NORTHWESTERN UNIVERSITY

Linguistic factors in speech-in-speech perception

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By

Kristin J. Van Engen

EVANSTON, ILLINOIS

December 2010

ABSTRACT

Linguistic factors in speech-in-speech recognition

Kristin J. Van Engen

Speech perception rarely takes place in quiet environments, and for people with hearing impairment, difficulty understanding speech in noisy conditions is often the primary complaint. However, neither traditional clinical hearing tests nor a person's ability to understand speech in quiet can adequately predict the amount of trouble s/he will have in noise. Noise is also more detrimental to speech understanding for non-native and bilingual listeners than for monolingual listeners of a given language. This variability in performance within and across clinical and linguistic populations suggests that speech intelligibility in noise is modulated by both the auditory system and by experience-related cognitive factors such as a person's language background. Thus far, most research efforts have been directed toward auditory function; my dissertation aims to develop a fuller understanding of the particularly *linguistic* aspects of speech-in-noise perception.

Building on the insight that speech-in-noise intelligibility may be modulated by listener experience, this dissertation has two primary aims: 1) to expand our understanding of the relationship between language experience and speech intelligibility in speech noise and 2) to identify parameters of auditory training that can facilitate speech-in-speech intelligibility. The first aim was addressed by assessing English sentence intelligibility in English and Mandarin babble by monolingual English speakers and non-native speakers of English whose native language is Mandarin. Results showed that both groups experienced greater difficulty in English

versus Mandarin babble, but that Mandarin listeners received a smaller release from masking in Mandarin babble relative to English babble. These findings indicate that both the similarity between the target and the noise and the language experience of the listener contribute to the masking imposed by speech noise. The second aim was addressed with a training study, in which listeners received speech-in-noise training in speech-shaped noise, Mandarin babble, or English babble. Post-test performance showed that listeners were able to take advantage of target talker familiarity, and, after babble training, that they improved most in coping with the babble language in which they were trained. Speech-in-noise training, therefore, can enhance processes associated with both "tuning in" to speech targets and "tuning out" speech noise.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor, Ann Bradlow, for teaching and mentoring me with such wisdom, wit, and grace. She has been instrumental both to this project and more broadly to my development as a scholar. Over the past several years, Ann has inspired, challenged, and supported me—always at the right time and always in the right way—giving generously of her time, energy, and advice. Working with Ann, in short, has been a delight.

Beyond Ann, I would also like to thank my other committee members, Sumit Dhar and Patrick Wong, for their helpful comments throughout the dissertation project and for challenging me to think about my work from the perspectives of multiple disciplines.

I am also grateful to the Linguistics Department at Northwestern, which has provided me with training and support in and out of the classroom, including opportunities for teaching, research, and administration that have laid a solid foundation for continuing this career.

In addition, I gratefully acknowledge the institutional support I have received during my graduate years. In particular, I would like to thank Northwestern University and the NIH-NIDCD for generous fellowships that have allowed me to do the research presented in this dissertation.

My work and life as a graduate student has been greatly enriched by supportive colleagues and a wonderful community of collaborators. I am grateful to each of the members of the Speech Communication Research group, and especially to Rachel Baker, Melissa Baese-Berk, Susanne

Brouwer, and Lauren Calandruccio. Chun Chan deserves many, many thanks for his technical support and patient teaching. I am also thankful to Matt Goldrick for statistical advice and to Page Piccinini and Kelsey Mok for assistance with data collection.

Finally, I am both thankful for and grateful to my family, friends, and church community. Their support and friendship over these years has been invaluable to me. It would be impossible to name everyone here, but I would especially like to thank my parents, Barb and Bruce, and my sisters, Mara and Alyssa. My son Simon has blessed this past year in countless ways. Lastly, I most want to thank Abram, whose encouragement and love have seen me through the challenges and joys of graduate school and strengthen me day by day.

TABLE OF CONTENTS

Abstract 3

Acknowledgements 5

Table of Contents 7

List of Figures 10

List of Tables 11

Chapter 1: Introduction 12

Significance of this research 18
Overview of the dissertation 18
Organization of the dissertation 22

Chapter 2: Two follow-up studies to Van Engen and Bradlow (2007) 23

Experiment 1: Real-word vs. Non-word babble 23
Introduction 23
Methods 24
Participants 24
Materials 24
Target sentences 24
Noise 25
Procedure 26
Data analysis 26
Results 27
Discussion 27
Experiment 2: An investigation of SNR manipulations methods 29
Introduction 29
Methods 30
Participants 30
Materials 30
Procedure 31
Data analysis 32
Results 32
Discussion 33

Chapter 3: Similarity and familiarity: second language sentence recognition in first- and second-language multi-talker babble 35

Abstract....................35
Introduction....................36
Methods....................45
Participants....................45
Monolingual English listeners....................45
L2 English listeners....................46
Materials....................47
Multi-talker babble....................47
Target sentences....................48
Targets + babble....................49
Procedure....................50
Data analysis....................53
Results....................53
HINT results....................53
Speech-in-babble results....................55
Discussion....................58
HINT thresholds....................58
Sentence intelligibility in 2-talker babble....................59

Chapter 4: Speech perception in speech noise by early bilingual listeners....................65

Introduction....................65
Methods....................67
Participants....................67
Materials....................68
Procedure....................68
Results....................68
HINT results....................68
Speech-in-babble results....................69
Discussion....................71
HINT thresholds....................71
Sentence intelligibility in 2-talker babble....................71

Chapter 5: Speech-in-speech recognition: a training study....................73

Abstract....................73
Introduction....................73
Experiment: multiple-talker training....................85
Methods....................85
Participants....................85
Materials....................86
Pre-testing....................86
Target sentences....................87
Noise....................88

Mixing target sentences and noise 88
Procedure 89
Pre-testing 89
Training 90
Testing 91
Data analysis 92
Results 93
Letter-number sequencing 93
HINT 93
Training 93
Test 97
Talker 98
Masker 99
Training condition 100
Working memory and SNR 101
Discussion 102
Performance during training 102
Performance in the speech-in-speech post-test 107
Talker 107
Masker 107
Training condition 108
Working memory 110
SNR 111
Future directions 113
Conclusion 115

Chapter 6: Conclusion 116
HINT normalization 117
Linguistic masking effects in speech-in-speech recognition 121
Auditory training 127
Individual differences 128
Conclusion 129

References 130

Appendices 140

LIST OF FIGURES

Figure 2.1: Keyword intelligibility scores (in RAU) in real-word and non-word noise at SNRs of 0 dB and -5 dB 27

Figure 2.2: Keyword identification scores (in RAU) from Van Engen and Bradlow (2007), Condition 4 (left panel) and in the current experiment (right panel). 33

Figure 3.1: HINT threshold scores (the SNR at which a listener can identify whole sentences 50 percent of trials) for the native English listeners and the native Mandarin listeners. The center line on each boxplot denotes the median score, the edges of the box denote the 25th and 75th percentiles, and the whiskers extend to data points that lie within 1.5 times the interquartile range. Points outside this range appear as outliers. 54

Figure 3.2: Mean intelligibility scores expressed as percentage of correct keyword identifications for native English listeners (left) and native Mandarin (L2 English) listeners (right). Error bars represent standard error....... 55

Figure 4.1: Boxplots showing HINT SNR-50 thresholds for monolingual English listeners and bilingual speakers of English and Spanish. In each boxplot, the black line represents the median value, the box extends to the 25th and 75th percentile, and whiskers extend to values within 2.5 times the interquartile range....... 69

Figure 4.2: Percentage of keywords identified by Monolingual English listeners and Spanish-English bilingual listeners in English and Mandarin 2-talker babble at four SNRs. Error bars represent standard error....... 70

Figure 5.1: Keywords identified by each training group on Days 1 and 2 of training. Day 1 data is presented in grey boxplots; Day 2 in white boxplots. In each plot, the black line represents the median value, the box extends to the 25th and 75th percentile, and the whiskers extend to data points within 2.5 times the interquartile range....... 94

Figure 5.2: Average percentage of keywords identified by listeners in each training condition in the post-test for target sentences spoken by the training talker and the new talker. Error bars indicate standard error....... 98

Figure 5.3: Average percentage of keywords identified by listeners in each training group in the post-test for target sentences embedded in Mandarin and English babble. Error bars indicate standard error. 100

LIST OF TABLES

Table 2.1: Differences in rms amplitude across SNR conditions using two methods for presenting different SNRs. Method 1 was used in Van Engen and Bradlow (2007); Method 2 was investigated in the current experiment. 31

Table 3.1: Native Mandarin participants: English learning and proficiency information 47

Table 3.2: Means and standard errors of English and Mandarin listeners' recognition scores in each noise condition (% keywords identified). 55

