0.554 SE = 0.189, Z = 2.935, p = 0.003) conditions. There was no difference across the two SNRs in the Both Unfamiliar condition when talker identification scores were incorrect. However, when talker identification scores were correct, error proportion was significantly higher when stimuli were presented at -5 dB ($p<0.001$). These findings suggest that the proportion of masker errors was higher in the more adverse SNR and when identification of target talker familiarity was incorrect. It is possible that listeners were biased towards responding to the "louder" voice (-5 dB SNR), which then resulted in an increase in errors when talker identification was incorrect. In the two conditions that had a familiar voice, either as a target or competing talker, error proportions did not vary greatly across SNR when the talker identification was correct. Thus, explicit target talker familiarity identification may be critical for reducing the effects of adverse SNRs.

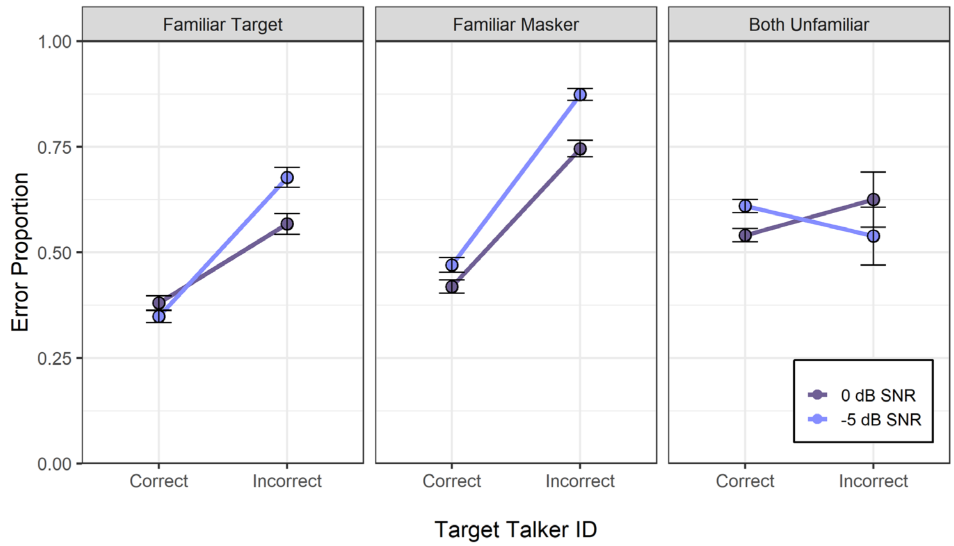


Figure 3.10. Proportion of masker errors for the monotic conditions, plotted as a function of when the talker identification score was “correct” or “incorrect. Scores were plotted separately for the 0 dB and -5 dB SNR conditions. Error bars represent ± 1 SE.

Dichotic Conditions: Target Talker and Two Competing Talkers

Speech Understanding. Speech understanding scores for the four dichotic listening conditions (1) Familiar Target, 2) Familiar Masker – TE, 3) Familiar Masker – OE, and 4) All Unfamiliar) were transformed to a RAU scale (Studebaker, 1985). A GLME model was conducted to determine the contribution of talker familiarity, talker ID performance, and age group on speech understanding scores on a dichotic listening task. The results from the final GLME model are shown in Table 3.3.

Table 3.3

GLME Model for Dichotic Conditions

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	0.010	0.190	0.055	0.956	
Condition: All Unfamiliar (ref)					
Familiar Target	1.648	0.175	9.437	<0.001	***
Familiar Masker (Target Ear)	0.221	0.155	1.422	0.155	
Familiar Target (Opposite Ear)	0.021	0.119	0.180	0.857	
Target Talker ID: Correct (ref)					
Incorrect	-1.277	0.607	-2.105	0.035	*
Age Group: YNH (ref)					
ONH	-0.871	0.272	-3.199	0.001	**
Interactions					
Familiar Target x ONH	-0.379	0.256	-1.481	0.139	
Familiar Masker (TE) x ONH	-0.228	0.229	-0.994	0.320	
Familiar Masker (OE) x ONH	-0.079	0.182	-0.435	0.663	
Familiar Target x Target Talker ID	-0.817	0.608	-1.345	0.179	
Familiar Masker (TE) x Target Talker ID	-0.823	0.616	-1.335	0.182	
Familiar Masker (OE) x Target Talker ID	0.065	0.666	0.098	0.922	
ONH x Target Talker ID	1.229	0.679	1.809	0.070	
Familiar Target x Target Talker ID x ONH	-1.052	0.686	-1.534	0.125	
Familiar Masker (TE) x Target Talker ID x ONH	-0.169	0.692	-0.244	0.807	
Familiar Masker (OE) x Target Talker ID x ONH	-0.594	0.747	-0.796	0.426	

Random Effects	Variance	SD
Subject Intercept	0.490	0.700
Subject Familiar Target Slope	0.260	0.510
Subject Familiar Masker (TE) Slope	0.221	0.470
Subject Familiar Masker (OE) Slope	0.103	0.321
Subject Target Talker ID Slope	0.308	0.555

***p<0.001. **p<0.01. *p<0.05.

The results of the GLME revealed significant main effects of condition, target talker identification, and age group, and no significant interactions. As shown in Figure 3.11, the YNH listeners performed better on the dichotic listening conditions than the ONH listeners (p<0.001). This effect did not vary by condition. The post-hoc releveling of the GLME model revealed that performance on the Familiar Target condition was significantly better than in all other listening conditions (p<0.001). The Familiar Masker conditions (Target Ear and Opposite Ear) and the All Unfamiliar condition did not differ significantly from each other. This suggests that in a dichotic listening task, both YNH and ONH listeners exhibit benefit from listening to a familiar target talker, but are not differentially affected when a familiar voice is used as a competing talker.

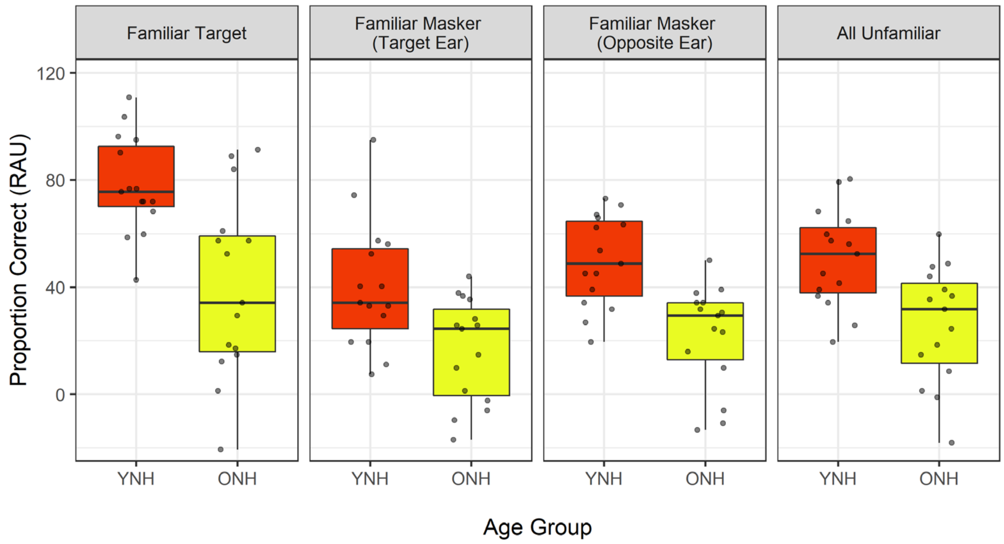


Figure 3.11. Speech understanding performance (RAU score) for each dichotic condition, plotted separately for YNH and ONH group. The black dots represent individual subject scores averaged across all trials.

Error Analysis. An analysis of the types of errors participants made in each of the dichotic conditions was conducted. The response errors were classified into three categories: masker (target ear), masker (opposite ear), and non-masker. The masker errors occurred when the listener responded with a word spoken by the masker, either in the masker from the target ear or the masker from the opposite ear, and the non-masker error occurred when the listener responded with a word that was not present in that particular trial. Because the masker errors were of main importance to this analysis, all words with non-masker errors were excluded from this analysis. A GLME model was conducted to determine the contribution of talker familiarity, error type, talker ID performance, and age group on word errors in the dichotic conditions. The results from the final GLME model are shown in Table 3.4.

Table 3.4

GLME Model for Dichotic Errors

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	-0.940	0.109	-8.595	<0.001	***
Condition: All Unfamiliar (ref)					
Familiar Target	-0.159	0.091	-1.740	0.082	
Familiar Masker (Target Ear)	0.024	0.077	0.309	0.758	
Familiar Masker (Opposite Ear)	-0.005	0.076	-0.063	0.950	
Error Type: Masker - Opposite Ear (ref)					
Masker - Target Ear	0.181	0.108	1.673	0.094	
Talker ID: Correct (ref)					
Incorrect	0.205	0.066	3.114	0.002	**
Age Group: YNH (ref)					
ONH	0.278	0.114	2.440	0.015	*
Interactions					
Error Type (TE) x ONH	-0.336	0.132	-2.558	0.011	*

Random Effects	Variance	SD
Subject Intercept	0.003	0.054

***p<0.001. **p<0.01. *p<0.05.

The proportion of errors that were categorized as Masker – Target Ear and Masker – Opposite Ear, varied across the YNH and ONH listener groups. Examination of the Error Type x Age interaction revealed that the YNH listeners did not differ in error proportion for the target ear and opposite ear masker errors (p>0.05). The ONH listeners, however, had a greater proportion of opposite ear errors than target ear errors (β = -0.156, SE = 0.075, Z = -2.068, p = 0.039), and had a greater proportion of opposite ear errors than the YNH listeners (p<0.05). This suggests that older adults committed more intrusion errors than younger adults when listening in a dichotic configuration. This may also indicate that older adults have increased difficulty suppressing competing speech, regardless of talker familiarity, in a complex environment. Additionally, the GLME analysis revealed a significant main effect of target talker identification. Consistent with previous analyses in this study, error proportion was lower when talker identification scores were correct vs. incorrect (p<0.01), and overall error proportion was lower for YNH listeners vs. ONH listeners (p<0.05).

Comparison of Monotic and Dichotic Conditions

Speech understanding scores were compared in monotic vs. dichotic configurations across the three listening conditions that were in common: Familiar Target, Familiar Masker, and All Unfamiliar (both target and masker talker are unfamiliar). In order to accomplish this comparison, performance scores in Familiar Masker – TE and Familiar Masker – OE were merged into a single condition of Familiar Masker. All scores were transformed to a RAU scale. A GLME model was conducted to determine the contribution of talker familiarity, configuration (monotic vs. dichotic), talker ID performance, and age group on speech understanding in a complex auditory task. The results from the final GLME model are shown in Table 3.5.

Table 3.5

GLME Model for the Monotic vs. Dichotic Conditions

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	1.058	0.158	6.679	<0.001	***
Condition: All Unfamiliar (ref)					
Familiar Target	1.403	0.188	7.478	<0.001	***
Familiar Masker	0.030	0.128	0.237	0.812	
Configuration: Monotic (ref)					
Dichotic	-1.061	0.117	-9.097	<0.001	***
Talker ID: Correct (ref)					
Incorrect	-1.237	0.398	-3.109	0.002	**
Age Group: YNH (ref)					
ONH	-1.009	0.221	-4.557	<0.001	***
Interactions					
Familiar Target x Dichotic	0.271	0.161	1.685	0.092	
Familiar Masker x Dichotic	0.076	0.126	0.600	0.548	
Familiar Target x Talker ID	-0.752	0.403	-1.868	0.062	
Familiar Masker x Talker ID	-1.496	0.390	-3.832	<0.001	***
Familiar Target x ONH	-0.131	0.252	-0.521	0.603	
Familiar Masker x ONH	0.409	0.177	2.311	0.021	*
Talker ID x Dichotic	0.724	0.413	1.753	0.080	
Talker ID x ONH	0.315	0.493	0.640	0.522	
Dichotic x ONH	0.210	0.164	1.277	0.202	
Familiar Target x Talker ID x Dichotic	-0.778	0.445	-1.748	0.080	
Familiar Masker x Talker ID x Dichotic	0.231	0.437	0.530	0.596	
Familiar Target x Talker ID x ONH	-0.116	0.473	-0.244	0.807	
Familiar Masker x Talker ID x ONH	0.595	0.461	1.291	0.197	
Familiar Target x Dichotic x ONH	-0.438	0.203	-2.156	0.031	*
Familiar Masker x Dichotic x ONH	-0.571	0.175	-3.258	0.001	**

Random Effects	Variance	SD
Subject Intercept	0.305	0.552
Subject Familiar Target Slope	0.278	0.527
Subject Familiar Masker Slope	0.090	0.301
Subject Dichotic Slope	0.079	0.282
Subject Talker ID Slope	0.381	0.618

***p<0.001. **p<0.01. *p<0.05.

There were significant findings for all main effects tested (listening condition, configuration, target talker ID, and age group). Main effects will be discussed in the context of their higher-order interactions. There were significant two-way interactions of Condition x Talker ID and Condition x Age, as well as a significant three-way interaction of Condition x Configuration x Age. The interaction of interest in this analysis was whether talker familiarity benefit varied across the listening configurations (Monotic vs. Dichotic). To analyze the three-way interaction of Condition x Configuration x Group, speech understanding scores were plotted as a function of configuration (monotic vs. dichotic) and age group (YNH vs. ONH) for each of the three listening conditions (Figure 3.12). Upon inspection, both groups performed more poorly on the dichotic conditions than the monotic conditions ($p<0.001$), and the ONH listeners performed more poorly than the YNH listeners in the monotic and dichotic presentation modes ($p<0.001$). The two age groups performed differently on the monotic vs. dichotic presentation configurations across the three listening conditions. Both groups exhibited a similar decline in speech understanding score from the monotic to dichotic presentation modes in the Familiar Target ($p>0.05$) and Both Unfamiliar conditions ($p>0.05$). However, in the Familiar Masker condition, the ONH group showed a significantly greater change in performance across configurations than the YNH listeners (β = -0.362, SE = 0.159, Z = -2.276, p = 0.023). This finding suggests that in the most challenging condition (Familiar Masker), the ONH listeners were more negatively impacted by a dichotic vs. monotic listening environment than the YNH listeners.

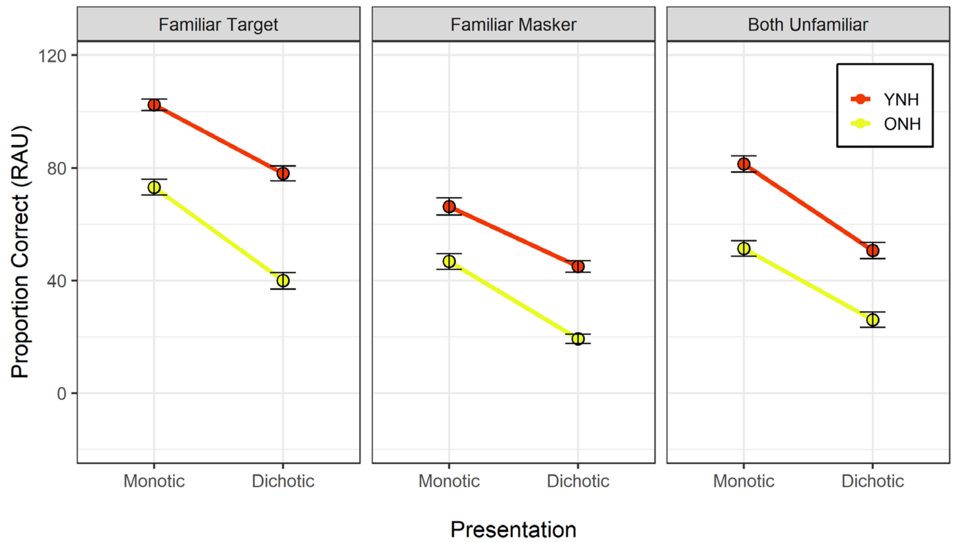


Figure 3.12. Speech understanding performance for the three listening conditions, plotted as a function of the Monotic and Dichotic configurations. Scores were plotted separately for the YNH and ONH groups. Error bars represent +/- 1 SE.

The interaction between Condition and Talker ID is plotted in Figure 3.13. Speech understanding scores were higher in the Familiar Talker condition than in the Familiar Masker (p<0.001) and Both Unfamiliar conditions (p<0.001) when talker identification was correct. When talker identification was incorrect, overall performance level was significantly lower than when talker identification scores were correct across all conditions, and speech understanding scores were significantly worse in the Familiar Masker condition than the Familiar Target (p<0.001) and Both Unfamiliar (p<0.001) conditions. Scores were significantly better when the target talker ID was correct vs. incorrect for each of the listening conditions. These results are consistent with the monotic speech understanding scores, which also showed that performance was poorest in the Familiar Masker condition when the talker identification was incorrect.

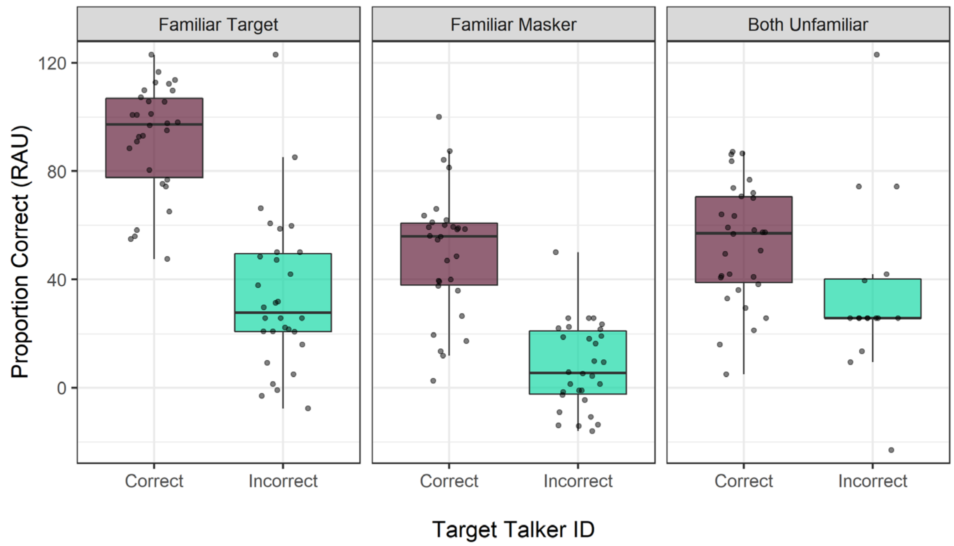


Figure 3.13. Speech understanding performance for the three listening conditions, combined across the Monotic and Dichotic configurations. Performance was plotted as a function of Talker Identification accuracy. The black dots represent individual subject scores averaged across all trials.

Discussion

Talker Familiarity Benefit for Monotic Conditions

The goals of this study were to: 1) to determine if listeners experienced a benefit in speech understanding in a complex auditory environment with one competing talker when the target talker was familiar; 2) to determine whether talker familiarity benefit was dependent on accurate identification of the familiarity of the target talker; 3) to determine if these familiarity effects persisted when the masker was familiar; and 4) to determine if these effects varied as a function of listener age.

It was hypothesized that younger and older listeners would demonstrate different patterns of speech understanding results with familiar vs unfamiliar target talkers, as well as familiar vs. unfamiliar maskers. This hypothesis derives from the notion that older adults often have reduced cognitive resources that support speech understanding (CHABA, 1988) and generally exhibit poorer speech understanding in difficult listening environments than younger adults (Dubno et al., 1984). As a result, it was expected that older listeners would rely more heavily on any cue that promoted speech segregation, including talker familiarity. If this hypothesis was correct, then ONH listeners would demonstrate better speech understanding scores with a familiar talker (re: an unfamiliar talker) and poorer speech understanding scores with a familiar masker (re: an unfamiliar masker) than YNH listeners, due to their reduced ability to inhibit the irrelevant signal with increasing age (Hasher & Zacks, 1988; Presacco et al., 2016b). Results revealed that YNH and ONH listeners had comparable familiarity benefit, such that performance on the Familiar Talker condition was better than on the Familiar Masker and the Both Unfamiliar conditions. The Familiar Masker and Both Unfamiliar conditions were not significantly different from one another for either listener group.

There were, however, different patterns of results noted between the listener groups when the talker identification accuracy was taken into account. In the Familiar Masker condition, which was expected to be a more challenging condition because it required listeners to inhibit the voice that was familiar, YNH listeners had a larger decline in performance when the talker identification accuracy was incorrect vs correct relative to the ONH listeners. This is in contrast to the hypothesis that ONH listeners would experience a greater decline in performance than the YNH listeners. One possible interpretation of this unexpected finding is that the YNH listeners' overall performance was superior to that of the ONH listeners, and therefore, the YNH listeners had more room for a larger score decrement. As a result, there was a greater difference in performance between the Familiar Target and Familiar Masker conditions for the YNH listeners compared to the ONH listeners. Previous studies have shown mixed evidence that a familiar masker is beneficial or harmful to speech understanding. Johnsrude et al. (2013) found that older adults demonstrated a familiarity benefit when the familiar voice was the masker vs. when both voices were unfamiliar. However, in a more recent study by Domingo, Holmes, and Johnsrude (2019), younger adults did not show this pattern of performance, and demonstrated no difference in speech understanding when the familiar voice was the masker vs. when both the target and masker were unfamiliar voices.

The error analysis conducted on data from the monotic conditions provides additional evidence to support the theory that when the YNH listeners accurately perceived the familiar voice, they could segregate it correctly as the target or masker. Results showed that when the target talker identification was correct, the proportion of masker errors did not differ across the Familiar Target and Familiar Masker conditions (Figure 3.9). However, when the talker identification score was incorrect, there was a significant increase in masker errors when the familiar voice was the masker vs. target talker. Additionally, the YNH listeners showed a greater increase in masker errors from the Familiar Target to the Familiar Masker condition than the ONH listeners. The ONH listeners had similar error patterns to the YNH listeners in the Familiar Target and

Familiar Masker conditions, regardless of whether the talker identification was correct or incorrect. This is inconsistent with the hypothesis that the ONH listeners would be more negatively impacted than YNH listeners by a familiar masker. Taken together, the familiar talker benefit for both YNH and ONH listeners was strongly impacted by whether the listener was able to correctly label the target talker as being familiar or unfamiliar. When listeners correctly identified the target talker as the familiar voice, they showed a talker familiarity benefit. However, when listeners incorrectly identified the target voice as being familiar, they did not demonstrate a talker familiarity benefit. In this instance, it was likely that listeners attended to the incorrect speech stream rather than not recognizing the voice as being familiar, which contributed to their very poor speech understanding performance. This is a different error as opposed to incorrectly assigning an unfamiliar voice selection to a familiar voice, which would suggest that the listener was unable to correctly recognize whether the voice was familiar or unfamiliar.

Previous studies showed that talker identification did not necessarily predict familiarity benefit. Yonan and Sommers (2000) trained younger and older adults to identify several different talkers, and then measured their speech intelligibility on each of those talkers. While the older adults performed more poorly than the younger adults in identifying the trained “familiar” talker, both groups received a significant intelligibility benefit when listening to the trained talker compared to the untrained talker, and thus the identification ability did not predict speech intelligibility performance. In the current study, participants were personally familiar with the “familiar” talker, which has been shown to result in better talker identification than when the familiar voice was trained (Fontaine et al., 2017). Furthermore, the talker identification task implemented by Yonan and Sommers (2000) required the listeners to recall the name of the target talker. The task in the current study assessed whether listeners could identify whether the target talker was familiar or unfamiliar. These differences in study design may have contributed to different findings across the two studies.

Talker Familiarity Benefit for Dichotic Conditions

It was hypothesized that the dichotic conditions (Familiar Target, Familiar Masker – TE, Familiar Masker – OE, and All Unfamiliar) would result in within- and across-ear word confusions between the target and competing talkers, and that these confusions would be modulated by the presence and location of a familiar voice. When the familiar voice was the target talker, performance was predicted to be better than in all other conditions. However, when a familiar masker was present in either the target or opposite ear, speech understanding performance was predicted to be significantly poorer than if all voices were unfamiliar.

The results from this study confirmed the familiarity benefit of a target talker for both YNH and ONH listeners when listening in a dichotic configuration (Figure 3.11). However, the familiarity of the masker did not influence performance. Previous studies on talker familiarity have only measured familiarity benefit in monotic configurations or binaural configurations where the same signals were presented to each ear (Domingo, Holmes, & Johnsrude, 2019;

Holmes et al., 2018; Johnsrude et al., 2013; Souza et al., 2013). The current study illustrates that in a cocktail-party like configuration simulated under headphones (Brungart & Simpson, 2002), both YNH and ONH listeners exhibited a talker familiarity benefit.

While the YNH and ONH listeners did not vary in their performance patterns across the four dichotic listening conditions, there was a significant difference in the type of errors they made across all of the conditions. Listeners could exhibit two types of masker-related errors in the dichotic conditions: Masker-Target Ear and Masker-Opposite ear. The Masker-Target ear confusion was when the listener reported the word spoken by the masker in the target ear, and the Masker-Opposite Ear confusion was when the listener reported the word spoken by the masker in the non-target ear. Across all dichotic conditions and regardless of talker familiarity of the target or competing talkers, the ONH listeners had a greater proportion of Masker-Opposite Ear errors than the YNH listeners. This finding is consistent with those reported in previous studies in which ONH listeners had greater difficulty segregating competing streams of speech (Helfer & Jesse, 2015; Jesse & Helfer, 2019), and were more likely to attend to irrelevant signals than the YNH listeners.

The greater intrusion errors from the masker in the opposite ear observed in the ONH listeners in the current study may be due to a reduction in inhibition by older adults (Hasher & Zacks, 1988; Presacco et al., 2016a; Tun et al., 2002). The ONH listeners may have had greater difficulty inhibiting the speech from the irrelevant ear (Familiar Masker – OE). Although listeners' performance on a test of inhibition (Flanker) did not significantly contribute to the models in this study, ONH listeners' poorer performance in the dichotic configuration relative to the monotic configuration may be related to their reduced ability to inhibit irrelevant stimuli. The Flanker assessment that was used to quantify inhibition is just one of many inhibitory measures available, and the results of this specific measure may not have fully captured listeners' ability to inhibit irrelevant stimuli.

Talker Familiarity Benefit for Monotic vs. Dichotic Configurations

It was hypothesized that talker familiarity benefit would be negatively impacted when listening in a dichotic vs. a monotic configuration. Based on previous studies, it was expected that an increase in the number of competing talkers would result in poorer speech understanding (Brungart, 2001b; Brungart & Simpson, 2002). The results from this study showed that overall performance on conditions in the dichotic configuration was significantly poorer than in the monotic configuration (Figure 3.12). Both YNH and ONH listeners exhibited this decline in performance across all test conditions. The decline in speech understanding performance in the dichotic vs. monotic conditions is consistent with findings from Brungart and Simpson (2002), who demonstrated that younger normal-hearing listeners exhibit a decline in speech understanding when a speech masker in the non-test ear was present versus when the masker was only present in the test ear. The present study confirms that the decline in speech understanding in the dichotic compared to monotic configurations is observed for both younger and older adults.

A benefit of a familiar talker on speech understanding was expected for both the monotic and dichotic configurations. The results for this study suggest that while the talker familiarity benefit is present for both configurations, certain familiarity conditions resulted in a greater change in performance for YNH and/or ONH listeners. The ONH listeners exhibited a greater decline in performance from monotic to dichotic configurations when listening in the Familiar Masker condition than the YNH listeners. Across all other conditions, the YNH and ONH listeners exhibited a similar change in performance from the monotic to dichotic configuration. This finding is consistent with the hypothesis that the masker in the non-test ear (present only in the dichotic configuration) would capture the listener's attention when that masker contained the familiar voice (Conway et al., 2001; Moray, 1959). Additionally, the differences in performance decrement across the listener groups with the dichotic configuration relative to the monotic configuration suggest that the ONH listeners were affected more than the YNH listeners with the addition of a second competing voice.

Conclusions

This study examined whether the presence of a familiar voice serving as either the target talker or as a masker would affect speech segregation in a simulated cocktail party listening task. It was hypothesized that speech understanding performance would improve when the familiar voice was the target and that a familiar masker would draw attention away from an unfamiliar target voice, resulting in a familiar talker "penalty" to speech understanding. It was also hypothesized that older listeners would benefit more from a familiar target voice and would experience a greater detrimental effect from a familiar masking voice, compared to younger adults. For the monotic conditions, the familiar talker benefit for both YNH and ONH listeners was affected by the listener's accuracy in judging the target talker as familiar or unfamiliar. When the talker identification performance was accurate, the difference between YNH and ONH listeners was consistent across all listening conditions. However, when the talker identification performance was inaccurate, the YNH listeners showed a greater decline in performance from the Familiar Target to the Familiar Masker than the ONH listeners. It also was hypothesized that when the stimuli were presented dichotically, there would be an increase in within-ear and across-ear word confusions between the target and competing talkers, especially when the masker was a familiar voice. The results showed that performance was significantly poorer in the dichotic vs. monotic configurations for both the YNH and ONH listeners. However, when listening in the dichotic configuration, both the YNH and ONH listeners had better speech understanding performance when the familiar voice was the target talker relative to all other dichotic conditions. Target talker identification did not differentially impact speech understanding on the dichotic conditions. Additionally, speech understanding scores did not differ when the familiar masker was in the target ear or was in the opposite ear. An analysis of the masker errors revealed that the ONH listeners had a greater proportion of across-ear confusions than the YNH listeners. This may be due to the ONH listeners' increased difficulty in inhibiting irrelevant speech signals.