Table 3.3: Summary of logistic regression on probability of correct response including participant as random intercept (Overall intercept: 0.606; St.Dev. of participant intercepts: 0.529). 56

Table 3.4: Mean differences with 95% confidence intervals for keywords identified in Mandarin - English babble (in percentage correct). 57

Table 5.1: Coefficients of the mixed-effects model and associated Z-statistics for performance during speech-in-noise training....................95

Table 5.2: Coefficients of the mixed-effects models and associated Z-statistics for performance on Days 1 and 2 of training for the group trained in speech-shaped noise. 96

Table 5.3: Coefficients of the mixed-effects models and associated Z-statistics for performance on Days 1 and 2 of training for the group trained in Mandarin babble. 96

Table 5.4: Coefficients of the mixed-effects models and associated Z-statistics for performance on Days 1 and 2 of training for the group trained in English babble. 96

Table 5.5: Coefficients of the mixed-effects model and associated z-statistics for performance on the speech-in-speech post-test....................98

CHAPTER 1: INTRODUCTION

Research on human speech perception has traditionally been conducted using carefully controlled speech signals (e.g., phonemes, syllables, words, sentences) that are presented to listeners in quiet environments. Given the complexity both of speech signals themselves and the auditory and cognitive processes involved in perceiving and recognizing them, this approach has been a necessary control in the development of our understanding of how acoustic signals are mapped onto linguistic representations. During everyday speech perception, however, listeners encounter a tremendous amount of variability in speech signals, and rarely do they have the benefit of listening in quiet conditions.

Several recent studies have shown that humans process speech signals differently when they are embedded in noise as opposed to presented in quiet. First, studies of phoneme recognition have shown that the presence of background noise can "re-rank" acoustic cues to linguistic categories so that cues that are secondary in quiet conditions become primary cues in noise (Parikh and Loizou, 2005; Jiang et al., 2006). Further, Mattys et al. (2005) have shown that, in word segmentation, listeners assign different weights to various cues (i.e., lexical, segmental, prosodic) when the speech signal is fully available as opposed to degraded by noise. Finally, noise has been shown to have asymmetrical effects on the intelligibility of native- versus foreign-accented speech (Rogers et al., 2004). As pointed out by Mattys and Liss (2008), "any successful model of speech processing must ultimately be able to accommodate the full range and variety of speech distortion that listeners encounter, and it must be able to explain how and when distortion interferes with speech processing." Since the presence of noise in the communicative environment is one of the most common sources of speech distortion, it is crucial that we continue to develop an understanding of how speech is processed in noise.

The study of speech perception in noise is often traced to Cherry's (1953) paper, which introduced the "cocktail party problem" (see Bronkhorst, 2000 and Darwin, 2008 for reviews of speech-in-noise research), and sparked research on a wide range of relevant phenomena such as selective listening, masking, binaural processing, and many aspects of signal segregation (Bronkhorst, 2000). This dissertation, therefore, joins a diverse and growing body of research across multiple disciplines (e.g., speech and hearing science, linguistics, psychology, neuroscience) in investigating the perception of speech in noisy conditions. The focus of the research in this dissertation is on human speech recognition in the presence of interfering speech (2-talker babble). In particular, it aims to establish that the intelligibility of speech in speech noise is modulated by listeners' experience – both long-term language experience and short-term auditory training experience.

For most people, speech communication takes place with relatively little effort, even in noisy situations. This ability to cope well with noise is typically attributed to listeners' ability to take advantage of redundancies in the speech signal, higher-level linguistic contextual information (e.g. lexical, syntactic, semantic, and pragmatic cues), visual cues, and spatial cues. However, difficulty understanding speech in noise is the primary complaint of people with hearing loss, and neither the traditional audiogram nor a person's ability understand speech in quiet can adequately predict the amount of trouble the person will have understanding speech in noise (e.g., Plomp and Mimpen, 1979; Smoorenburg, 1992; Killion and Niquette, 2000). Furthermore, noise is more detrimental to speech understanding for non-native and bilingual listeners than for monolingual listeners of a given language (e.g., Mayo et al., 1997; von Hapsburg et al., 2004; Rogers et al., 2006). This variability in performance within and across clinical and linguistic populations suggests that the ability to understand speech in noise is modulated by both the functionality of the auditory system and by experience-related cognitive factors such as a person's language learning background. Thus far, most research efforts have been directed toward auditory function; the research in this dissertation aims to develop a fuller understanding of linguistic aspects of speech-in-noise perception.

The goal of the research begun in this dissertation is to understand the role of language (i.e., listeners' language experience and the linguistic content of target and noise signals) in modulating speech-in-noise intelligibility and to specify the contribution of linguistic factors to the relatively underspecified notion of "informational masking." The objective of the experiments in the dissertation itself is to understand how long-term language experience (native, non-native, and bilingual) affects speech-in-speech understanding and how short-term training can improve it. Since this scenario requires listeners to process, at some level, speech signals in both targets and noise, it is hypothesized that listener experience with the language(s) of speech targets and speech noise will modulate the intelligibility of target signals. Further, short-term speech-in-noise training experience may also be able to enhance listeners' ability to "tune in" to targets and "tune out" noise.

Although most listeners are very good at coping with noise during speech recognition, the presence of noise in the acoustic environment can render the task of speech perception more difficult for any listener. Hearing scientists often characterize the interference imposed on target signals by noise in terms of energetic and informational masking (Kidd et al., 2007). Noise of any type imposes energetic masking on speech signals when spectral and temporal overlap between the noise and the signal leads to interference in the auditory periphery, rendering the signal inaudible and, consequently, reducing the available acoustic and linguistic cues relevant to speech understanding. Informational masking typically refers to any reduction in target intelligibility that cannot be explained by energetic masking (though see Durlach, 2006 and Kidd et al., 2007 for a nuanced discussion of this definition). Informational masking is said to occur, for example, when both target speech and noise are audible but a listener has trouble separating one from the other. In speech-in-speech scenarios, therefore, the noise can impose both energetic masking and informational masking on target speech intelligibility.

A key strategy for investigating the effects of noise on speech processing and for ultimately developing a principled account of these effects is to compare different types of noise, which vary with respect to the kind and degree of interference they impose on speech signals. To this end, linguists and audiologists have employed a variety of noise types, including single-talker maskers, multi-talker babble with various numbers of talkers, and non-speech noise (i.e., speech-shaped noise and white noise, with and without temporal modulations). For speech noise in particular, such studies have shown that greater similarity between masker and target voices with respect to characteristics such as vocal tract size and fundamental frequency decreases intelligibility (Brungart et al., 2001). Target intelligibility also generally decreases as additional voices are added to multi-talker babble (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart et al., 2001; Rhebergen et al., 2005; Simpson and Cooke, 2005), although a recent study on consonant identification by Simpson and Cooke (2005) showed that this relationship was non-monotonic. Interestingly, Simpson and Cooke also found that intelligibility scores were lower in natural babble than in babble-modulated noise when there were more than two talkers in the noise. Similarly, Sperry et al. (1997) showed that listeners had more difficulty on a word recognition task in the presence of multi-talker babble compared to both reversed babble and to matched, amplitude-modulated speech-shaped noise. Since natural babble can induce masking effects over and above those observed for energetically matched noise, the additional masking can be primarily considered a form of informational masking.

Since speech noise signals crucially contain linguistic information, the informational masking imposed by them is likely due to linguistic interference at some level of higher auditory and/or cognitive processing. That is, the presence of linguistically meaningful signals in competing speech or multi-talker babble likely diverts processing resources from target speech identification and/or disrupts such processing. This diversion of processing resources may occur at any of a number of levels of linguistic processing – prosodic, phonological, lexical, or semantic. If the processing of linguistic information in the noise drives informational masking, as suggested here, then it is expected that listeners' experience with the language of the noise will affect the extent to which that noise interferes with target speech recognition. This dissertation aims, therefore, to investigate how listeners' long-term language experience both with the target language and the noise language may modulate linguistic interference, as well as whether short-term training can mitigate it.

One useful method for assessing the effects of linguistic factors in speech perception in noise is to examine the effects of speech maskers of different languages on a listener population. Since the long-term average speech spectrum is similar across languages (Byrne et al., 1994), such comparisons roughly equate energetic masking while varying maskers' linguistic content and its meaningfulness to listeners. Using this method, recent studies have shown that, indeed, the *language* spoken in speech noise affects the amount of interference experienced by listeners (Rhebergen et al., 2005; Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Calandruccio et al., in press). In each of these studies, native speakers of the target language had more difficulty coping with native- versus foreign-language babble. That is, the masking of speech signals by speech noise is, it appears, modulated by the linguistic content of the signal and the noise with respect to a listener's language background. These results provide additional evidence that an understanding of linguistic interference as it relates to listener experience is crucial both to a complete account of speech-in-speech perception and to the effective development of testing and rehabilitative techniques for people with difficulty understanding speech in noise.

Building on the insight that speech intelligibility in speech noise is subject to experience-related modification as suggested by the observation of differential effects of different noise languages, this dissertation has two primary aims: 1) to expand our understanding of the relationship between language experience and speech intelligibility in speech noise and 2) to identify parameters of auditory training that can benefit speech-in-speech intelligibility. The first aim is addressed by assessing English sentence intelligibility in English and Mandarin babble by monolingual English speakers, non-native (late bilingual) speakers of English whose native language is Mandarin, and early bilingual speakers of Spanish and English (Chapters 3 and 4). The second aim is addressed with a speech-in-noise training study (Chapter 5).

Significance of this research

Research on the linguistic factors involved in the processing of speech in speech noise contributes to the fields of both linguistics and hearing science. As this dissertation will show, the effects of speech noise on target speech recognition can be modulated by language experience and by training experience. Models of speech perception that aim to accommodate and explain the various distortions of speech signals that listeners encounter in day-to-day speech processing, therefore, must take such interactions into account. Further, the needs of populations with special hearing and language needs will be better met once we have a clear understanding of the linguistic factors involved in speech-in-speech perception. The research presented in this dissertation provides new information about how listeners' language background and the language of interfering speech contribute to speech-in-speech intelligibility. Furthermore, the training study provides insight into parameters of auditory training that may be useful for improving speech-in-speech understanding. This research, therefore, advances our understanding of the everyday problem of speech perception in noise and contributes valuable information for the development of speech-in-noise testing and training programs.

Overview of the dissertation

The goal of this dissertation is to extend our understanding of speech recognition in speech noise, with a particular focus on the effects of language experience and short-term auditory training. The groundwork for the experiments presented here can be found in Van Engen and Bradlow

(2007), which showed that native English listeners had greater difficulty on keyword recognition in sentences embedded in English 2-talker babble versus Mandarin 2-talker babble at difficult signal-to-noise ratios (SNRs). The studies in this dissertation build on this result and the methodology employed in that study. In Chapter 2, I first present two experiments that directly followed up on the findings of Van Engen and Bradlow. The first experiment investigated whether the noise language effect is primarily driven by the presence and activation of real English content words in the noise (i.e., whether interference takes place primarily at the level of lexical processing) by comparing the effects of babble composed of sentences with real English content words versus sentences with non-words in place of real content words. This comparison revealed no significant difference in listeners' performance on the same sentence recognition task used in Van Engen and Bradlow (2007), ruling out a strict lexical explanation for the noise language effect observed in that study. The second follow-up experiment (also described in Chapter 2) investigated the effect of overall stimulus level in Van Engen and Bradlow (2007). In that study, different SNRs were generated by adjusting the noise level relative to a constant signal level. A decrease in SNR, therefore, entailed an increase in the overall level of the stimuli (signal + noise). The follow-up study replicates Van Engen and Bradlow (2007) using stimuli that were equated for overall RMS amplitude after the signal and noise were mixed. As in Van Engen and Bradlow (2007), keyword identification scores on the English sentence recognition task were lower in English 2-talker babble than Mandarin 2-talker babble at an SNR of -5 dB, but not at 0 dB. There were no significant differences between listeners' performance in the two experiments. This finding shows that the interaction of the noise language effect with SNR observed in Van Engen and Bradlow and the general difference across SNRs could not be attributed to overall energy changes across SNRs, but rather revealed true speech-in-noise processing effects.

In Chapter 3, I investigated the effects of English and Mandarin 2-talker babble on English keyword identification in sentences at several SNRs for native speakers of English and native speakers of Mandarin. This study showed, first of all, that the non-native listeners (the Mandarin-speaking group) required significantly higher SNRs to achieve performance levels similar to those of the native listeners on the speech-in-speech task. Performance levels were successfully normalized using a novel method: each listener was tested at SNRs selected relative to his/her performance on the Hearing in Noise Test (HINT), a standardized test of full sentence recognition in speech-shaped noise. With respect to noise language effects, both listener groups showed lower keyword identification scores in English babble versus Mandarin babble, suggesting that a match between the target and noise languages may be the primary contributor to the previously-observed noise language effects. In addition, however, the English listeners received a greater release in masking in the Mandarin babble relative to the English babble than did the Mandarin listeners, suggesting that the native language status of the noise also contributes significantly to the interference listeners' experience when listening to speech in competing speech noise.

Chapter 4 employed the methods from Chapter 3 in a study of early Spanish-English bilingual listeners with the goal of assessing the effects of early bilingualism on speech-in-speech recognition. This group of listeners did not differ significantly from native English listeners on the task of sentence recognition in English and Mandarin 2-talker babble, suggesting that this particular task may not be sensitive to differences between monolingual and bilingual speech perception.

In Chapter 5, the effects of speech-in-noise training in speech-shaped noise, Mandarin 2-talker babble, and English 2-talker babble were compared. After two days of training, listeners from these three training groups (along with an untrained control group) were given a speech-in-speech post-test that included English and Mandarin babble and two target talkers: a familiar talker (from training) and a novel talker. Performance during the training sessions showed that listeners were able to improve their performance from day 1 to day 2 in babble, but not in speech-shaped noise. Test performance showed that, first of all, listeners who had received training were able to take advantage of talker familiarity in the post-test, identifying more keywords from the familiar talker than from the novel talker. With respect to the different training conditions, listeners who were trained in babble generally outperformed those who were trained in speech-shaped noise on the post-test, and interaction effects suggest that the benefit of babble training was specific to the language of the training babble. In addition to training effects, this study also showed that, after selecting test SNRs based on listeners' performance on the HINT, SNR was still a significant predictor of success on the speech-in-speech test. That is, listeners with higher (i.e., worse) thresholds on the HINT received correspondingly easier SNRs for the speech-in-speech test, and their performance on the post-test was generally higher than that of listeners with better HINT scores (who therefore received more difficult SNRs for the post-test). Working memory capacity was also a significant predictor of correct keyword identification.

Organization of the dissertation

This document is divided into six chapters. The current chapter (1) provides the reader with a general introduction and overview of the research presented in chapters 2-5. Chapters 3 (speech-in-noise recognition by native and non-native listeners) and 5 (speech-in-noise training) comprise the primary empirical content of the dissertation. They were written as independent research papers, and therefore contain their own full introduction, methods, results, and discussion sections. Chapters 2 and 4, which present data from experiments that were not included in the primary papers, are presented here as additional contributions to the overall research program. Chapter 6 serves as a general conclusion, discussing implications of the studies presented in the dissertation and directions for future research.

CHAPTER 2: TWO FOLLOW-UP STUDIES TO VAN ENGEN AND BRADLOW (2007)

This chapter reports briefly on two different experiments that were conducted to follow up on questions raised by the results in Van Engen and Bradlow (2007). Experiment I probes the effect of noise language observed in Van Engen and Bradlow by comparing the effects of 2-talker babble comprised of sentences containing real words and sentences containing non-words. Experiment II assesses the method of SNR manipulation used in Van Engen and Bradlow (changing noise levels relative to a static target level) by replicating a condition from that study using stimuli that were re-equated for overall output level after being mixed with noise. These two experiments are presented below, with their own short introductions, methods, results, and discussion sections.

Experiment I: Real-word vs. Non-word babble

Introduction

Van Engen and Bradlow (2007) showed that, particularly at difficult SNRs, native speakers of English had more difficulty identifying keywords in sentences in the presence of English 2-talker babble versus Mandarin 2-talker babble. This result – that the language of interfering noise can affect the intelligibility of target speech – was taken as evidence that, under certain conditions, linguistic interference plays a role in the perception of speech in speech noise. The precise aspects of the linguistic content of noise that may contribute to the observed language effect, however, remain to be determined. One possibility is that the language effect is primarily a lexical effect, such that hearing and activating English words from the babble is what made

English a more effective masker than Mandarin for native English listeners. Indeed, participants often transcribed entire words from the babble tracks in their responses, indicating that interference did occur at this level.

The current study aims to begin to identify the locus of the noise language effect by comparing the effects of English 2-talker babble with English 2-talker babble composed of sentences whose content words are, in fact, non-words (words that are phonologically legal in English but are not real words). This study will allow us to determine whether the simple presence of English content words in the background noise drives the greater interference listeners experienced with English babble or whether sub-lexical features of English noise (still present in the non-word noise) may be equally as detrimental to target sentence recognition as real words.

Methods

Participants

Seventeen participants were recruited from the Northwestern community and paid for their participation. All participants were between the ages of 18 and 35, were native speakers of American English, and reported no history of problems with speech or hearing. The data from one participant were excluded from analysis because the individual was outside the required age range.