Study 3: Effect of Talker Familiarity on Working Memory Capacity in a Competing Talker Task

Introduction

In complex auditory environments, speech understanding is affected by both auditory and cognitive factors (Arlinger et al., 2009; Humes et al., 2012). Stimulus factors such as the talker's gender, background noise level, and familiarity, contribute to a listener's ability to segregate target speech from background noise (Best et al., 2018; Brungart, 2001b; Darwin et al., 2003; Domingo, Holmes, Macpherson, et al., 2019; Kidd et al., 2005). These auditory factors contributing to speech understanding segregation can vary with respect to the bottom-up information available, as well as the top-down linguistic content. In particular, listeners rely on top-down processing to enhance speech understanding when the bottom-up message is degraded. Top-down processing can be impacted by cognition, which puts older adults at a disadvantage for understanding speech in a complex auditory environment.

Older adults require a higher SNR for maintaining speech understanding in noise compared to younger adults (Dubno et al., 1984; Helfer & Freyman, 2008). Dubno et al. (2002) measured the effect of masker noise modulation on speech understanding for younger and older adults. The results showed that while older adults had poorer performance in steady-state noise compared to younger adults, the older adults had a greater predicted benefit from listening in the presence of a modulated noise masker relative to steady-state noise.

However, the observed data showed the opposite, where younger adults had a greater benefit from listening in the presence of a modulated vs. steady-state noise masker. It was suggested that the difference between the predicted and observed data was due to age-related changes in threshold in steady-state noise. The use of speech as a competing signal has also revealed significant performance differences between younger and older adults. Helfer and Freyman (2008) evaluated the effect of speech understanding in the presence of two competing talkers where the competing talkers were either the same or the opposite sex from the target talker. For both younger and older adults, speech understanding in the presence of a same-sex masker resulted in poorer performance compared to maskers that were the opposite sex from the target talker. Interestingly, older adults were less able to ignore the competing speech when the maskers were of the younger adults.

The familiarity of a talker or masker voice can also improve a listener's speech understanding performance in noise. A benefit of 10-15% in speech intelligibility in noise was observed when listeners attended to a familiar talker in the presence of background noise (Johnsrude et al., 2013; Souza et al., 2013). Souza et al. (2013) recorded familiar participant pairs (spouses and friends) saying sentences from the Institute of Electrical and Electronics Engineers (IEEE) sentence corpus (IEEE, 1969). Speech intelligibility was measured in quiet and at two signal-to-noise ratios (SNR; +2 and +6 dB SNR) in the presence of competing six-talker babble or speech-shaped noise. Listeners showed a significant benefit of familiarity across all listening conditions and no differences across noise conditions.

Older adults appear to benefit from talker familiarity. Johnsrude et al. (2013) measured the impact of talker familiarity on a speech segregation task with one competing talker for older normal-hearing adults. Results showed a familiarity benefit, where intelligibility was best when the target was a familiar talker compared to when both the target and masker were unfamiliar. Additionally, listeners also exhibited a familiarity benefit, although smaller and only at the most adverse target-to-masker ratios (TMRs), when the target talker was unfamiliar and the masker was familiar. This suggests that listeners were able to use familiarity as a cue for segregation, even when the familiar talker was not the intended target. This finding is unique in the literature, and has not been replicated in subsequent studies (Domingo, Holmes, & Johnsrude, 2019).

There is an abundance of evidence to support the hypothesis that cognitive factors, particularly working memory and inhibitory control, have a strong influence on speech understanding in a noisy environment (Akeroyd, 2008; Ronnberg et al., 2008; Zekveld et al., 2013). Working memory is a limited capacity system that allows for the storage and manipulation of information in short term memory (Baddeley, 2000; Baddeley & Hitch, 1974). The Ease of Language Understanding (ELU) model suggests that when bottom-up constraints such as signal degradation, or the top-down process of reduced working memory capacity are present, speech recognition performance may be reduced (Rudner & Ronnberg, 2008).

Inhibitory control, which is the ability to suppress competing irrelevant signals while selectively attending to a target signal, has been shown to decline with age (Hasher & Zacks, 1988). Reduced inhibition can result in increased intrusion errors from competing signals and can result in reduced speech understanding in noise (Presacco et al., 2016a; Tun et al., 2009; Tun et al., 2002). Processing speed, which refers to the speed at which an individual can perform a particular task, has also been shown to decline with age (Salthouse, 1996). Processing speed and inhibition can be assessed with the Flanker Task (Eriksen & Eriksen, 1974). This test assesses accuracy and response time on a non-verbal task that requires the participant to select the orientation of an arrow (left vs. right) that is surrounded by arrows that are either facing the same direction (congruent) or opposite directions (incongruent) from the target arrow.

Working memory capacity can be quantified through several different types of methods, including the complex span and n-back paradigms (Conway et al., 2005). The R-SPAN test (Daneman & Carpenter, 1980) requires individuals to read a set of sentences and make a semantic judgement (was the statement true vs. false) about each sentence. At the end of a series of sentences, they are prompted to recall the final word from a set of the previous sentences. This test was adapted to create the L-SPAN, which conducts the span test in an auditory modality (Pichora-Fuller et al., 1995). Both the R-SPAN and L-SPAN measures have been correlated with listeners' ability to understand speech in noise, where higher span scores were associated with better speech understanding performance (Gordon-Salant & Cole, 2016; Schurman et al., 2014).

The n-back test is another paradigm for measuring working memory capacity. A series of stimuli are presented to the listener, who is asked to recall the item presented a certain number of trials previously (n-back), while concurrently attending to new stimuli presented. Kidd and Humes (2015) used a modified version of the n-back paradigm to measure spatial recall performance by younger and older adults. Word items were presented to different spatial locations and the listener had to identify when a word presentation was repeated at the same spatial location. The number of trials between repeated presentations to that spatial location (n-back) was varied. One of the results from this study was that the decline in recall performance was consistent for younger and older adults, suggesting that younger and older adults were not differentially impacted by the increasing memory load on the n-back test. However, Schurman et al. (2014) used a different modified version of the n-back procedure and found a significant effect of age on recalling sentences in a 1-back paradigm. Performance on the 1-back test was measured for high- and anomalous-probability sentences, presented in either speech-shaped noise or competing speech maskers. Speech intelligibility was equated to 80% performance using an adaptive tracking procedure where the competing signal was adjusted based on the number of incorrect responses on an immediate recall task, and recall accuracy on the 1-back paradigm was tested at those noise levels. The results showed that older adults had greater difficulty on the 1-back test than the younger adults, and that the younger adults derived a greater benefit from semantic context of the sentences than the older listeners on the 1-back test.

The authors concluded that the effect of age may be due to reduced working memory capacity for the older adults on this complex speech understanding task, particularly because performance on the 1-back test was highly correlated with performance on the LSPAN. Thus, the method used by Schurman et al. (2014) was a sensitive measure of working memory capacity and was able to differentiate performance between younger and older adults.

In addition to providing a speech intelligibility benefit, listening to a familiar voice may reduce the use of limited cognitive resources. The cognitive benefit could potentially result in higher working memory capacity when the target talker is a familiar vs. unfamiliar voice. Ingvalson and Stoimenoff (2015) assessed working memory ability of young adults on a visual digit recall task while participants concurrently repeated back sentences (primary task). Results showed that participants exhibited a familiarity benefit when the cognitive load was increased (i.e., in the conditions where they concurrently completed the speech recognition test and the number recall test). Familiarity benefit was not significant in the conditions where the listeners only completed the speech recognition test. This finding suggests that when the target talker was a familiar voice, there was a reduction in use of cognitive resources by younger adults, which resulted in more available resources for completing the digit recall task. Older adults were not tested on this paradigm, and thus, the impact of talker familiarity on working memory ability for older listeners is not yet known.

Research Questions and Hypotheses

The objective of this study was to measure the extent to which a familiar voice reduces the demand on available cognitive resources as measured by a test of working memory. Adults with a larger working memory capacity have better speech understanding in noise compared to their age-matched peers with lower WM capacity (Gordon-Salant & Cole, 2016). The working hypothesis is that listeners will perform better on a 1-back test (e.g., high memory load) when the target talker is a familiar voice than when the target talker is an unfamiliar voice, reflecting reduced demand on cognitive resources with the familiar voice. Performance on the 1-back task was expected to correlate with performance on a standard test of WM, the LSPAN (Schurman et al., 2014). It was expected that older adults would perform more poorly on the 1-back test than younger adults, but would show a larger benefit when the target talker was a familiar voice. This finding would be important because it would establish a connection between familiarity benefit and cognitive resource allocation.

Method

Participants

The same participants from Study 2 were recruited for Study 3. Pairs of normal-hearing adults aged 18-80 years old who were highly familiar with each other were recruited. A total of 15 younger normal-hearing adults (YNH; $\bar{x} = 32.7$ yrs ± 3.8), and 15 older normal-hearing adults (ONH; $\bar{x} = 62.9$ yrs ± 4.4) completed this study. Participants were native speakers of English and passed a screening test for mild cognitive impairment (MoCA).

Stimuli

The same stimuli recorded in Study 2 were used also in Study 3. Refer to Study 2 for a detailed description of the stimuli parameters. The target stimuli were defined by the name category in each sentence. In this study, the participants were instructed to listen for the sentence with “Mike” as the name. Target sentences from Study 2 were not repeated in the current study.

Procedures

Data collection was completed over two sessions. In the first session, informed consent was obtained, followed by the hearing and cognitive screening to verify if the participant met the inclusion criteria.

Cognitive Measures. Participants completed several cognitive measurements in the domains of WM, selective attention and inhibition, and speed of processing. All subtests , with the exception of the LSPAN, were measured using the tablet-based NIH Toolbox (Gershon et al., 2013). The raw scores and response times from these measures were included in the planned analyses (see Data Analysis section).

Working memory capacity was measured using the LSPAN test (Daneman & Carpenter, 1980). As previously described in Studies 1 and 2, participants listened to sets of two to seven sentences and completed a semantic question at the end of each sentence. Following a set of *n* sentences, the listener was asked to recall the final word of each sentence in the order in which it was presented. Selective attention and inhibition were measured using the Flanker test (as described in Study 2) (Eriksen & Eriksen, 1974).

Speed of processing was measured using the Flanker test and the Pattern Comparison Processing Speed test (PCPS; Carlozzi et al., 2015). For the PCPS test, participants indicated if two pictures were identical ("yes" response) or different ("no" response). The accuracy score reflected the total number of correct responses achieved in 90 seconds.

Tracking to 80% Speech Understanding. Speech intelligibility performance of each listener was equated at a level of 80% correct, in order to minimize differences in intelligibility as a confound. Each participant completed an adaptive tracking procedure to measure their speech recognition threshold (SRT) for 80% speech understanding accuracy (SRT_{80}) on an immediate sentence recall task. The target sentence was fixed at 65 dB SPL and the masker was adjusted adaptively in level across 20 trials. The masker was reduced by 8 dB times the proportion of correctly recalled words in the sentence, and then increased by 2 dB times the proportion of incorrectly recalled words in the sentence. This paradigm converged at an SNR that reflected 80% accuracy, and was repeated twice for each listening condition (Schurman et al., 2014). In this study, the familiar voice served as the target talker in the presence of speech shaped noise (FAM + SSN condition) or a competing unfamiliar talker (FAM + UNF condition). The familiar voice also served as the masker in the presence of an unfamiliar target talker (UNF + FAM condition). Baseline conditions with an unfamiliar target talker and SSN masker (UNF + SSN), and unfamiliar target and unfamiliar masker (UNF + UNF) were also tested. The target and masker were presented to the same ear for all conditions.

The 1-Back Memory Recall Task. The 1-back trials required the participant to listen to a sentence and hold it in memory, listen to a second sentence, and then recall the first sentence presented. The 1-back trials were tested at each individual listener's sequence of SNRs that produced a mean score of 80% speech intelligibility (SNR_{80}). Two blocks of 1-back trials were administered for each of the five conditions described above (FAM + SSN, UNF + SSN, FAM + UNF, UNF + FAM, UNF + UNF). In each block, 10 trials of immediate recall trials (0-back) were completed. The participant was then instructed that the next series of trials (n = 11) would be 1-back trials. Speech recognition testing was completed in approximately one hour.

Data Analysis

The SRT_{80} and the 1-back recall scores from each participant were analyzed separately using GLME models. As described in the previous chapters, the buildmer package (Voeten, 2020) was used to determine the final maximal model. Each model input included the fixed effects and interactions of listening condition and listener group (YNH vs. ONH), as well as predictors that included the Flanker, LSPAN, PCPS, years of familiarity, and the maximal number of random effects possible (Hox et al., 2017). All continuous factors (LSPAN, Flanker, PCPS, years of familiarity) were grand-mean centered. The main effect of interest was talker familiarity, and findings were expected to indicate whether performance on a test of WM capacity was modulated by the presence of a FV as either a target talker or a masker.

Results

Speech Recognition Thresholds in the 0-Back Task

The sound level of the competing signal (SSN or speech) was adjusted adaptively to determine the speech recognition thresholds for 80% intelligibility (SRT_{80}). Both the YNH and ONH listeners had SRTs that ranged from -5 to -7 dB SNR. A LME model was run to analyze the SRT_{80} scores (dependent variable) using the buildmer package (Voeten, 2020), with factors of age group, listening condition (Familiar Target + SSN, Unfamiliar Target + SSN, Familiar Target (1T), Familiar Masker (1T), and Both Unfamiliar (1T), and cognitive scores from the Flanker, PCPS, and LSPAN tests. Maximal random slope and intercept values were entered into the random effects structure. The results from the final model are shown in Table 4.1.

Table 4.1

LME Model for SRT80.

Fixed Effects	Estimate	SE	T	P	
(Intercept)	-5.617	0.942	-5.966	<0.001	***
Condition: Unfamiliar Target - SSN (ref)					
Familiar Target (SSN)	-0.355	0.858	-0.414	0.679	
Familiar Target (1T)	4.153	0.858	4.842	<0.001	***
Familiar Masker (1T)	7.556	0.858	8.809	<0.001	***
Both Unfamiliar (1T)	8.113	0.858	9.459	<0.001	***
Age Group: YNH (ref)					
ONH	0.767	1.196	0.641	0.522	
LSPAN (standardized)	-0.186	0.746	-0.250	0.803	
Flanker (standardized)	-0.263	0.817	-0.322	0.748	
Interactions					
Familiar Target (SSN) x LSPAN	-0.082	0.914	-0.089	0.929	
Familiar Target (1T) x LSPAN	-2.306	0.914	-2.524	0.012	*
Familiar Masker (1T) x LSPAN	-1.753	0.914	-1.918	0.055	
Both Unfamiliar (1T) x LSPAN	-0.314	0.914	-0.344	0.731	
Familiar Target (SSN) x Flanker	0.327	0.914	0.358	0.721	
Familiar Target (1T) x Flanker	-2.066	0.914	-2.261	0.024	*
Familiar Masker (1T) x Flanker	-1.069	0.914	-1.170	0.242	
Both Unfamiliar (1T) x Flanker	-1.699	0.914	-1.859	0.063	
Random Effects	**Variance**	**SD**			
Subject Intercept	3.028	1.74			

***p<0.001. **p<0.01. *p<0.05.

The primary goal of this analysis was to determine if listeners exhibited a talker familiarity benefit when speech intelligibility was fixed at a high level. There was a significant main effect of Talker Condition and significant two-way interactions of Condition x LSPAN and Condition x Flanker. Neither interaction included age group, and therefore data were collapsed across both YNH and ONH listener groups. The results from each listening condition are plotted as a function of LSPAN score in Figure 4.1. The left panel of the figure includes data from the two conditions with SSN maskers, and the right panel is a plot of the three competing speech conditions. The SRT_{80} scores for the SSN conditions did not vary as a function of target talker familiarity or increasing LSPAN score

(p>0.05). The competing speech conditions, however, elicited a different performance pattern. Higher LSPAN scores resulted in lower (better) SRT_{80} thresholds for the Familiar Target (β = -2.492, SE = 0.746, t = -3.34, p<0.01) and Familiar Masker (β = -1.939, SE = 0.746, t = -2.598, p<0.05) conditions. Performance on the Both Unfamiliar condition did not vary as a function of LSPAN. The improvement in score shown in the Familiar Talker condition was not significantly different than the Familiar Masker condition (β = .553, SE = 0.914, t = 0.605, p=0.546). Thus, performance on both the Familiar Talker and Familiar Masker conditions improved with increasing LSPAN score.

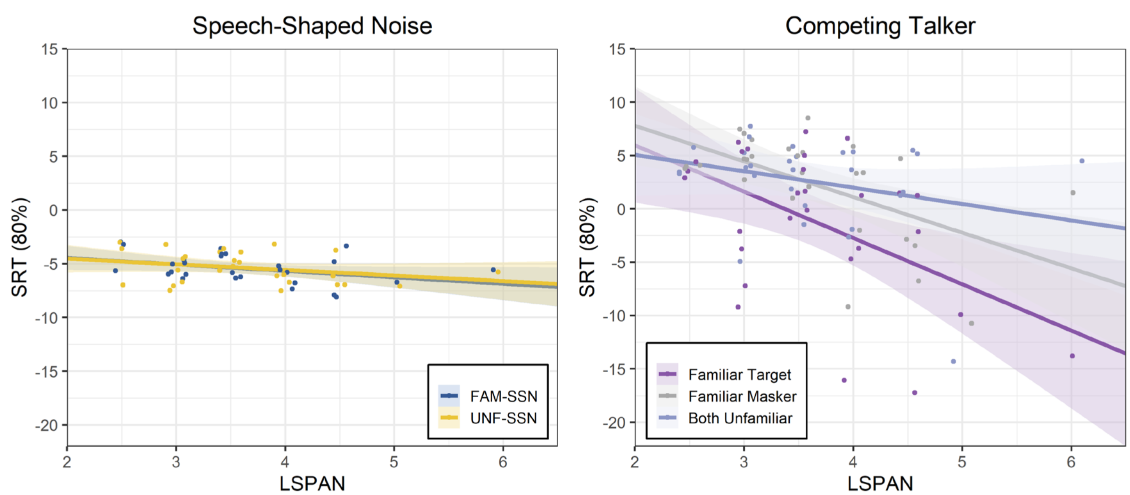


Figure 4.1. Scatterplot of SRT_{80} thresholds as a function of LSPAN score. SRT_{80} above 0 dB: target talker level > competing signal level; SRT_{80} thresholds below 0 dB: target talker level < competing signal level. Left panel: competing signal was SSN. Right panel: competing signal was 1 talker. Shaded area represents 95% confidence intervals.

The measure of inhibitory control was also found to be a significant predictor of SRT_{80} performance across the listening conditions. The interaction of Condition x Flanker is shown in Figure 4.2. The results from each condition are plotted as a function of the Flanker score (not corrected for participant age) from the NIH toolbox. Data were collapsed across the two listener groups. As described above, the left panel of the figure includes data from the two SSN conditions and the right panel includes data from the three conditions with a single competing talker. Performance in the SSN conditions did not vary as a function of Flanker score. However, as was the case with the LSPAN score, performance in certain conditions with a single competing talker did improve with increasing Flanker score. Both the Familiar Target (β = -2.329, SE = 0.817, t = -2.850, p<0.01) and Both Unfamiliar (β = -1.962, SE = 0.817, t = -2.400, p<0.05) conditions exhibited improved SRT_{80} thresholds with higher Flanker scores. Performance improvements were not significantly different between the Familiar Target and Both Unfamiliar conditions (p>0.05). However, in both conditions, listeners showed significantly greater improvement in SRT_{80} thresholds than in the two SSN conditions (p<0.05). Performance in the Familiar Masker condition did not vary significantly as a function of Flanker score (p>0.05), although there is a strong similarity in slope of the regression line between this condition and the Both Unfamiliar condition. These results suggest that individuals with higher (better) scores on the cognitive measures have better performance on certain competing speech conditions in an immediate recall task, whereas performance on the SSN conditions did not differ significantly with higher cognitive scores.

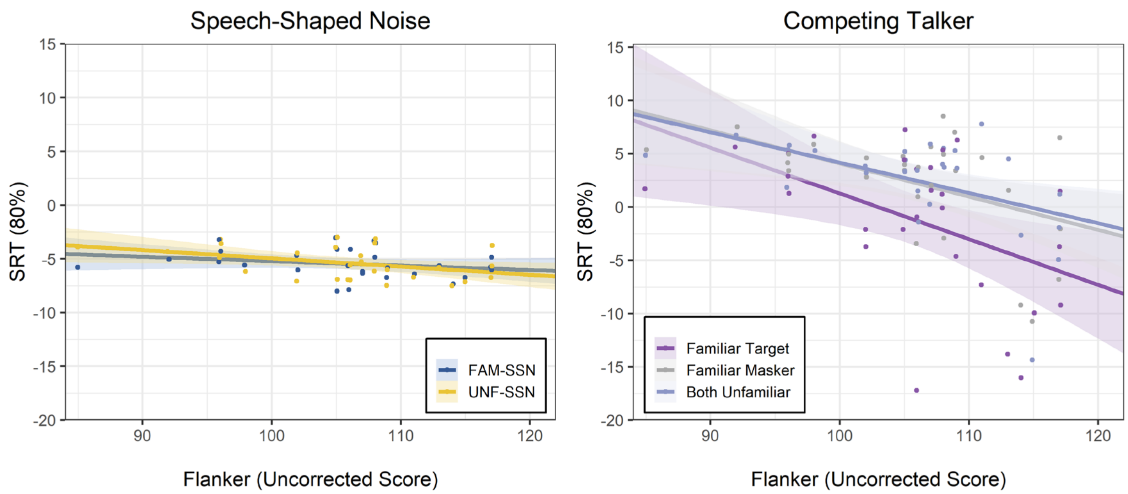


Figure 4.2. Scatterplot of SRT_{80} thresholds as a function of Flanker score (uncorrected for age). SRT_{80} above 0 dB: target talker level > competing signal level; SRT_{80} below 0 dB: target talker level < competing signal level. Left panel: competing signal was SSN. Right panel: competing signal was 1 talker. Shaded area represents 95% confidence intervals.

Proportion Correct Word Recall in the 1-Back Task

A GLME model was conducted to analyze performance on the 1-back test. The model included factors of age group, listening condition, SNR, cognitive scores from the Flanker, PCPS, and LSPAN tests, as well as maximal random slope and intercept values in the random effects structure. The results from the final GLME model are shown in Table 4.2.

Table 4.2

GLME Model for 1-back score.

Fixed Effects	Estimate	SE	Z	P	
(Intercept)	-0.069	0.135	-0.512	0.609	
Condition: Unfamiliar Target - SSN (ref)					
Familiar Target (SSN)	0.382	0.159	2.398	0.016	*
Familiar Target (1T)	-0.429	0.122	-3.504	<0.001	***
Familiar Masker (1T)	-0.541	0.132	-4.080	<0.001	***
Both Unfamiliar (1T)	-0.513	0.127	-4.055	<0.001	***
Age Group: YNH (ref)					
ONH	0.467	0.216	2.167	0.030	*
Flanker (Standardized)	0.289	0.095	3.031	0.002	**
LSPAN (standardized)	0.479	0.125	3.838	<0.001	***
SNR (standardized)	1.130	0.181	6.225	<0.001	***
Interactions					
Familiar Target (SSN) x ONH	-0.375	0.233	-1.609	0.108	
Familiar Target (1T) x ONH	-0.693	0.193	-3.580	<0.001	***
Familiar Masker (1T) x ONH	-1.054	0.229	-4.595	<0.001	***
Both Unfamiliar (1T) x ONH	-1.185	0.227	-5.216	<0.001	***
Familiar Target (SSN) x Flanker	-0.274	0.084	-3.267	0.001	**
Familiar Target (1T) x Flanker	-0.273	0.086	-3.184	0.001	**
Familiar Masker (1T) x Flanker	-0.137	0.085	-1.621	0.105	
Both Unfamiliar (1T) x Flanker	-0.084	0.086	-0.975	0.330	
Familiar Target (SSN) x LSPAN	0.084	0.146	0.572	0.567	
Familiar Target (1T) x LSPAN	-0.388	0.134	-2.883	0.004	**
Familiar Masker (1T) x LSPAN	-0.409	0.126	-3.244	0.001	**
Both Unfamiliar (1T) x LSPAN	-0.267	0.116	-2.308	0.021	*
Familiar Target (SSN) x SNR	0.308	0.275	1.118	0.263	
Familiar Target (1T) x SNR	-1.048	0.189	-5.552	<0.001	***
Familiar Masker (1T) x SNR	-1.016	0.197	-5.163	<0.001	***
Both Unfamiliar (1T) x SNR	-0.948	0.195	-4.875	<0.001	***
ONH x SNR	0.820	0.279	2.939	0.003	**
ONH x LSPAN	-0.290	0.179	-1.624	0.104	
LSPAN x SNR	0.631	0.176	3.593	<0.001	***
Familiar Target (SSN) x ONH x LSPAN	0.465	0.164	2.841	0.004	**
Familiar Target (1T) x ONH x LSPAN	0.516	0.186	2.769	0.006	**
Familiar Masker (1T) x ONH x LSPAN	0.777	0.176	4.408	<0.001	***
Both Unfamiliar (1T) x ONH x LSPAN	0.486	0.165	2.943	0.003	**
Familiar Target (SSN) x ONH x SNR	-0.530	0.415	-1.278	0.201	
Familiar Target (1T) x ONH x SNR	-0.744	0.294	-2.532	0.011	*
Familiar Masker (1T) x ONH x SNR	-0.055	0.318	-0.172	0.863	
Both Unfamiliar (1T) x ONH x SNR	-0.063	0.318	-0.199	0.842	
Familiar Target (SSN) x LSPAN x SNR	-0.079	0.244	-0.325	0.745	
Familiar Target (1T) x LSPAN x SNR	-0.796	0.185	-4.292	<0.001	***
Familiar Masker (1T) x LSPAN x SNR	-0.894	0.189	-4.742	<0.001	***
Both Unfamiliar (1T) x LSPAN x SNR	-0.752	0.183	-4.112	<0.001	***

Random Effects	Variance	SD
Subject Intercept	0.084	0.290

****p*<0.001. ***p*<0.01. **p*<0.05.

The results of the GLME analysis indicate that most main effects and interactions were significant. Of primary interest is the effect of condition on the listeners' 1-back score and how that varied with age. While the main effects of condition and age were significant, there were significant two-way interactions involving all main effects, as well as three-way interactions of Condition x Age x LSPAN, Condition x Age x SNR, and Condition x LSPAN x SNR.

Performance on the 1-back task, which was calculated as the proportion of words in a sentence that were recalled correctly, was found to vary by listeners' performance on the Flanker test. The interaction of Condition x Flanker is plotted in Figure 4.3, such that 1-back performance is plotted as a function of Flanker score for the SSN conditions (left panel) and the competing speech conditions (right panel). Performance on the 1-back improved with increasing Flanker score in the Unfamiliar Talker + SSN ($p<0.01$) and the Both Unfamiliar conditions (β = 0.205, SE = 0.097, Z = 2.124, $p<0.05$). Notably, these were the two conditions that did not include a familiar voice. The amount of improvement in 1-back performance with increasing Flanker score did not vary between the Unfamiliar Talker + SSN and the Both Unfamiliar conditions ($p=0.330$).

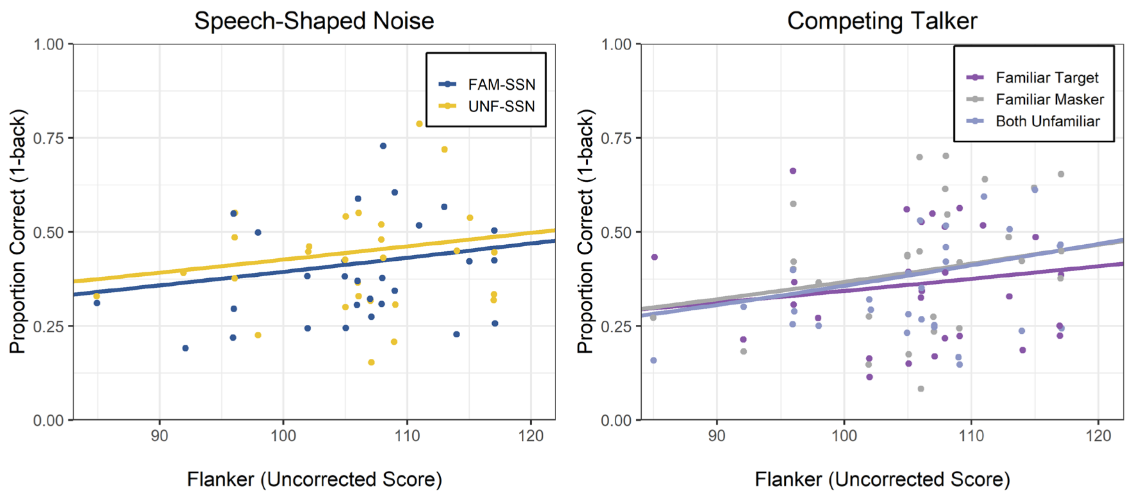


Figure 4.3. Scatterplot of 1-back proportion correct scores plotted as a function of Flanker score (uncorrected for age). Left panel: competing signal was SSN. Right panel: competing signal was 1 talker.