Materials

Target sentences

The same target sentence recordings were used as in Van Engen and Bradlow (2007). The recordings, produced by a female, native speaker of American English were made for an unrelated study (Bent and Bradlow, 2003). The sentences were from the Revised Bamford-Kowal-Bench Standard Sentence Test (BKB sentences), lists 7-10. Each list contains 16 simple, meaningful English sentences (e.g., *The children dropped the bag*.) and 50 keywords (3-4 per sentence). Lists 7, 8, 9, and 10 were chosen for Van Engen and Bradlow (2007) and this study based on their approximately equivalent intelligibility scores for normal children (Bamford and Wilson, 1979). All sentence recordings were equated for RMS amplitude.

Noise

The semantically anomalous sentences used to generate babble in Van Engen and Bradlow (2007) (20 sentences created by Smiljanic and Bradlow, 2005) were recorded by two new female speakers of American English. These talkers also recorded a second version of the sentence set, in which all of the content words were converted to non-words by manipulating onsets, codas, or vowels. For example, the real-word sentence “Your tedious beacon lifted our cab” was converted to “Your bedious reacon loofted our bab.” (The two sentence lists can be found in Appendix A.) Four short babble tracks were created from each sentence set following the procedure described in Van Engen and Bradlow (2007) and again in Chapter 3 of this dissertation. The noise tracks were equated for RMS amplitude at three levels relative to the level of the target sentences to produce SNRs of +5 dB, 0 dB, -5 dB.

Procedure

Listeners were seated in a sound-attenuated booth facing a computer monitor. Stimuli were presented diotically over headphones (Sennheiser HD 580). Listeners were presented with a set of 16 target sentences (containing 50 keywords for scoring) in non-word babble and then 16 sentences in real-word babble at SNRs of +5 dB, 0 dB, and -5 dB (in that order). The sentence sets were counter-balanced across the different SNR conditions. Listeners were told they would be listening to sentences in noise and were asked to write down what they heard. They were told to guess if they were unsure and to report individual words if that was all they could identify. The task was self-paced, and listeners advanced from sentence to sentence by pressing the space bar on the computer.

The easiest (+5 dB) block was included as a familiarization phase so that listeners could adjust to the task before the harder SNRs were presented. Performance was analyzed for the SNRs of 0 dB and -5 dB – the levels at which noise language effects were observed in Van Engen and Bradlow (2007) for English versus Mandarin babble.

Data analysis

An experimenter scored listeners' response sheets by hand. Perception scores were determined by a strict keyword identification count. Each set of sentences contained 50 keywords for scoring, and listeners received credit for all correctly transcribed keywords. Obvious spelling errors or homophones were considered correct, but any words with added or deleted morphemes were considered incorrect. Raw scores were converted to percentage correct scores, and then to rationalized arcsine units (Studebaker, 1985).

Results

The data from the 0 dB and -5 dB conditions were analyzed using a repeated measures analysis of variance (ANOVA), with noise type (word vs. non-word) and SNR (0 dB vs. -5 dB) as within-subjects factors. This analysis revealed a main effect of SNR [$F(1, 15) = 146.668$, $p < .0001$], but no significant effect of noise type and no interaction between the factors. The intelligibility scores in RAU are shown in Figure 2.1 below.

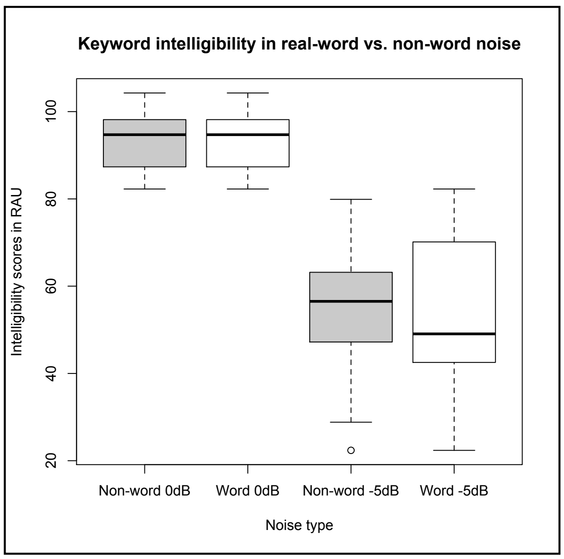

Figure 2.1. Keyword intelligibility scores (in RAU) in real-word and non-word noise at SNRs of 0 dB and -5 dB.

Discussion

There was no significant difference in target speech intelligibility in real-word versus non-word noise, a finding that provides evidence against a strict lexical explanation for the noise language effect observed in Van Engen and Bradlow (2007). However, the method employed to generate the non-word noise (i.e., converting real words to non-words by manipulating onsets, codas, or vowels) meant that there was a high degree of similarity between the non-words in the babble and real English words. Given this similarity, there may still have been activation of lexical items in the non-word babble condition. This interpretation is supported by the fact that participants frequently transcribed real words that sounded like the non-words that were actually present in the babble. Further, the fact that word-like non-words interfered just as much as real words may, if anything, suggest a deep influence of lexical processes in the interference imposed by speech noise.

Future research is required to further delineate the linguistic features that may be most important in driving the noise language effect. For example, such studies might employ other types of non-words (that are less similar to real English words), various accents of the target language, other languages, or babble constructed from non-sentential materials, such as word lists or syllable strings.

Some of these noise types have been utilized in ongoing collaborative research at Northwestern University and cooperating institutions. In collaborations with Lauren Calandruccio, Susanne Brouwer, Sumit Dhar and Ann Bradlow, we have investigated, for example, maskers in different speech styles, languages (e.g. English, Dutch, Mandarin, Croatian) and with varying amounts of semantic content at the sentence level.

Experiment 2: An investigation of SNR manipulation methods

Introduction

In speech-in-noise tests and experiments that utilize a range of signal-to-noise ratios (SNRs), SNR is usually manipulated either by changing the noise level relative to a constant signal level or by changing the signal level relative to a constant noise level. In the Hearing in Noise Test (HINT), for example, noise level is held constant while target sentence level changes. Such manipulations entail differences in the overall energy delivered to the ear across SNR conditions. It is possible, then, that overall energy changes are partially responsible for behavioral differences observed across SNRs. For example, Van Engen and Bradlow (2007) showed that English 2-talker babble was more detrimental than Mandarin 2-talker babble for English sentence recognition by native English listeners at difficult SNRs but not at easier SNRs. Specifically, in a condition where listeners were presented with SNRs of +5 dB and 0 dB, the language effect was significant at 0 dB only, and in a condition where listeners received SNRs of 0 dB and -5 dB, the effect was significant at -5 dB only. In Van Engen and Bradlow (2007), SNR was made more difficult by raising the noise level relative to the target sentence level. In order to investigate whether overall energy changes may have been responsible for the results in Van Engen and Bradlow (2007), the current study replicates one condition from the study (0 dB and -5 dB), re-leveling the mixed signal + noise files so that the rms amplitude is consistent across the two SNRs that were presented to listeners.

Methods

Participants

Seventeen native English speakers were recruited from the Northwestern University Linguistics Department subject pool and received class credit for their participation. No participant reported any speech or hearing problems.

Materials

Target sentences and 2-talker babble in English and Mandarin were taken from Van Engen and Bradlow (2007). The process for constructing this babble is also described in Chapter 3 of this dissertation. The target sentences were BKB sentences spoken by a female native English talker (Bamford and Wilson, 1979), and the babble was constructed from semantically anomalous sentences developed in English by Smiljanic and Bradlow (2005) and translated into Mandarin by Van Engen and Bradlow (2007). These sentences were spoken by female native talkers of English and Mandarin.

The different SNRs were generated as follows:

1. All individual sentence files were leveled to a given rms amplitude (65 dB).
2. All noise tracks were then leveled to the rms amplitudes required to produce 0 and -5 dB SNRs (65 dB, 70 dB).
3. The sentences were mixed with the babble files, and the resultant files were again equated for rms amplitude. In this way, the two different SNRs could be played to participants at the same overall output level. The differences in overall output across the two SNR conditions for Van Engen and Bradlow (2007) (Method 1) and the present

experiment (Method 2) are shown in Table 2.1 below. By design, stimuli produced according to Method 2 do not differ in rms amplitude across SNR conditions. The stimuli produced using Method 1 showed an overall amplitude increase of approximately 4 dB in the -5 dB SNR condition as compared to the 0 dB SNR condition (3.91 dB for stimuli with Mandarin babble and 3.96 dB for stimuli with English babble).

Differences in average overall rms amplitude across SNR conditions (0 dB SNR to -5 dB SNR)		
	Mandarin	English
Method 1	3.91 dB	3.96 dB
Method 2	0 dB	0 dB

Table 2.1. Differences in rms amplitude across SNR conditions using two methods for presenting different SNRs. Method 1 was used in Van Engen and Bradlow (2007); Method 2 was investigated in the current experiment.