Fitted lines for 1-back scores in each listening condition were plotted as a function of LSPAN score for the SSN and 1T competing talker maskers, and separately for the YNH and ONH listeners (Figure 4.4). The YNH listeners showed an increase in 1-back score with increased LSPAN score for only the Familiar Talker + SSN ($p<0.001$) and Unfamiliar Talker + SSN conditions ($p<0.001$). However, in the competing speech conditions, the YNH listeners did not show an effect of LSPAN on n-back performance. Performance in the two SSN conditions did not differ significantly from each other ($p>0.05$), nor did they differ in the competing talker conditions ($p>0.05$). The ONH listeners showed a slightly different pattern of performance. Higher LSPAN scores resulted in higher n-back scores in all but the Unfamiliar Talker + SSN condition. Performance on the Familiar Talker + SSN condition significantly improved with higher LSPAN score ($p<0.001$), and the listeners exhibited a greater improvement in performance in the Familiar Talker + SSN condition than the competing speech conditions of Familiar Target (p<0.01) and Both Unfamiliar (p<0.05). The impact of higher LSPAN score on n-back performance was relatively similar across for both the YNH and ONH listeners. However, in the Familiar Masker condition, the ONH listeners had a greater improvement in n-back score with increasing LSPAN score than the YNH listeners (β = -0.437, SE = 0.176, Z = -2.487, p<0.05).

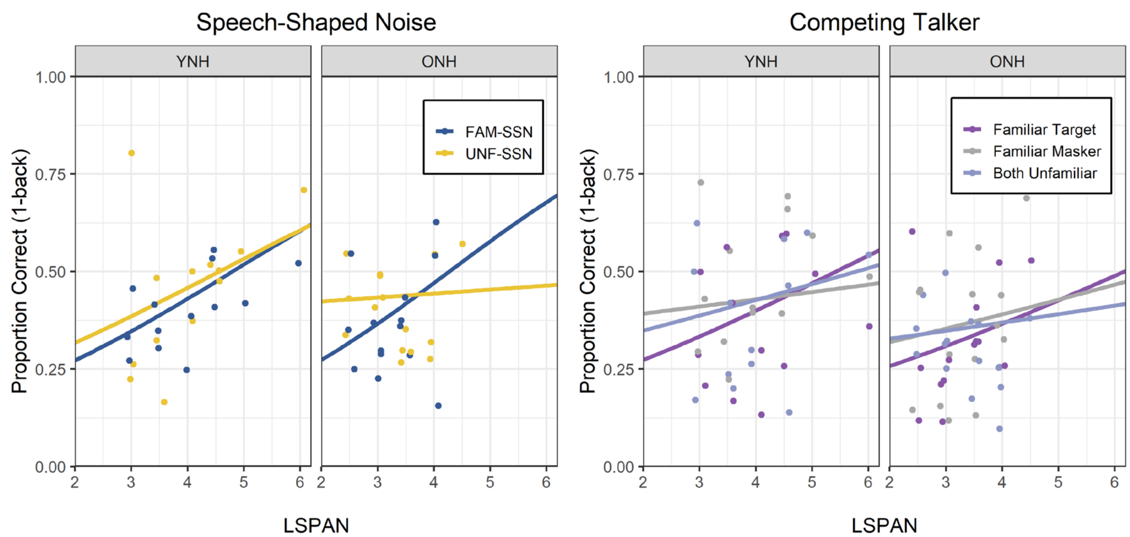


Figure 4.4. Scatterplot of 1-back proportion correct scores plotted as a function of LSPAN score, for YNH and ONH listeners. Left panel: SSN conditions. Right panel: competing talker conditions.

The interaction of listening condition and age also varied as a function of the SNR at which each 1-back trial was administered. The three-way interaction of Condition x Age x SNR is plotted in Figure 4.5. Fitted lines for 1-back scores in each listening condition were plotted as a function of SNR separately for the SSN masker (two left panels) and 1T competing talker maskers (two right panels), and separately for the YNH and ONH listeners (left and right panels for each masker type). Additionally, individual trial data points were plotted for each participant.

The figures should be interpreted with caution, as curves were drawn from the predicted model. There are regions within the figure, particularly for the ONH listeners, where no actual data were measured, however the regression curve shows a predicted value in that region.

There were significant differences within and across listener groups with changes in SNR for each of the conditions. The YNH listeners exhibited an improvement in 1-back score with higher SNRs in the Familiar Talker + SSN condition ($p<0.001$), Unfamiliar Talker + SSN condition ($p<0.01$), and the Both Unfamiliar condition($p<0.05$). The YNH listeners exhibited a greater increase in performance with increasing SNR in the two SSN conditions than in the competing talker conditions ($p<0.001$). There was no significant difference in 1-back score across the two SSN conditions, nor across the competing talker conditions ($p>0.05$). The ONH listeners demonstrated a significant improvement in performance with increasing SNR for all conditions. The benefit of increasing SNR was greater in the three competing talker conditions than the Familiar and Unfamiliar Talker + SSN conditions. Furthermore, within the competing talker conditions, the ONH listeners had less of an SNR benefit in the Familiar Talker condition than the Familiar Masker (β = 0.692, SE = 0.141, Z = 4.922, $p<0.001$) and Both Unfamiliar (β = 0.731, SE = 0.141, Z = 4.971, $p<0.001$) conditions. Comparing across listener groups, the ONH listeners exhibited a higher 1-back score with increasing SNR than the YNH listeners in three conditions: Unfamiliar Talker + SSN ($p<0.01$), Familiar Masker ($p<0.001$), and Both Unfamiliar ($p<0.001$) conditions. These results suggest that in all conditions that featured an unfamiliar target, ONH listeners exhibited a greater SNR benefit than YNH listeners on 1-back performance.

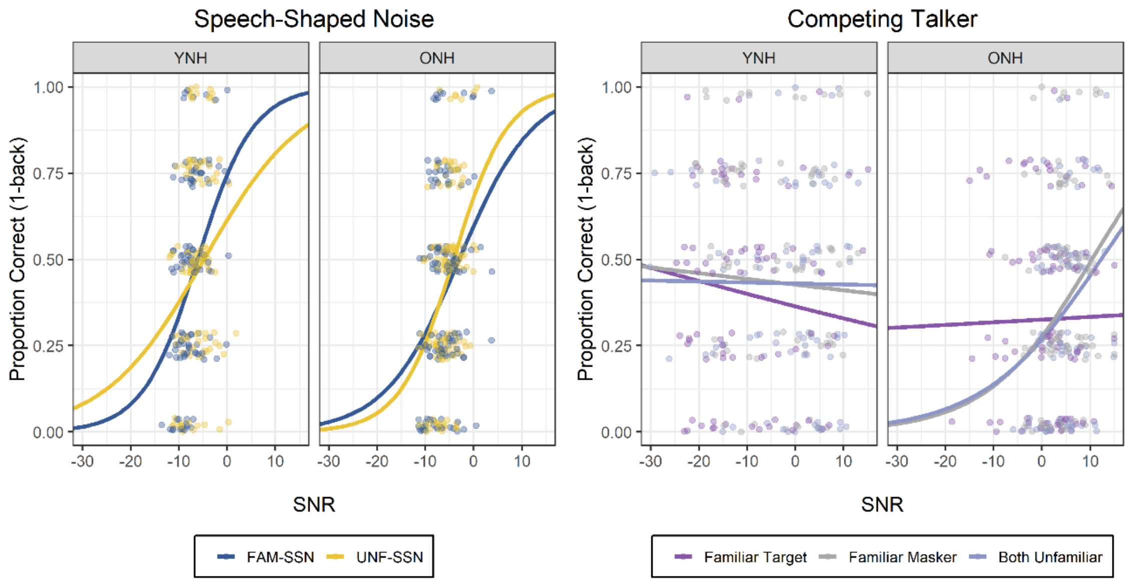


Figure 4.5. Scatterplot of 1-back proportion correct scores plotted as a function of SNR, for YNH and ONH listeners. Left panel: SSN conditions. Right panel: Competing talker conditions.

In addition to listener age group, the interaction of listening condition and SNR also varied as a function of working memory capacity. As shown in Figure 4.6, the 1-back performance in each listening condition was plotted as a function of SNR (data collapsed across listener groups). To depict the influence of LSPAN performance, the LSPAN scores were divided into values above the median (High LSPAN > 4) and at or below the median (Low LSPAN ≤ 4) score. Similar to Figure 4.5, the individual trial data for each participant were plotted to differentiate actual performance on the task with predicted performance (curve) from the GLME model.

There were significant differences in 1-back score with varying SNR and LSPAN score in each condition. The 1-back performance improved with increasing SNR and increasing LSPAN in the Familiar + SSN ($p<0.001$) and the Unfamiliar + SSN conditions ($p<0.001$). However, a different pattern of performance emerged when the competing signal was a single talker. Performance in the Familiar Target condition decreased with increasing SNR for listeners with higher LSPAN scores, and remained relatively stable across SNR for individuals with lower LSPAN scores (β = -0.140, SE = 0.055, Z = -2.521, $p<0.05$). The 1-back scores increased in the highest SNRs for individuals with lower LSPAN scores for the Familiar Masker (β = -0.223, SE = 0.059, Z = -3.751, $p<0.001$) and Both Unfamiliar conditions (β = -0.102, SE = 0.046, Z = -2.239, $p<0.05$). When comparing across the listening conditions, the Familiar + SSN and Unfamiliar + SSN conditions elicited a greater improvement in performance with higher SNRs and higher LSPAN scores than the competing talker conditions ($p <0.001$). The SSN conditions were not significantly different from each other, and the competing talker conditions were not significantly different from each other. Taken together, when listening in the SSN conditions, individuals with higher LSPAN scores were better able to take advantage of the more favorable SNRs than in the competing speech conditions.

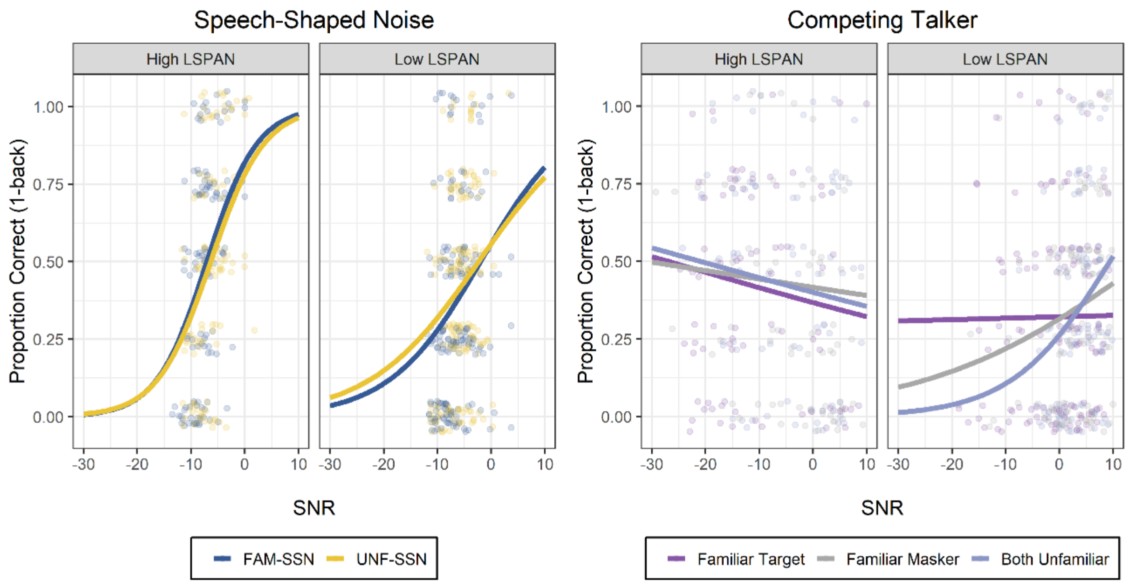


Figure 4.6. Scatterplot of 1-back proportion correct scores plotted as a function of SNR, for High (> 4) and Low (≤ 4) LSPAN scores. Left panel: SSN conditions. Right panel: competing talker conditions.

Discussion

The aim of this study was to measure the extent to which a familiar voice reduced the demand on available cognitive resources as measured by a test of working memory. Prior to measuring listener performance on the 1-back task (which was designed to test working memory), speech intelligibility performance was tracked to 80% correct on a 0-back task. The 1-back trials were then administered at SNRs used to maintain 80% speech understanding accuracy (SRT_{80}). The SRT_{80} was measured when the target talker was familiar or unfamiliar in the presence of a masker that was either a speech shaped noise (SSN) or a single competing talker (speech). The findings on the 0-back task generally showed a significant talker familiarity benefit when the competing signal was speech, but not when the competing signal was speech- shaped noise. Performance on the 1-back task was influenced by listener age and cognitive function. ONH listeners with higher LSPAN scores exhibited a familiarity benefit on the 1-back test, such that scores increased with increasing LSPAN score on the Familiar Talker + SSN condition but not in the Unfamiliar Talker + SSN condition.

Performance on the SRT_{80} Tracking Procedure

Cognitive Variables. The results on the 0-back task showed that greater available cognitive resources, particularly working memory capacity and inhibition, correlated with better SRTs when listening in a competing talker condition. There were no significant interactions of Flanker or LSPAN score in the SSN conditions, suggesting that performance in the SSN conditions did not vary as a function of cognitive scores. Furthermore, when taking listeners' cognitive scores into account, there was no difference in performance across the Familiar + SSN and Unfamiliar + SSN. In other words, when cognitive performance was assessed, individuals did not show a talker familiarity benefit when listening in the presence of a SSN. Souza et al. (2013) measured talker demonstrated that older hearing-impaired adults exhibit a familiarity benefit at +2 and +6 dB SNR. The listeners in that study may have exhibited a familiarity benefit because of their limited audibility associated with hearing loss. Participants in the current study, who had normal audiometric hearing thresholds, had SRT_{80} thresholds of -5 to -7 dB SNR, which are significantly lower (better) SNRs than the test SNRs presented in the Souza et al. (2013) study. Furthermore, when cognitive abilities were taken into account in the current study, listeners did not exhibit a benefit of talker familiarity for SSN. Previous studies on talker familiarity benefit have not included cognitive function as a test measure (Domingo, Holmes, & Johnsrude, 2019; Johnsrude et al., 2013; Souza et al., 2013).

In the competing speech conditions, performance on the Familiar Talker condition improved with higher LSPAN scores, as well as with higher Flanker scores. Similarly, improvement was observed in the Familiar Masker condition with increasing LSPAN score, and in the Both Unfamiliar condition with increasing Flanker score. These findings suggest listeners rely more on cognitive resources in competing speech conditions vs. when listening to speech in a steady-state noise masker. Zekveld et al. (2013) measured speech understanding performance for younger normal-hearing adults when speech intelligibility was adaptively tracked to 29% and 71% accuracy. The target speech was presented in the presence of a single competing talker, steady state noise, and fluctuating noise. They found that working memory capacity as measured through the RSPAN test (Daneman & Carpenter, 1980) was greatest for the competing speech masker at both intelligibility performance levels than for the noise maskers. This is in agreement with previous studies that evaluated the influence of working memory capacity on speech intelligibility in noise, and found that working memory was significantly correlated with performance when the masker was a competing talker. (Ronnberg et al., 2010; Rudner et al., 2012).

Performance in the Familiar Talker condition improved significantly for individuals with higher LSPAN and Flanker scores. This was observed for both younger and older listeners. Additionally, performance on the Familiar Masker condition improved with increasing LSPAN score, and performance on the Both Unfamiliar condition improved with increasing Flanker score. However, neither improvement in the Familiar Masker and Both Unfamiliar conditions significantly differed from that shown in the Familiar Target condition. Thus, it appears that individuals with higher working memory capacity or greater inhibitory processing were better able to take advantage of the target talker familiarity benefit than individuals with poorer cognitive processing. This is a novel finding, as previous studies that have assessed the benefit of listening to a familiar voice on speech segregation did not evaluate whether the magnitude of benefit was influenced by listeners' cognitive capacity (Domingo, Holmes, & Johnsrude, 2019; Domingo, Holmes, Macpherson, et al., 2019; Holmes et al., 2018; Johnsrude et al., 2013).

Age Group. The effect of age group did not interact with cognitive score or listening condition. Specifically, the ONH listeners did not require a higher SNR to achieve 80% intelligibility than the YNH listeners, and performance for either listener group did not vary by talker familiarity. Previous studies have shown an inconsistent effect of listener age on talker familiarity benefit. Johnsrude et al. (2013) assessed talker familiarity benefit across a range of TMRs (-6 to +6 dB TMR) for normal-hearing adults who ranged in age from 44-79 years old. Their analysis revealed that the Younger (<55 years) and Older (≥ 55 years) listener groups did not differ in performance across the test conditions. However, when performance was analyzed when listener performance was at 76% speech intelligibility and age was used as a continuous variable, there was a significant correlation between age and talker familiarity. Performance in the

Familiar Target condition did not vary across the span of ages. However, speech understanding performance was significantly poorer in the Both Unfamiliar condition with increasing age.

The variance in results between the current study and Johnsrude et al. (2013), may due to differences in the age range of the younger adult participants. The current study included YNH participants that ranged in age from 27-40 years, whereas the younger adults in the Johnsrude et al. (2013) ranged from 44-54 years old. Additionally, the current study analyzed data using a between-subjects factor of listener group (YNH and ONH). The Johnsrude et al. (2013) did not find an effect of age group when age was analyzed as a grouping variable. When they conducted correlation analyses with age (years) as a continuous variable, Johnsrude et al. (2013) found that speech understanding performance did not vary with increasing age when the target talker was familiar. Conversely, Domingo, Holmes, and Johnsrude (2019) found that speech understanding in noise performance of a familiar target talker declined with increasing age (28 – 82 years). The two previous studies did not directly measure cognitive function, thus it may be that the age effects seen may be more related to cognitive function of the individual participants, as was seen in the current study.

Performance on the 1-back Test

Influence of Inhibitory Control, Working Memory Capacity, and SNR. Performance on the 1-back test was influenced by individuals' performance on cognitive measures. For both listener groups, 1-back scores on the Unfamiliar + SSN and the Both Unfamiliar conditions improved with increasing Flanker score.

Interestingly, performance in conditions that included a familiar voice as either a target talker or as a masker did not vary significantly with increasing Flanker score. One possibility for why the Flanker did not correlate with performance on the competing talker conditions is because a different cognitive process, such as working memory capacity, was more related to recall ability with a speech masker.

It was hypothesized that performance on the 1-back task would correlate with performance on a test of working memory (LSPAN). Previous studies have shown that adults with a higher working memory capacity have better speech understanding in noise compared to their age-matched peers with lower working memory capacity (Gordon-Salant & Cole, 2016). The results of the current study showed a significant three-way interaction between listening conditions, LSPAN score, and SNR (Figure 4.6). When completing the 1-back task in the presence of SSN, individuals with higher LSPAN scores were better able to take advantage of the more favorable SNRs, and thus had higher 1-back scores than individuals with lower LSPAN scores. However, when listening in the competing talker conditions, performance for the individuals with higher LSPAN scores decreased with increasing SNR. This is surprising, because individuals with higher working memory capacity demonstrated better speech understanding in noise in the current study and in those reported previously (Fullgrabe & Rosen, 2016; Gordon-Salant & Cole, 2016).

The finding that individuals with higher LSPAN scores had poorer 1-back recall with higher SNRs may be related to the listeners' performance on the 0-

back test. As previously mentioned, the SNRs for the 1-back test were selected from the trials in the 0-back test that were used to track 80% intelligibility. Thus, the individuals with lower LSPAN scores may have required higher SNRs to maintain 80% intelligibility than the individuals with higher LSPAN scores. While the SNRs elicited equal intelligibility across listeners, the SNRs may still have been sufficiently aversive to increase cognitive demand on the memory recall task. Schurman et al. (2014) measured 1-back performance in competing speech and speech-shaped noise and reported significant correlations between 1-back performance and LSPAN score. However, they did not analyze performance relative to the test SNR, as was done in the current study. Thus, it is unclear whether these findings are consistent with previous studies that utilized a 1-back method for assessing working memory capacity.

Age Group. It was hypothesized that younger and older listeners would perform better on a 1-back test (e.g., high memory load) when the target talker was a familiar voice compared to an unfamiliar voice, reflecting reduced demand on cognitive resources with the familiar voice. The results generally showed that performance on the 1-back test did not vary with the familiarity of the target talker's voice, except for the ONH listeners in the SSN conditions. Both younger and older listeners exhibited improved 1-back performance with increasing LSPAN score. This pattern was observed in all of the SSN conditions for both groups. However, the 1-back scores for the ONH listeners were higher in the Familiar vs. Unfamiliar Talker + SSN condition. This finding is novel and adds to the evidence that talker familiarity is beneficial for understanding speech in noise

(Domingo, Holmes, & Johnsrude, 2019; Domingo, Holmes, Macpherson, et al., 2019; Johnsrude et al., 2013), and that talker familiarity can improve performance when listening in a situation that causes a greater strain on cognitive resources, as simulated on the 1-back task.

There was a talker familiarity benefit when performance on the 1-back test interacted with SNR. Individuals that required a higher SNR to achieve 80% intelligibility were then tested at these higher SNRs in the 1-back paradigm, and individuals that required a lower SNR were tested at those more adverse levels. Performance on the Familiar Masker condition improved with increasing SNR for the ONH listeners, and this improvement was larger than that shown in the Familiar Target condition. Thus, higher SNRs were more beneficial to ONH listeners when the competing voice was their spouse, whereas the YNH listeners only showed an improvement in 1-back score with increasing SNR in the two SSN conditions.

The finding that YNH and ONH listeners have differences in talker familiarity benefit in a high memory load task is novel. Previous studies on talker familiarity benefit have not evaluated performance in relation to working memory capacity. Additionally, previous studies that examined the effect of age on 1-back recall did not evaluate this effect with regard to individual trial SNRs. The SNRs at which the YNH and ONH listeners were tested in the current study ranged from -20 to +10 dB SNR. The YNH listeners exhibited a wide range in SNRs, whereas the ONH listeners were primarily tested at 0 to +10 dB SNR. Schurman et al. (2014) implemented a similar 1-back task, and measured performance for

YNH and ONH listeners when sentence context and type of background noise were manipulated. They found that ONH listeners exhibited poorer performance on a similar 1-back test as compared to YNH listeners for essentially all test conditions. However, performance was not evaluated with regards to trial-by-trial SNRs.

One possible reason for the ONH listeners' improved performance on the 1-back test with a familiar masker may be due to the range of SNRs in which they were tested. The SNR for each 1-back trial was based on SNRs that were used to adaptively track to 80% speech intelligibility in the 0-back test. As shown in Figure 4.5, the ONH listeners required SNRs that were generally above 0 dB when listening in competing speech, whereas the YNH listeners' SNRs were spread across a wide range in values. Thus, the familiar masker benefit may have been generated from a restricted range of SNRs at which the ONH listeners were tested. Conversely, the YNH listeners did not exhibit a talker familiarity benefit on the 1-back test, regardless of SNR or LSPAN score. While performance in several of the conditions was significantly higher than for the ONH listeners, there was not an instance where performance by YNH listeners on a familiar talker condition was greater than when the target talker was unfamiliar. The YNH listeners may not rely as heavily on the familiarity cue for maintaining accuracy on a task of increased working memory load.

Conclusions

The objective of this study was to measure the extent to which a familiar voice reduces the demand on available cognitive resources as measured by tests of working memory and inhibition. It was hypothesized that the familiarity of a talker would impact performance on a 1-back test, such that performance would improve when the target talker was a familiar vs. unfamiliar voice, and that ONH listeners would show a larger familiarity benefit than the YNH listeners. On the immediate recall task, where the SNR was adapted to determine listeners' speech intelligibility performance at 80% correct, the talker familiarity benefit was only present when the competing signal was a 1-talker masker. Specifically, performance on the Familiar Talker condition improved with higher LSPAN scores, as well as with higher Flanker scores. Talker familiarity and cognitive function did not impact performance on the SSN conditions. These findings suggest that when listening to speech in the presence of a competing talker, more cognitive resources are required for speech intelligibility compared to when listening in the presence of a speech shaped noise masker. On the 1-back recall test, the ONH listeners exhibited a significant talker familiarity benefit. Performance was higher in the Familiar + SSN condition with increasing LSPAN than the Unfamiliar + SSN condition. This finding is novel and provides evidence that ONH listeners can benefit from listening to a familiar talker in the presence of background noise.

This study also examined whether the effect of talker familiarity on the 1-back test was influenced by cognitive function. It was hypothesized that older adults would perform more poorly on the working memory test than younger adults, but would show a larger benefit when the target talker was a familiar voice. The findings confirmed that performance on the 1-back task varied with listener age group, SNR, and LSPAN score. For the SSN conditions, individuals with higher LSPAN scores were better able to take advantage of the more favorable SNRs, and thus had higher 1-back scores than individuals with lower LSPAN scores. However, when listening in the competing talker conditions, performance of the individuals with higher LSPAN scores decreased with increasing SNR. This may be due to the trend that primarily individuals with higher LSPAN scores were able to achieve very low SRTs. Taken together, these findings suggest that listening to a familiar voice can improve speech understanding in noise when the background masker is composed of speech, and that ONH listeners can take advantage of the talker familiarity benefit to improve performance on a test that increased demands on working memory capacity.

General Discussion

The long-term goal of this research was to understand the cognitive mechanisms responsible for a FV benefit in real-world environments, and to develop means to exploit the FV benefit to increase saliency of the attended speech for older who have difficulty understanding speech in noise. Speech understanding in a complex environment requires the listener to extract the target speech from the background noise by selectively attending to the target and filtering out the other competing sounds (Cherry, 1953). Speech understanding ability in noise is dependent on three main factors: 1) the environment in which an individual is listening; 2) individual subject factors of the listener; and 3) characteristics of the talker(s). The approach for this research was to measure speech understanding for familiar partners (spouses) in a cocktail-party scene. While adults have shown a speech intelligibility benefit when listening to a familiar voice in a controlled laboratory setting, it is unknown whether this effect is observed in a real-world environment. Speech intelligibility of familiar and unfamiliar talkers was examined in a real-world environment (Study 1). Speech segregation of a familiar voice was evaluated when the familiar voice served as either the target or competing talker in a laboratory environment, where the spatial separation of the target talker and maskers varied (Study 2). Lastly, the allocation of cognitive resources when listening to familiar speech was measured through a recall test with an increased memory load (Study 1 and 3).

Study 1 examined the effect of talker familiarity on speech understanding and working memory tasks in a real-world environment. Speech intelligibility was measured for pairs of familiar partners (spouses) in a local restaurant using a self-administered speech perception task. On a given trial, the target talker was a familiar voice for one of the participants and was as an unfamiliar voice for the other participants. Results suggested that both younger and older adults experienced a talker familiarity benefit in a real-world environment, and that this benefit was highly correlated with working memory capacity (i.e., on the LSPAN test), such that older adults with higher working memory scores exhibited a greater familiarity benefit than older adults with lower working memory scores. The listeners also completed a working memory task that required each participant to recall a particular word from each of the previous four sentences. The results showed that talker familiarity did not influence recall on the memory task. Thus, while talker familiarity was a salient cue for improved speech understanding in noise, it did not result in a reduction in cognitive resource consumption on the working memory task employed.

Study 2 examined the effect of talker familiarity on speech segregation for younger and older adults in a simulated cocktail-party environment. Stimuli were recorded from each participant's spouse (familiar voice), and were presented monotically or dichotically under headphones to simulate a cocktail-party environment in a laboratory setting. Speech understanding performance was measured for conditions where the FV was the target talker, and conditions where the FV is a masker. When speech was presented monotically, talker familiarity benefit was influenced by whether the listener was able to correctly identify whether the target talker was familiar or unfamiliar. When younger and older listeners were able to correctly identify the talker, they exhibited a talker familiarity benefit when the target talker was familiar. However, when the listener incorrectly identified the familiarity of the talker, the listeners exhibited a significant decline in performance when the familiar voice was a masker than when the voice was the target talker. This decline was significantly greater for the YNH than for the ONH listeners. Speech understanding performance was significantly poorer when the stimuli were presented dichotically, although both groups demonstrated a talker familiarity benefit. Furthermore, ONH listeners had more across-ear errors in the dichotic conditions than the YNH listeners, suggesting that they were less able to inhibit irrelevant stimuli than the YNH listeners.