The assessment of performance in English and Mandarin babble across SNRs of 0 and -5 dB using Method 2, therefore, allows us to investigate whether the approximately 4 dB difference in overall stimulus level across SNRs in Van Engen and Bradlow (2007) may have been partially responsible for the speech-in-noise effects observed in that study.

Procedure

The testing procedure was exactly the same as in condition 4 of Van Engen and Bradlow (2007). Each participant listened to one BKB list (16 sentences, 50 keywords for scoring) in Mandarin babble at an SNR of 0 dB, then one list in English babble at 0 db, then one list in Mandarin at -5 dB, and finally one list in English at -5 dB. Since the English babble was predicted to be more difficult than the Mandarin babble, it was presented after the Mandarin in each SNR condition in order to stack the cards against the predicted language effect. That is, any learning effects (task learning, adaptation to the target talker) would, if anything, reduce the predicted difference between the two babble languages by elevating performance in English babble.

Data analysis

Analysis was the same as in Van Engen and Bradlow (2007). Perception scores were determined by a strict count of keywords correctly identified. Listeners received credit for each keyword that was transcribed perfectly. Words with added or deleted morphemes were counted as incorrect, but homophones or obvious spelling errors were counted as correct. Raw scores were converted to rationalized arcsine units (RAU) for statistical analysis (Studebaker, 1985).

Results

Listener performance on this experiment was compared to listener performance on Condition 4 from Van Engen and Bradlow (2007), in which the harder SNR was generated by raising the level of the noise relative to the target and *not* re-leveling the resultant stimuli to equate rms amplitude across the 0 dB and -5 dB SNR stimuli. A three-way repeated measures ANOVA with SNR (easy vs. hard) and noise language (English vs. Mandarin) as within-subjects factors and Experiment (original vs. re-leveled noise) as a between-subjects factors showed a significant main effect of SNR [$F(1,30) = 139.013$, $p < 0.0001$], a significant main effect of noise language [$F(1,30) = 16.633$, $p = 0.0003$] and a significant interaction between SNR and Language [$F(1,30) = 16.361$, $p = 0.0003$). There was no significant main effect of Experiment and no significant 2- or 3-way interactions involving Experiment. The analysis, therefore, reveals no significant differences in listener performance across the two experiments. The results are shown in Figure 2.2 below.

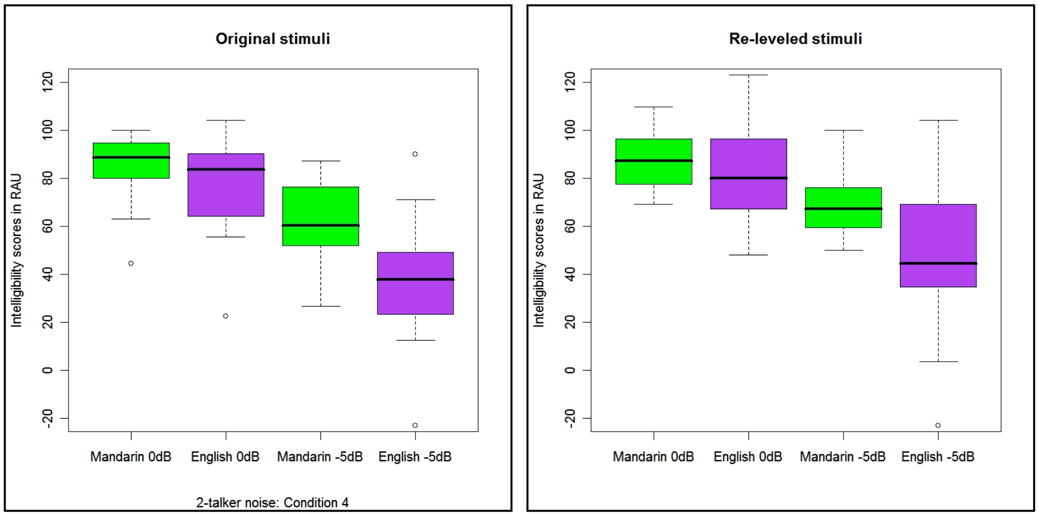

Figure 2.2. Keyword identification scores (in RAU) from Van Engen and Bradlow (2007), Condition 4 (left panel) and in the current experiment (right panel).

In both experiments, post-hoc analyses showed that listener performance was significantly worse in English versus Mandarin babble at the harder SNR (-5 dB), but not at the easier SNR (0 dB). Note that, in Van Engen and Bradlow (2007), a significant noise language effect *was* observed at an SNR of 0 dB when that SNR was in the second half of the experiment. This pattern of results suggests that the novelty of the task in block one may have mitigated noise language effects that could otherwise be observed at an SNR of 0 dB.

Discussion

This study shows that, for young adult listeners with normal hearing, behavioral results on a test of sentence intelligibility in 2-talker babble were not affected by whether or not the mixed signal + noise stimuli at different SNRs were re-leveled to equate rms amplitude. This result suggests that the interaction of the noise language effect with SNR observed in the Van Engen and Bradlow (2007), and the general difference across “easy” (0 dB) and “hard” (-5 dB) SNRs cannot be attributed to overall energy changes across SNRs, but rather reveal speech-in-noise processing effects. More generally, these findings suggest that behavioral results with normal-hearing young adults are relatively stable across these two different methods of manipulating SNR.

CHAPTER 3: SIMILARITY AND FAMILIARITY: SECOND LANGUAGE SENTENCE RECOGNITION IN FIRST- AND SECOND-LANGUAGE MULTI-TALKER BABBLE

Published in *Speech Communication*, 2010

Kristin J. Van Engen

Abstract

The intelligibility of speech in noisy environments depends not only on the listener's auditory system, but also on cognitive factors such as language learning experience. Previous studies have shown that listeners attending to a non-native language have more difficulty identifying speech targets in noisy conditions than do native listeners. Furthermore, native listeners have more difficulty understanding speech targets in the presence of interfering noise in their native language versus a foreign language. The present study addresses the role of listener language experience with both the target and noise languages by examining second-language sentence recognition in first- and second-language background noise. Native English speakers and L2 English speakers whose L1 is Mandarin were tested on English sentence recognition in English and Mandarin 2-talker babble. Results show that both listener groups experienced greater difficulty in English versus Mandarin babble, but that native Mandarin listeners experienced a smaller release from masking in Mandarin versus English babble as compared with the native English listeners. These results indicate that both the acoustic and/or linguistic similarity between target and noise and the language experience of the listeners contribute to the amount of interference listeners experience when listening to speech in the presence of interfering speech noise.

Keywords: speech-in-noise perception, informational masking, multi-talker babble, bilingual speech perception

Introduction

During everyday speech communication, listeners must cope with a variety of competing noises in order to understand their interlocutors. While it is well known that trouble understanding speech in noisy environments is a primary complaint for listeners with hearing loss, the ability to process speech in noise depends not only on the peripheral auditory system, but also on cognitive factors such as a listener's language experience (e.g., Mayo et al., 1997). For normal-hearing native-speaking listeners, speech intelligibility remains relatively robust even in adverse conditions. Such listeners are able to take advantage of redundancies in the speech signal (e.g., Cooke, 2006 and citations therein), as well as contextual cues at higher levels of linguistic structure, such as lexical, syntactic, semantic, prosodic, and pragmatic cues (e.g., Bradlow and Alexander, 2007). When people listen to speech in a *second* language, however, they have greater difficulty identifying speech signals (phonemes, words, sentences) in noisy conditions than do native speakers (Nábělek and Donohue, 1984; Takata and Nábělek, 1990; Mayo et al., 1997; Garcia Lecumberri and Cooke, 2006; Cooke et al., 2008; Cutler et al., 2008). Some recent data suggest that even bilinguals who acquired both languages before age 6 may have greater difficulty recognizing words in noise and/or reverberation than monolingual listeners (Rogers et al., 2006). Furthermore, when the interfering noise is also a speech signal (as in the case of multi-talker babble or a competing speaker), listeners' experience with the language of the *noise* seems to modulate their ability to process target speech: native language noise has been shown to be more detrimental than foreign-language noise for listeners' identification of native language speech targets (Rhebergen et al., 2005; Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Calandruccio et al., in press). The present study further investigates the role of listeners' language experience in the perception of speech in noise, extending the research on both non-native speech perception and on the effects of different noise languages by examining the effects of first- and second-language noise on sentence intelligibility for listeners who are processing their second language.