Study 3 examined the effect of talker familiarity on a measure of auditory working memory in a laboratory environment. Similar to Study 2, stimuli were recorded from couples and were presented under headphones. Listeners completed an immediate recall task that adapted the SNR to obtain speech intelligibility performance at 80% correct. These SNRs were then used to test listeners on a task of auditory recall that required them to hold a spoken sentence in memory and then recall it after hearing another sentence (i.e., 1-back task). Consistent with results in Studies 1 and 2, the results for the immediate recall task showed a significant talker familiarity benefit when the competing signal was speech, but not when the competing signal was speech-shaped noise. However, in the competing speech conditions, performance on the Familiar Talker condition improved for individuals with higher cognitive scores

(LSPAN and Flanker), which suggests that when listening to speech in the presence of a competing talker, more cognitive resources are required for speech intelligibility compared to when listening in the presence of a SSN masker.

The results on the 1-back test in Study 3 suggest that individuals with higher working memory capacity and individuals with higher inhibitory control generally perform better on a speech recognition task with a relatively high memory load (i.e., the 1-back task) than those with lower cognitive ability. Performance on the 1-back task improved for ONH listeners with higher LSPAN scores on the Familiar Talker + SSN condition but not in the Unfamiliar Talker + SSN condition. This finding is novel and suggests that ONH listeners with higher working memory capacity can benefit from listening to a familiar voice in a situation that requires increased use of cognitive resources. However, the familiarity benefit measured in the SSN conditions was not present when the competing signal was speech. The presence of an informational masker may have increased cognitive demand such that listeners were unable to utilize their limited remaining cognitive resources to access a familiarity benefit on a memory recall task.

Real-world Environments

Target talker familiarity improved speech understanding in real-world environments. Study 1 demonstrated that when listening in an actual real-world environment, individuals were able to take advantage of the familiarity cue. This finding is novel, and suggests that both younger and older listeners can benefit from listening to a familiar voice beyond the effects seen in laboratory measures (Domingo, Holmes, & Johnsrude, 2019; Holmes et al., 2018; Johnsrude et al., 2013). When listeners were tested under headphones in Study 2, results showed that adults achieved higher speech understanding scores when the target talker was familiar in a simulated cocktail-party type scene (dichotic condition). These findings support previous literature that demonstrated that talker familiarity is an important cue for improving speech understanding in noise (Domingo, Holmes, & Johnsrude, 2019; Holmes et al., 2018; Johnsrude et al., 2013). Furthermore, when speech intelligibility was fixed at a high performance level (Study 3), individuals with higher LSPAN or Flanker scores exhibited a talker familiarity benefit when the competing signal was a single talker. When the competing signal was speech-shaped noise, however, there was not a significant effect of talker familiarity, regardless of listener age or cognitive ability. This masker-dependent familiarity benefit may be due to the fact that listening in the presence of a competing talker is significantly more challenging than when listening in the presence of fluctuating noise (Brungart, 2001b; Freyman et al., 2001). Thus, presence of a familiar target talker may be most beneficial when the competing signal provides some amount of informational masking on speech understanding task.

The presence of a familiar masker can potentially impair speech understanding performance. In Studies 2 and 3, listeners were presented with target stimuli that were mixed with one (Study 2 and 3) or two (Study 2) competing talkers. When the competing voice was familiar, listeners' ability to segregate target from masker speech varied based on the complexity of the auditory scene (number of competing talkers) and talker identification accuracy. When speech was presented monotically, the familiar masker did not negatively impact speech understanding if listeners were able to correctly identify the target talker. However, when the identification was incorrect, speech understanding performance was significantly poorer when the familiar voice was the masker. Performance in the Both Unfamiliar condition did not vary when the target talker identification was correct vs. incorrect. This suggests that with a familiar masker, the listener selected the incorrect speech stream as the target and was biased to follow that familiar voice. Thus, there was no benefit of a familiar voice as a masker, and furthermore, the familiar masker was detrimental to speech understanding.

When speech was presented dichotically, performance did not decrease with the addition of a familiar voice as the masker in either the Masker-Target Ear or the Masker – Opposite conditions, compared to when all competing voices were unfamiliar. This suggests that the decrement in performance for the dichotic conditions was not further impaired when one of the competing talkers was a familiar voice. In Study 3, when speech understanding performance was adapted to be high (80% correct intelligibility), the familiar masker did not reduce speech understanding performance in relation to the Both Unfamiliar condition. However, in the 1-back task, the ONH listeners exhibited an improvement in performance with increasing SNR in the Familiar Masker condition. One possible reason for the ONH listeners' improved performance on the 1-back test with a familiar masker may be due to the range of SNRs in which they were tested. The ONH listeners required SNRs that were generally above 0 dB when listening in competing speech, whereas the YNH listeners' SNRs were more dispersed and generally lower (less favorable). The findings across the three studies suggest that a familiar masker is only a detriment to speech understanding when there is an explicit confusion about the target talker identity, in which case the listener is more likely to attend to the salient familiar voice.

Aging

The benefits of talker familiarity on speech understanding were not specific to a particular age group. Both younger and older adults exhibited a familiarity benefit when listening in a real-world environment (Study 1) and when the complex environment was presented under headphones (Study 2 and Study 3). When speech intelligibility was fixed at 80% correct, both YNH and ONH listeners showed better SRT_{80} scores in the competing speech conditions when the target talker was familiar vs. unfamiliar (Study 3). Irrespective of talker familiarity, the ONH listeners showed a greater decline in performance when listening in the dichotic vs. monotic conditions than the YNH listeners. This is in agreement with studies that have shown that older adults are less able than young adults to ignore competing speech (Dubno et al., 1984; Helfer & Freyman, 2008).

In Study 2, the ONH listeners demonstrated both a benefit of a familiar target talker (monotic and dichotic configurations), and also a benefit of a familiar masker in the monotic configuration. Grouped across SNR (0 and -5 dB SNR), speech understanding for the ONH listeners in the Familiar Masker condition was significantly better than for the Both Unfamiliar condition when target talker identification was correct. This is in agreement with Johnsrude et al. (2013), who showed a familiar masker benefit for adults aged 40 – 80 years old in adverse SNRs (-6, -3, and 0 dB SNR).

Younger adults appeared to ignore the familiar voice when it served as the masker. One possible reason for this finding is that the YNH listeners were able to attend well to the call sign cue. The older adults, however, may have been attending to both streams and in instances where they correctly identified the target talker, they were better able to segregate and report what was said by the target talker even when the familiar voice was the masker. This is supported by the error analysis conducted in Study 2, which showed that across all dichotic conditions, the ONH listeners had a greater proportion of errors than YNH listeners for masker words reported from the non-target ear. The across-ear intrusion errors may be due to a reduction in inhibition by older adults (Hasher & Zacks, 1988; Presacco et al., 2016a). This finding is consistent with those reported in previous studies in which ONH listeners had greater difficulty segregating competing streams of speech (Helfer & Jesse, 2015; Jesse & Helfer, 2019),

Working Memory Capacity and Inhibitory Control

Studies in the area of aging and speech understanding have repeatedly highlighted the importance of cognition, especially during difficult listening conditions. Auditory working memory and inhibitory control significantly contribute to listeners' ability to understand speech in the presence of background noise (Conway et al., 2001; Janse, 2012; Ronnberg et al., 2010; Souza et al., 2015). The Ease of Language Understanding (ELU) model suggests that working memory is dependent on top-down processes such as language processing when attempting to understand speech in a noisy environment (Ronnberg et al., 2019; Ronnberg et al., 2008). The impact of increased working memory demand during a speech understanding task was evaluated in Studies 1 and 3. Speech understanding in the real-world environment (Study 1) was predicted by LSPAN scores. In trials with higher noise levels, speech understanding performance was better for individuals with higher vs. lower LSPAN scores, which is consistent with the premises of the ELU model (Ronnberg et al., 2019; Ronnberg et al., 2013). Older adults with lower LSPAN scores were less able to take advantage of the familiar target talker benefit in this environment than older adults with higher LSPAN scores. Similarly, the results from Study 3 indicated that higher cognitive performance as measured on the LSPAN and Flanker tests resulted in lower (better) SRT_{80} thresholds for the competing talker conditions. Neither measure was predictive of performance in the speech-shaped noise conditions.

The Flanker and LSPAN scores were also predictive of performance on recall tests with increased memory load. In Study 1, individuals with higher LSPAN scores had better n-back recall performance than those with lower LSPAN scores, particularly for the 1-back condition. Similarly, performance on the 1-back test in Study 3 correlated with cognitive measures, which included the Flanker and the LSPAN cognitive measures. Participants with higher Flanker scores demonstrated better 1-back recall in the Unfamiliar Talker + SSN and Both Unfamiliar conditions. Because inhibitory control was not identified as a significant factor in the Familiar Talker conditions, it appears that inhibitory control was not predictive of talker familiarity benefit. Conversely, ONH listeners with higher LSPAN score demonstrated better 1-back scores in the Familiar Talker vs. Unfamiliar Talker + SSN condition. This evidence suggests that the presence of a familiar target talker can reduce consumption of cognitive resources. However, this effect was not seen in Study 1. These findings are consistent with previous studies that have shown that speech understanding in noise is highly correlated with cognitive function, and in this specific case, auditory working memory (Gordon-Salant & Cole, 2016; Pichora-Fuller et al., 1995; Ronnberg et al., 2019; Schurman et al., 2014).

The difference in familiarity benefit seen in Studies 1 and 3 may be due to differences in methodology. In Study 1, participants were instructed to recall the CRM call sign (Baron, Ringo, etc.) from the previous four trials. The methodology used in Study 3 was different in that participants were instructed to recall the sentence (four words) that was presented in the previous trial (e.g., 1-back). When comparing the 1-back scores across the two studies, both younger and older adults exhibited better recall ability in Study 1 than Study 3. Thus, the reduced memory load of remembering one vs. four words may have resulted in a decrease in reliance on the familiarity cue in Study 1.

Conclusions

Across the three studies, talker familiarity was a salient cue that both younger and older adults utilized when listening in complex environment, such as a restaurant or crowded gathering, or in simulations of these complex environments. Understanding the role that familiar voices may have on the allocation of cognitive resources will result in improved aural rehabilitation strategies and will ultimately facilitate improvements in partner communication in complex real-world environments (Preminger, 2003; Tye-Murray et al., 2016). For example, utilizing familiar voices (i.e., spouses), in aural rehabilitation training programs could help to alleviate some of the cognitive demands involved during auditory training. In turn, these cognitive resources could then be allocated towards the specific training task, such as learning to differentiate an acoustic cue or contrast, with the ultimate goal of improving overall speech understanding. Thus, if audiologists can reduce some of the barriers to listening in a complex environment by training patients on certain tasks while using a familiar voice as the stimulus, the result could be quite positive.

The long-term goal of this research is to understand the cognitive mechanisms responsible for a FV benefit in real-world environments, and to develop means to exploit the FV benefit to increase saliency of the attended speech for older adults with hearing loss. While it was demonstrated that listening to a familiar voice in a complex environment can result in improved speech understanding, and potentially improved recall ability in a high-memory load task, the mechanisms contributing to these benefits are still unknown.

Different methodologies for measuring the cognitive mechanisms of talker familiarity could include use of physiologic measures such as pupillometry and electroencephalography (EEG), and should be used to further evaluate how the brain encodes familiar speech. Overall, the results suggest that younger and older adults receive a significant benefit to speech understanding when attending to a familiar voice as the target talker, and these findings can contribute to the growing body of research into the development of attention- or cognitively driven hearing aids, which rely on a listener's focus of attention to selectively amplify a target signal.

Bibliography

Akeroyd, M.A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol, 47*(s2), S53-71.

ANSI. (1969). *Calculation of the articlation index*. (ANSI S3.5-1969). New York: American National Standards Institute.

ANSI. (2017). *American national standard for methods of calculation of the speech intelligibility index*. (ANSI S3.5-1997 (R2017)). New York: American National Standards Institute.

ANSI. (2018). *American national standard for specification for audiometers*. (ANSI/ASA S3.6-2018). New York: American National Standards Institute.

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M.K. (2009). The emergence of cognitive hearing science. *Scand J Psychol, 50*(5), 371-384.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends Cogn Sci, 4*(11), 417-423.

Baddeley, A., & Hitch, G.J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89).

Barker, B.A., & Elliott, E.M. (2019). The role of talker familiarity in auditory distraction. *Exp Psychol, 66*(1), 1-11.

Bernstein, J.G., & Grant, K.W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am, 125*(5), 3358-3372.

Best, V., Ahlstrom, J.B., Mason, C.R., Roverud, E., Perrachione, T.K., Kidd, G., Jr., & Dubno, J.R. (2018). Talker identification: Effects of masking, hearing loss, and age. *J Acoust Soc Am, 143*(2), 1085.

Best, V., Keidser, G., Buchholz, J.M., & Freeston, K. (2015). An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment. *Int J Audiol, 54*(10), 682-690.

Bolia, R.S., Nelson, W.T., Ericson, M.A., & Simpson, B.D. (2000). A speech corpus for multitalker communications research. *J Acoust Soc Am, 107*(2), 1065-1066.

Bregman, A.S. (1990). *Auditory scene analysis : The perceptual organization of sound*. Cambridge, Mass.: MIT Press.

Bronkhorst, A.W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica, 86*(1), 117-128.

Brungart, D.S. (2001a). Evaluation of speech intelligibility with the coordinate response measure. *J Acoust Soc Am, 109*(5 Pt 1), 2276-2279.

Brungart, D.S. (2001b). Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am, 109*(3), 1101-1109.

Brungart, D.S., & Simpson, B.D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *J Acoust Soc Am, 112*(6), 2985-2995.

Brungart, D.S., Simpson, B.D., Ericson, M.A., & Scott, K.R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am, 110*(5 Pt 1), 2527-2538.

Carlozzi, N.E., Beaumont, J.L., Tulsky, D.S., & Gershon, R.C. (2015). The nih toolbox pattern comparison processing speed test: Normative data. *Arch Clin Neuropsychol, 30*(5), 359-368.

CHABA. (1988). Speech understanding and aging. Working group on speech understanding and aging. Committee on hearing, bioacoustics, and biomechanics, commission on behavioral and social sciences and education, national research council. *J Acoust Soc Am, 83*(3), 859-895.

Cherry, E.C. (1953). Some experiments on the recognition of speech with one and with two ears. *J Acoust Soc Am, 25*(5), 975-979.

Clopper, C.G., & Bradlow, A.R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Lang Speech, 51*(Pt 3), 175-198.

Cohen, J.I., & Gordon-Salant, S. (2017). The effect of visual distraction on auditory-visual speech perception by younger and older listeners. *J Acoust Soc Am, 141*(5), EL470.

Compton-Conley, C.L., Neuman, A.C., Killion, M.C., & Levitt, H. (2004). Performance of directional microphones for hearing aids: Real-world versus simulation. *J Am Acad Audiol, 15*(6), 440-455.