Given that the effects of noise on speech perception can vary based on non-peripheral factors such as a listener's language experience, it is useful to consider the contrast drawn by hearing scientists between energetic and informational masking (see Kidd et al., 2007 and citations therein). Noise imposes energetic masking on auditory speech targets when mechanical interference occurs in the auditory periphery: components of the speech signal are rendered inaudible where there is spectral and temporal overlap between the noise and the signal. Energetic masking, therefore, is dependent on the interaction of the acoustics of the speech signal and the noise signal, and it results in the loss of acoustic and linguistic cues relevant to speech understanding. Any reduction in target speech intelligibility that is *not* accounted for by energetic masking (e.g., when both target and noise are audible, but a listener has trouble separating them) is typically described as informational masking.[1] This contrast between energetic and informational masking will be useful as we consider the effects of interfering speech on speech perception by listeners with varying language backgrounds: the relative energetic masking

[1] Durlach (2006) observed that this very broad use of the term 'informational masking' reflects a lack of conceptual and scientific certainty or clarity. Kidd et al. (2007) provide a useful history and overview of the terms energetic and informational masking.

imposed by two noise types (English and Mandarin babble in this study) is necessarily static across listener groups, whereas the relative *informational* masking of the two noises may be modulated by listeners' language experience.

Before proceeding to a discussion of previous literature on speech perception in noise by non-native listeners, it is helpful to clarify the terms that will be used in this paper to discuss various types of noise. 'Noise' is intended to refer to any sounds in an auditory environment other than the speech to which a listener is attending. 'Masker' will be used to refer to noise that is used in experimental settings. A single interfering talker is referred to as 'competing speech', and more than one interfering talker is 'multi-talker babble' or 'babble'. 'Non-speech noise' refers to any noise that is not comprised of speech (e.g., white noise), and 'speech-shaped noise' is a type of non-speech noise that is generated by filtering broadband noise through the long-term average spectrum of speech.

With respect to the effects of noise on non-native speech perception, several studies have shown that native listeners perform better than non-natives on speech perception tasks in stationary (i.e., without amplitude modulations) non-speech noise and multi-talker babbles containing several talkers (Mayo et al., 1997; Hazan and Simpson, 2000; Bradlow and Bent, 2002; Van Wijngaarden et al., 2002; Cutler et al., 2004; Garcia Lecumberri and Cooke, 2006; Rogers et al., 2006). These studies employed listeners from a wide range of native language backgrounds and used a variety of types of speech targets and noise. Mayo et al. used English target sentences in English 12-talker babble, and their listeners were native speakers of English, native speakers of Spanish, and early learners of both English and Spanish. Hazan and Simpson used English speech targets (VCV syllables), speech-shaped noise, and listeners who were native speakers of English, Japanese and Spanish. Bradlow and Bent (2002) used English sentences in white noise and listeners who were native speakers of English and a wide range of other languages. Cutler et al. (2004) used English CV and VC syllables as targets, English 6-talker babble, and native English and native Dutch listeners. Van Wijngaarden et al.'s (2002) targets were English, Dutch, and German sentences in speech-shaped noise, and their listeners were native speakers of Dutch, English, and German. Garcia Lecumberri and Cooke (2006) used English VCV syllables and native English and Spanish listeners. This study employed non-speech noise, English 8-talker babble, and competing speech in both English and Spanish. Only this study, therefore, investigated the effects of native and second-language noise (in the form of competing speech) on native and non-native listeners of a given language. Those results will be discussed in greater detail below.

The above studies all show, in general, poorer performance by non-native listeners on speech perception tasks in noise relative to native speakers. As noted by Cooke et al. (2008), estimates of the relative size of the native listener advantage across different levels of noise have differed across these studies. While some show that the native listener advantage increases with increasing noise levels, others show constant native listener advantages across noise levels. The size of these effects seems to be related to the nature of the speech perception task (tasks in these studies range from phoneme identification to keyword identification in sentences) and/or the precise methods used (Cutler et al., 2008). Differences aside, however, all of these studies show that non-native listeners have more difficulty identifying speech targets in noise than native listeners.

Many of the noise types used in these studies would induce primarily energetic masking (many used non-speech noises or babbles with many talkers). The specific effect of informational masking on non-native listeners of English was more recently investigated by Cooke et al. (2008). In this study, Cooke et al. explicitly investigated the roles of energetic and informational masking by comparing the effects of a primarily energetic masker (stationary non-speech noise) with a primarily informational masker (single competing talker). They found that increasing levels of noise in both masker types affected non-native listeners more adversely than native listeners. Further, a computer model of the energetic masking present in the competing talker condition showed that the intelligibility advantage for native listeners could not be attributed solely to energetic masking. The authors conclude, therefore, that non-native listeners are more affected by informational masking than are native listeners.

Cooke et al. (2008) also respond to Durlach's (2006) observation regarding the lack of specificity in the term 'informational masking' by identifying several potential elements of informational masking: misallocation of audible masker components to the target, competing attention of the masker, higher cognitive load, and interference from a "known-language" masker. In the discussion of their observed effects of informational masking on non-native listeners, then, they suggest that such listeners might suffer more from target/masker misallocation, since their reduced knowledge of the target language (relative to native listeners) might lead to a greater number of confusions. Furthermore, they suggest that influence from the non-native listeners' native language (L1) might also result in more misallocations of speech sounds. In addition to misallocation, they also suggest that the higher cognitive load in their competing talker task (relative to the stationary noise task) may affect non-native listeners more than native listeners, given that some aspects of processing a foreign language are slower than processing a native language (Callan et al., 2004; Clahsen and Felser, 2006; Mueller, 2005). Finally, they suggest that the tracking and attention required to segregate speech signals may be compromised in non-native listeners since they have a reduced level of knowledge of the useful target language cues and/or may experience interference based on cues that are relevant for segregation of signals in the L1.

Since their study focused on a comparison of stationary noise and competing speech in the target language, Cooke et al. (2008) did not address the potential effects on non-native listeners of their final proposed aspect of informational masking: interference from a "known-language" masker. In this study, we specifically investigate this aspect of informational masking by comparing the effects of native (L1) and second-language (L2) babble on L2 sentence recognition. Native, monolingual English listeners and L2 English listeners whose L1 is Mandarin were tested on English target sentences in the presence of English 2-talker babble and Mandarin 2-talker babble. While it has been shown that English-speaking monolinguals have greater difficulty with English-language maskers as compared to foreign-language maskers (Rhebergen et al., 2005; Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Calandruccio et al., in press), the effects of different language maskers on L2 listeners have not been thoroughly examined.

As mentioned above, the one study in which L2 listeners were tested on speech targets in L1 and L2 noise is Garcia Lecumberri and Cooke (2006). This study investigated the performance of L1

Spanish listeners on L2 (English) consonant identification in L1 and L2 competing speech, and found no difference between the listeners' performance in the two noise languages. The authors suggest that while L1 noise might be generally more difficult than L2 noise to tune out, the task of identifying L2 targets might increase interference from L2 noise, thereby eliminating the difference between the masking effects of the two languages. The present study further investigates the effects of noise language on non-native listeners by asking whether these listeners are differentially affected by L1 and L2 babble when identifying L2 sentences.

In addition to simulating an ecologically valid listening situation, sentence-length materials contain all the acoustic, phonetic, lexical, syntactic, semantic, and prosodic cues of everyday speech, and may, therefore, reveal differences between the effects of different noise types that would not be observable at the level of a phoneme identification task. With sentences, listeners are able to use redundancies in the speech signal as well as contextual linguistic cues that aid speech understanding in real-world situations. Such cues may aid perception in noise in general, but if informational masking occurs at higher levels of linguistic processing, sentence materials may also make it possible to observe differences in the effects of different noise languages. As suggested by Cutler et al. (2004), non-native listeners' difficulty in noise may reflect an accumulation of difficulties across levels of speech processing. In this case, differential effects of noise languages which may not be observable at the level of phoneme identification could be observed using materials that require the processing of more levels of linguistic structure.

By including participants who speak both babble languages (i.e., the native Mandarin group), the current study addresses another open question regarding the previously-observed differential masking by native- versus foreign-language noise on native language targets. In the studies that showed this effect, the target language was the native language of the listeners, so the native-language babble (or competing speech) also matched the language of the target speech. It is possible, therefore, that the greater acoustic and/or linguistic similarity between the target and noise signals contributes importantly to the increased masking by native- versus foreign-language babble, regardless of the listeners' experience with the languages. With respect to acoustics, English target speech and English babble may, for example, have more similar spectral properties, leading to greater energetic masking. As for linguistic factors, English target speech and English noise share a wide range of properties (e.g., phonemes, syllable structures, prosodic features), which may make the segregation of English speech targets from English noise much more difficult—i.e., shared linguistic features may lead to greater informational masking, regardless of the native-language status of English. The present study will enable us to begin to understand, then, whether the noise language effect is primarily a *same-language* effect (i.e., similarity between target and noise leads to increased masking) or primarily a *native-language* effect (i.e., native language noise necessarily imposes more masking than another language). For the English listeners, English babble is their native language and it matches the target language. For the Mandarin listeners, however, English babble matches the target language, but Mandarin babble is their native language.

Using a different target talker from Van Engen and Bradlow (2007), we expect to replicate the finding that native English listeners have greater difficulty understanding English sentences in English versus Mandarin babble. This replication would provide additional support for the validity of the previously-observed noise language effect by showing that the effect cannot be attributed solely to the acoustic properties of a particular target voice and its interaction with the two babbles.

The performance of the Mandarin listeners in the two babble languages will allow for the comparison of interference from native-language noise versus noise in the language of the speech targets. If differential noise language effects are primarily driven by the native language status of noise, then the Mandarin babble may be more disruptive than the English babble. If such effects are primarily a result of the similarity between the target and noise languages, then the English babble may be more disruptive than the Mandarin. Finally, we may see evidence for important roles of both factors in modulating the interference that listeners experience from interfering speech. In this case, the similarity of the English babble to the target speech would make it more difficult than Mandarin babble for all listeners (due to increased energetic and/or informational masking), but the Mandarin listeners would show a smaller release from masking in Mandarin babble (i.e., a smaller performance gain in Mandarin babble relative to English babble) than the native English listeners. Crucially, this study investigates whether there are, indeed, different effects of L1 and L2 babble on L2 sentence recognition, and further, compares such effects across L1 and L2 listeners. The relative energetic masking imposed by the two noise languages is constant across the two groups, but their language experience varies and may, therefore, modulate informational masking.

Since the relative effects of English and Mandarin babble on the two listener populations is of primary interest, it was important to test both groups at signal-to-noise ratios (SNRs) that would result in similar levels of performance with respect to tolerance for energetic masking. To achieve this, listeners were tested at SNRs that were chosen relative to their individual performance on a standard speech perception test in stationary, speech-shaped noise (the Hearing in Noise Test (HINT), Nilsson et al., 1994). By normalizing the listeners according to their tolerance for energetic masking alone, the effects of two babble languages on two listener populations could be investigated.

Methods

The ability of each listener to understand sentences in non-speech noise (speech-shaped white noise) was measured with the Hearing in Noise Test (HINT), which employs an adaptive presentation method to estimate the SNR at which a listener can correctly repeat full sentences 50% of the time. This score was used to determine testing levels for the speech-in-babble test. Listeners were then presented with 4 blocks of 32 target sentences in 2-talker babble. Block 1 was presented at an SNR of HINT score +3 dB; Block 2 at HINT score +0 dB; Block 3 at HINT score -3 dB; and Block 4 at HINT score -6 dB. This range of SNRs was selected in order to observe performance at relatively easy and difficult noise levels and to avoid ceiling and floor effects. In each block, listeners heard a randomized set that included 16 sentences in English babble and 16 sentences in Mandarin babble (50 keywords in each). Methodological details are presented below.

Participants

Monolingual English listeners

Twenty-six undergraduate participants were recruited from the Northwestern University Linguistics Department subject pool and received course credit for their participation in the study. For the following reasons, 6 were omitted from the analysis presented here: 3 were bilingual; 2 had studied Mandarin; and 1 encountered a computer error. The remaining 20 participants were native speakers of English between the ages of 18 and 22 (average = 19.5), and all reported normal hearing. Four participants reported having received speech therapy in early childhood.

L2 English listeners

Twenty-one native speakers of Mandarin Chinese who speak English as a second language were recruited and paid for their participation. All of these participants were first-year graduate students at Northwestern University who were participating in the Northwestern University International Summer Institute, an English language and acculturation program that takes place during the month prior to the start of the academic year. One participant was excluded from analysis because she had lived in Malaysia for a number of years during childhood and therefore had had significantly different experience with English compared to the other participants, all of whom grew up in mainland China or Taiwan. The 20 included participants ranged from 22 to 32 years of age (average = 24.5), and none reported a history of problems with speech or hearing.

While English proficiency is not entirely uniform within this group, all participants had attained the required TOEFL scores for admission to the Northwestern University Graduate School and participated in the study within 3 months of their arrival in Evanston, Illinois.[2] In order to further

[2] Most participants had not spent a significant amount of time in an English-speaking country (0-2 months), but 4 participants reported having spent 2-3 years in the U.S. at an earlier time in their lives. These listeners were included because their HINT scores fell within the range of the other

characterize the L2 English participants' experience and proficiency in Mandarin and English, each person completed a lab-internal language history questionnaire and the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007). Table 3.1 provides basic information regarding the participants' English learning and proficiency.

Native Mandarin listeners	Mean (SD)
Age of initial English learning (age in years)	11.3 (1.3)
Years of formal English training	9.6 (2.2)
Self-reported TOEFL score (iBT)[3]	106.2 (6.4)
Self-reported proficiency – understanding English	6.1 (1.7)[4]
Self-reported proficiency – speaking English	5.8 (1.7)[5]

Table 3.1. Native Mandarin participants: English learning and proficiency information.

Materials

Multi-talker babble

Two-talker babble was used for this experiment, largely because significant effects of babble language have been observed for sentence recognition by native English speakers using 2-talker babble. In Van Engen and Bradlow (2007), for example, significant differences were observed between English and Mandarin babble for 2-talker babble but not for 6-talker babble. Calandruccio et al. (in press) also showed significant differences in the effects of English versus Croatian 2-talker babble. Finally, Freyman et al. (2004) showed maximal informational masking

---

participants', meaning they were not outliers with respect to the task of English sentence recognition in noise.

[3] 1 participant did not report any TOEFL scores; 3 reported paper-based test scores and 1 reported a computer-based score. All scores were converted to internet-based test scores using comparison tables from the test administration company (ETS, 2005).

[4] As rated by participants on a scale from 0 (*none*) to 10 (*perfect*). A rating of 6 indicates *slightly more than adequate*.

[5] As rated by participants on a scale from 0 (*none*) to 10 (*perfect*). A rating of 6 indicates *slightly more than adequate.*

effects in 2-talker babble as compared with 3-, 4-, 6-, and 10-talker babble. While the use of a single noise type (2-talker babble only) limits investigation of the particular contributions of energetic and informational masking in this study, the primary goal is to examine the *relative* effects of two noise languages on listener groups with different experience in the two languages. The relative energetic masking imposed by the two languages is constant across both groups, so differences in their relative effects can be attributed to informational masking.

Four 2-talker babble tracks were generated in English and in Mandarin (8 tracks in total). The babble was comprised of semantically anomalous sentences (e.g. *Your tedious beacon lifted our cab.*) produced by two adult females who were native speakers of English and two adult females who were native speakers of Mandarin. The sentences were created in English by Smiljanic and Bradlow (2005) and translated into Mandarin by Van Engen and Bradlow (2007). Female voices were used for the maskers and the target in order to eliminate the variable of gender differences, which can aid listeners in segregating talkers (e.g., Brungart et al. 2001). Babble tracks were created as follows: for each talker, two sentences (a different pair for each talker) were concatenated to ensure that the noise track duration would exceed the duration of all target sentences. 100 ms of silence were added to the start of one of the two talkers' files in order to stagger the sentence start times of the talkers once they were mixed together. The two talkers' files were then mixed using Cool Edit (Syntrillium Software Corporation), and the first 100 ms (in which only one talker was speaking) were removed so that the track only included portions where both people were speaking. The RMS amplitude was equalized across the finished babble tracks (4 in English; 4 in Mandarin) using Level16 (Tice and Carrell, 1998).

Target sentences

The target sentences come from the Revised Bamford-Kowal-Bench (BKB) Standard Sentence Test (lists 1, 5, 7-10, 15, and 21). These particular lists were selected based on their approximately equivalent intelligibility scores for normal hearing children as reported in Bamford and Wilson (1979). The BKB sentences were chosen for this study because they use a limited vocabulary that is appropriate for use with non-native listeners (see Bradlow and Bent (2002) for familiarity ratings from a highly similar population of non-native listeners). Each list contains 16 simple, meaningful English sentences and a total of 50 keywords (3-4 per sentence) for intelligibility scoring. An adult female speaker of American English produced the sentences. She was instructed to speak in a natural, conversational style, as if she were speaking to someone familiar with her voice and speech. Recording took place in a sound-attenuated booth in the Phonetics Laboratory at Northwestern University. The sentences appeared one at a time on a computer screen, and the speaker read them aloud, using a keystroke to advance from sentence to sentence. She spoke into a Shure SM81 Condenser microphone, and was recorded directly to disk using a MOTU Ultralight external audio interface. The recordings were digitized at a sampling rate of 22050 Hz with 24 bit accuracy. The sentences were then separated into individual files using Trigger Wave Convertor, an automatic audio segmentation utility developed in the Department of Linguistics at Northwestern University. The resultant files were trimmed to remove silence on the ends of the sentence recordings, and then the RMS amplitudes of all sentences were equalized using Level16 (Tice and Carrell, 1998).