Conway, A.R., Cowan, N., & Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychon Bull Rev, 8*(2), 331-335.

Conway, A.R., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev, 12*(5), 769-786.

Cruickshanks, K.J., Wiley, T.L., Tweed, T.S., Klein, B.E., Klein, R., Mares-Perlman, J.A., & Nondahl, D.M. (1998). Prevalence of hearing loss in older adults in beaver dam, wisconsin. The epidemiology of hearing loss study. *Am J Epidemiol, 148*(9), 879-886.

Daneman, M., & Carpenter, P.A. (1980). Individual-differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Darwin, C.J., Brungart, D.S., & Simpson, B.D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J Acoust Soc Am, 114*(5), 2913-2922.

Dirks, D.D., Morgan, D.E., & Dubno, J.R. (1982). A procedure for quantifying the effects of noise on speech recognition. *J Speech Hear Disord, 47*(2), 114-123.

Domingo, Y., Holmes, E., & Johnsrude, I.S. (2019). The benefit to speech intelligibility of hearing a familiar voice. *J Exp Psychol Appl*.

Domingo, Y., Holmes, E., Macpherson, E., & Johnsrude, I.S. (2019). Using spatial release from masking to estimate the magnitude of the familiar-voice intelligibility benefit. *J Acoust Soc Am, 146*(5), 3487-3494.

Dubno, J.R., Dirks, D.D., & Morgan, D.E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am, 76*(1), 87-96.

Dubno, J.R., Horwitz, A.R., & Ahlstrom, J.B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *J Acoust Soc Am, 111*(6), 2897-2907.

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143-149.

Fontaine, M., Love, S.A., & Latinus, M. (2017). Familiarity and voice representation: From acoustic-based representation to voice averages. *Front Psychol, 8*, 1180.

Freyman, R.L., Balakrishnan, U., & Helfer, K.S. (2001). Spatial release from informational masking in speech recognition. *J Acoust Soc Am, 109*(5 Pt 1), 2112-2122.

Freyman, R.L., Balakrishnan, U., & Helfer, K.S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am, 115*(5 Pt 1), 2246-2256.

Freyman, R.L., Helfer, K.S., McCall, D.D., & Clifton, R.K. (1999). The role of perceived spatial separation in the unmasking of speech. *J Acoust Soc Am, 106*(6), 3578-3588.

Fullgrabe, C., Moore, B.C., & Stone, M.A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Front Aging Neurosci, 6*, 347.

Fullgrabe, C., & Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Front Psychol, 7*(1268), 1268.

Gershon, R.C., Wagster, M.V., Hendrie, H.C., Fox, N.A., Cook, K.F., & Nowinski, C.J. (2013). Nih toolbox for assessment of neurological and behavioral function. *Neurology, 80*(11 Suppl 3), S2-6.

Gordon-Salant, S., & Cole, S.S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear Hear, 37*(5), 593-602.

Gordon-Salant, S., & Fitzgibbons, P.J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Hear Res, 36*(6), 1276-1285.

Gordon-Salant, S., & Fitzgibbons, P.J. (1999). Profile of auditory temporal processing in older listeners. *J Speech Lang Hear Res, 42*(2), 300-311.

Gordon-Salant, S., Fitzgibbons, P.J., & Friedman, S.A. (2007). Recognition of time-compressed and natural speech with selective temporal enhancements by young and elderly listeners. *J Speech Lang Hear Res, 50*(5), 1181-1193.

Gordon-Salant, S., Yeni-Komshian, G., & Fitzgibbons, P. (2008). The role of temporal cues in word identification by younger and older adults: Effects of sentence context. *J Acoust Soc Am, 124*(5), 3249-3260.

Gordon-Salant, S., Yeni-Komshian, G.H., & Fitzgibbons, P.J. (2010). Recognition of accented english in quiet and noise by younger and older listeners. *J Acoust Soc Am, 128*(5), 3152-3160.

Gordon-Salant, S., Yeni-Komshian, G.H., Fitzgibbons, P.J., Cohen, J.I., & Waldroup, C. (2013). Recognition of accented and unaccented speech in different maskers by younger and older listeners. *J Acoust Soc Am, 134*(1), 618-627.

Gosselin, P.A., & Gagne, J.P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *Int J Audiol, 50*(11), 786-792.

Gosselin, P.A., & Gagné, J.P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research, 54*(3), 944-958.

Grant, K.W., & Seitz, P.F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am, 108*(3 Pt 1), 1197-1208.

Hasher, L., & Zacks, R.T. (1988). Working memory, comprehension, and aging: A review and a new view. In *Psychology of learning and motivation* (pp. 193-225): Elsevier Science & Technology.

Helfer, K.S., & Freyman, R.L. (2005). The role of visual speech cues in reducing energetic and informational masking. *J Acoust Soc Am, 117*(2), 842-849.

Helfer, K.S., & Freyman, R.L. (2008). Aging and speech-on-speech masking. *Ear Hear, 29*(1), 87-98.

Helfer, K.S., & Jesse, A. (2015). Lexical influences on competing speech perception in younger, middle-aged, and older adults. *J Acoust Soc Am, 138*(1), 363-376.

Holmes, E. (2018). Speech recording videos (Version v1.0.0) [Computer code]: Zenodo.

Holmes, E., Domingo, Y., & Johnsrude, I.S. (2018). Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychol Sci, 29*(10), 1575-1583.

House, A.S., Williams, C., Hecker, M.H., & Kryter, K.D. (1963). Psychoacoustic speech tests: A modified rhyme test. Techn docum rep esd-tdr-63-403. *Tech Doc Rep U S Air Force Syst Command Electron Syst Div, 86*(11), 1-44.

Hox, J.J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*: Routledge.

Humes, L.E., & Dubno, J.R. (2010). Factors affecting speech understanding in older adults. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The aging auditory system* (pp. 211-257). New York, NY: Springer New York.

Humes, L.E., Dubno, J.R., Gordon-Salant, S., Lister, J.J., Cacace, A.T., Cruickshanks, K.J., . . . Wingfield, A. (2012). Central presbycusis: A review and evaluation of the evidence. *J Am Acad Audiol, 23*(8), 635-666.

IEEE. (1969). Ieee recommended practice for speech quality measurements. *Ieee Transactions on Audio and Electroacoustics, 17*(3), 225-246.

Ingvalson, E.M., & Stoimenoff, T.L. (2015). *Greater benefit for familiar talkers under cognitive load.* Paper presented at the Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK: University of Glasgow.

Iyer, N., Brungart, D.S., & Simpson, B.D. (2010). Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task. *J Acoust Soc Am, 128*(5), 2998-2910.

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 19*(6), 741-758.

Jesse, A., & Helfer, K.S. (2019). Lexical influences on errors in masked speech perception in younger, middle-aged, and older adults. *J Speech Lang Hear Res, 62*(4S), 1152-1166.

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

Kidd, G., Jr., Arbogast, T.L., Mason, C.R., & Gallun, F.J. (2005). The advantage of knowing where to listen. *J Acoust Soc Am, 118*(6), 3804-3815.

Kidd, G., Jr., Mason, C.R., Rohtla, T.L., & Deliwala, P.S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J Acoust Soc Am, 104*(1), 422-431.

Kidd, G., Mason, C.R., Richards, V.M., Gallun, F.J., & Durlach, N.I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143-189). Boston, MA: Springer US.

Kidd, G.R., & Humes, L.E. (2015). Keeping track of who said what: Performance on a modified auditory n-back task with young and older adults. *Front Psychol, 6*, 987.

Krause, J.C., & Braida, L.D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *J Acoust Soc Am, 112*(5 Pt 1), 2165-2172.

Kreitewolf, J., Mathias, S.R., & von Kriegstein, K. (2017). Implicit talker training improves comprehension of auditory speech in noise. *Front Psychol, 8*, 1584.

Kuchinsky, S.E., Ahlstrom, J.B., Vaden, K.I., Jr., Cute, S.L., Humes, L.E., Dubno, J.R., & Eckert, M.A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology, 50*(1), 23-34.

Lee, J.H., & Humes, L.E. (2012). Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background. *J Acoust Soc Am, 132*(3), 1700-1717.

Lipnicki, D.M., Crawford, J.D., Dutta, R., Thalamuthu, A., Kochan, N.A., Andrews, G., . . . Cohort Studies of Memory in an International, C. (2017). Age-related cognitive decline and associations with sex, education and apolipoprotein e genotype across ethnocultural groups and geographic regions: A collaborative cohort study. *Plos med, 14*(3), e1002261.

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille et du Larynx, 37*, 101-119.

Maibauer, A.M., Markis, T.A., Newell, J., & McLennan, C.T. (2014). Famous talker effects in spoken word recognition. *Atten Percept Psychophys, 76*(1), 11-18.

Marrone, N., Mason, C.R., & Kidd, G., Jr. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *J Acoust Soc Am, 124*(5), 3064-3075.

Mason, H.M. (1946). Understandability of speech in noise as affected by region of origin of speaker and listener. *Communications Monographs, 13*(2), 54-58.

Moray, N. (1959). Attention in dichotic-listening - affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56-60.

Nasreddine, Z.S., Phillips, N.A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The montreal cognitive assessment, moca: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc, 53*(4), 695-699.

Newman, R.S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics, 35*(1), 85-103.

Nygaard, L.C., & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Percept Psychophys, 60*(3), 355-376.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *J Exp Psychol Learn Mem Cogn, 19*(2), 309-328.

Park, D.C., Lautenschlager, G., Hedden, T., Davidson, N.S., Smith, A.D., & Smith, P.K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychol Aging, 17*(2), 299-320.

Pearson, J.D., Morrell, C.H., Gordon-Salant, S., Brant, L.J., Metter, E.J., Klein, L.L., & Fozard, J.L. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *J Acoust Soc Am, 97*(2), 1196-1205.

Pearsons, K., Bennett, R., & Fidell, S. (1977). Speech levels in various noise environments. Report no. Epa-600/1-77-025. *U.S. Enivronmental Protection Agency*.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1985). Speaking clearly for the hard of hearing i: Intelligibility differences between clear and conversational speech. *J Speech Hear Res, 28*(1), 96-103.

Pichora-Fuller, M.K., Schneider, B.A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97*(1), 593-608.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *J Speech Hear Res, 29*(2), 146-154.

Plomp, R., & Mimpen, A.M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am, 66*(5), 1333-1342.

Plomp, R., & Mimpen, A.M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences. *Acustica, 48*, 325-328.

Preminger, J.E. (2003). Should significant others be encouraged to join adult group audiologic rehabilitation classes? *J Am Acad Audiol, 14*(10), 545-555.

Presacco, A., Simon, J.Z., & Anderson, S. (2016a). Effect of informational content of noise on speech representation in the aging midbrain and cortex. *J Neurophysiol, 116*(5), 2356-2367.

Presacco, A., Simon, J.Z., & Anderson, S. (2016b). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J Neurophysiol, 116*(5), 2346-2355.

Revit, L.J., Schulein, R.B., & Julstrom, S.D. (2002). Toward accurate assessment of real-world hearing aid benefit. *Hear Rev, 9*, 34-38.

Ronnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int J Audiol, 42*(s1), S68-S76.

Ronnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int J Audiol, 58*(5), 247-261.

Ronnberg, J., Lunner, T., Zekveld, A., Sorqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The ease of language understanding (elu) model: Theoretical, empirical, and clinical advances. *Front Syst Neurosci, 7*, 31.

Ronnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (elu). *Int J Audiol, 47 Suppl 2*, S99-105.

Ronnberg, J., Rudner, M., Lunner, T., & Zekveld, A.A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise & Health, 12*(49), 263-269.

Rossi-Katz, J., & Arehart, K.H. (2009). Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *J Speech Lang Hear Res, 52*(2), 435-453.

Rudner, M., Lunner, T., Behrens, T., Thoren, E.S., & Ronnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology, 23*(8), 577-589.

Rudner, M., & Ronnberg, J. (2008). The role of the episodic buffer in working memory for language processing. *Cogn Process, 9*(1), 19-28.

Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol Rev, 103*(3), 403-428.

Schurman, J., Brungart, D., & Gordon-Salant, S. (2014). Effects of masker type, sentence context, and listener age on speech recognition performance in 1-back listening tasks. *J Acoust Soc Am, 136*(6), 3337.

Schweinberger, S.R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *J Speech Lang Hear Res, 40*(2), 453-463.

Sheffert, S.M., Pisoni, D.B., Fellowes, J.M., & Remez, R.E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *J Exp Psychol Hum Percept Perform, 28*(6), 1447-1469.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. *Trends Cogn Sci, 12*(5), 182-186.

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol, 26*(2), 183-196.

Souza, P., Arehart, K., & Neher, T. (2015). Working memory and hearing aid processing: Literature findings, future directions, and clinical applications. *Front Psychol, 6*, 1894.

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *J Am Acad Audiol, 24*(8), 689-700.

Studebaker, G.A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res, 28*(3), 455-462.

Tulsky, D.S., Carlozzi, N., Chiaravalloti, N.D., Beaumont, J.L., Kisala, P.A., Mungas, D., . . . Gershon, R. (2014). Nih toolbox cognition battery (nihtb-cb): List sorting test to measure working memory. *J Int Neuropsychol Soc, 20*(6), 599-610.

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

Tun, P.A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol Aging, 17*(3), 453-467.

Tye-Murray, N., Spehar, B., Sommers, M., & Barcroft, J. (2016). Auditory training with frequent communication partners. *J Speech Lang Hear Res, 59*(4), 871-875.

van Rooij, J.C., Plomp, R., & Orlebeke, J.F. (1989). Auditive and cognitive factors in speech perception by elderly listeners. I: Development of test battery. *J Acoust Soc Am, 86*(4), 1294-1309.

Vaughan, N.E., Furukawa, I., Balasingam, N., Mortz, M., & Fausti, S.A. (2002). Time-expanded speech and speech recognition in older adults. *J Rehabil Res Dev, 39*(5), 559-566.

Voeten, C.C. (2020). Buildmer: Stepwise elimination and term reordering for mixed-effects regression (Version 1.6): R Foundation for Statistical Computing. Retrieved from https://CRAN.R-project.org/package=buildmer

Walden, B.E., Busacco, D.A., & Montgomery, A.A. (1993). Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. *J Speech Hear Res, 36*(2), 431-436.

Wingfield, A., McCoy, S.L., Peelle, J.E., Tun, P.A., & Cox, L.C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J Am Acad Audiol, 17*(7), 487-497.

Yonan, C.A., & Sommers, M.S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychol Aging, 15*(1), 88-99.

Zekveld, A.A., Rudner, M., Johnsrude, I.S., & Ronnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *J Acoust Soc Am, 134*(3), 2225-2234.