Targets + Babble

The full set of target sentences was mixed with each of the 8 babble tracks using a utility that was developed in the Northwestern University Linguistics Department for the purpose of mixing large sets of signals. The targets and babble were mixed at a range of SNRs so that each participant could be tested at four SNRs relative to his/her HINT score (HINT +3 dB, +0 dB, -3 dB, and -6 dB). The various SNRs were generated by RMS-equalizing the babble tracks at various levels relative to a static target sentence level. This basic approach to SNR manipulation has been utilized in a large number of speech-in-noise studies (e.g., Mayo et al., 1997; Sperry et al., 1997; Van Wijngaarden et al., 2002; Rogers et al., 2006) and has the advantage of maintaining a constant target level across the entire experiment. Although this method entails that the overall level of the stimuli increases as SNR decreases (that is, when the noise becomes louder with respect to the signal), previous work showed that behavioral results on this type of task were unaffected when the mixed files were re-equalized (Van Engen, 2007).

The resulting stimuli each contained a 400 ms silent leader followed by 500 ms of babble, the target and the babble, and then a 500 ms babble trailer.

Procedure

In order to determine the SNRs at which participants would be tested in the speech-in-babble experiment, the Hearing in Noise Test (HINT) was administered first (details regarding the test and its materials can be found in Nilsson et al., 1994). Using an adaptive method of presentation, the HINT estimates the SNR at which a listener can understand 50% of entire sentences in speech-shaped noise (SNR-50). (For each sentence, listeners receive an all-or-nothing score, with some allowances for errors in short, frequently reduced function words, such as ‘a’ versus ‘the’.)

Listeners respond to each sentence by repeating it orally. A 20-sentence version of this test[6] was administered diotically through Sony MDR-V700DJ earphones, and listeners were seated in a sound-attenuated booth with an experimenter. HINT thresholds were rounded to the nearest whole number for the purposes of selecting the SNRs for the speech-in-speech test.[7]

For the speech-in-babble test, listeners were seated at a desk in the sound-attenuated booth. Stimuli were presented diotically over headphones at a comfortable level. Participants were presented with a total of 132 trials—four practice sentences followed by four experimental blocks containing 32 sentences each. Each block was comprised of two BKB lists – one mixed with English babble (four of the sentences with each of the four noise tracks), the other with Mandarin babble (four of the sentences with each of the four noise tracks). Within each block, all stimuli were randomized. Listeners were instructed that they would be listening to sentences mixed with noise, and that they should write down what they heard on a provided response sheet.[8] They were told to write as many words as they were able to understand, and to provide

[6] HINT lists 1 and 2 were used for HINT testing. Note that the original BKB sentences were used for the development of the HINT test (Nilsson et al., 1994), so there is considerable overlap between the two sets of sentences. In order to avoid the repetition of any sentence between the HINT test and the speech-in-babble test, matching sentences were removed from the selected BKB lists (1, 5, 7-10, 15, 21) and replaced with similar (matching in number of keywords and in sentence structure where possible) sentences from list 20. This amounted to a total of 7 replacements.

[7] The experiment-running software that was used for the speech-in-babble experiment required that signals and noise be mixed in advance of experimentation. Sentence targets and babble were mixed at whole-number SNRs to limit the number of required sound files to a manageable number.

[8] Written responses were used for this experiment because it can be problematic to score oral responses from non-native speakers of English. Given the difficulties native listeners have in understanding foreign-accented speech, there may be discrepancies between what a non-native participant intends to say and what the experimenter hears. Furthermore, it may be difficult to

their best guess if they were unsure. The task was self-paced; participants pressed the spacebar on a computer keyboard to advance from sentence to sentence. They heard each sentence only once.

Before the test began, listeners were familiarized to the task and the target speaker by listening to two sentences in English babble and two sentences in Mandarin babble, all at the SNR at which they would receive the first block of testing (their HINT score +3 dB). They were told that the target talker begins speaking one-half second after the noise comes on. The experimenter played these stimuli as many times as the listener needed in order to repeat the target sentence correctly. A few listeners were unable to repeat the target after several repetitions. In these cases, the experimenter read the target to the listener, who was then given another opportunity to hear the stimulus. At this point, all listeners were able to recognize the target. After listening to the familiarization stimuli, listeners were reminded that they would be listening to the same target voice throughout the experiment.

The order of the experimental blocks was the same for every listener in that each person received the four SNRs in descending order: HINT score +3 dB, HINT score +0 dB, HINT score -3 dB, HINT score -6 dB[9]. This was done to avoid floor and ceiling effects by pitting any task or talker

---

determine whether listeners are reporting words they have understood or mimicking sounds or partial words.

[9] For Mandarin listeners whose HINT scores were above +7 dB (n = 5), speech-in-babble testing was done at SNRs of +10, +7, +4, and +1. It was determined, on the basis of other experiments run in this laboratory, that easier SNRs would make the speech-in-babble task too easy too reveal differences in performance across the two babble languages. Analysis of the performance of these individuals showed that, overall, they performed similarly to the others and did not show worse performance as a result of this limitation on the normalization scheme.

learning effects against SNR difficulty and possible fatigue effects[10]. The same two sentence lists were presented in each block for each person (e.g. lists 1 and 5 were always the target lists in Block 1), but the language of the noise mixed with each list was counterbalanced.

Data analysis

Intelligibility scores were determined by a strict keyword-correct count. Keywords with added or deleted morphemes were counted as incorrect responses, but obvious spelling errors or homophones were considered correct.

Results

HINT results

An unpaired, one-tailed t-test confirmed that, as predicted, monolingual English listeners had significantly lower HINT thresholds than L2 listeners ($p < .0001$) ($t = -15.0031$, $df = 27.594$, $p < .0001$). The mean scores for the two groups differed by approximately 8 dB (English mean: -2.31; Mandarin mean: 5.66). They are shown in Figure 3.1 below.

[10] It should be noted, too, that the babble used in this experiment is “frozen babble” (i.e., there were just four different short babble samples that listeners heard for each language during the experiment). Felty et al. (2009) compared the use of frozen and randomly varying babble (samples taken from random time points in a long babble track) on a word recognition task and found that listeners had a steeper learning curve in the frozen babble condition. This finding suggests at least one type of perceptual learning that may have occurred over the course of the experiment but would have been countered by the increasingly difficult SNRs.

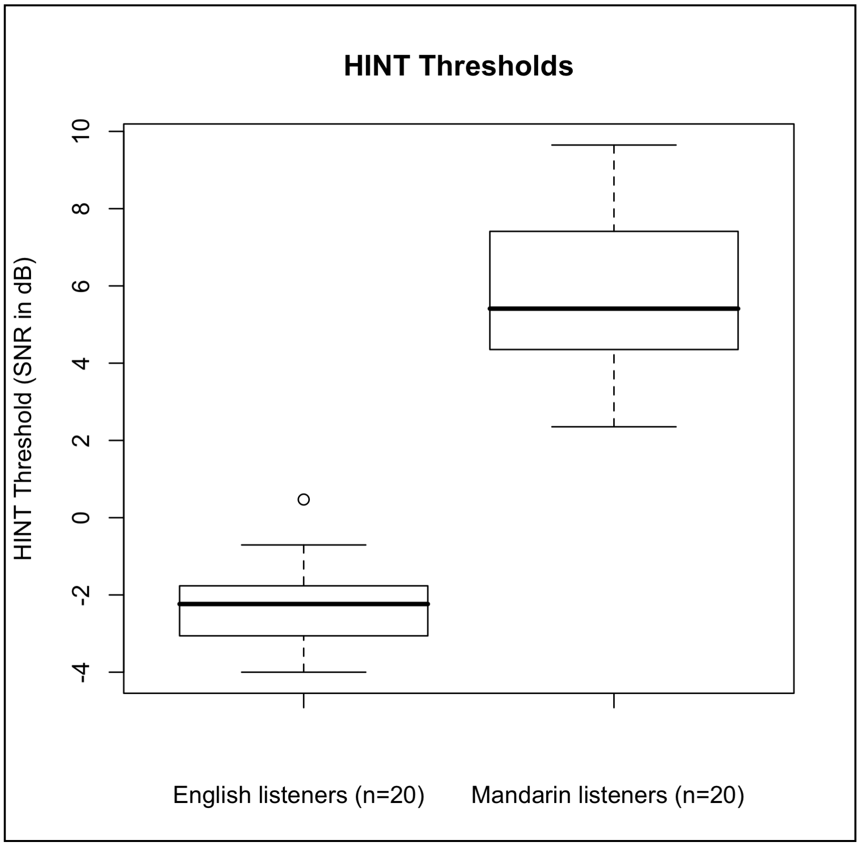

Figure 3.1. HINT threshold scores (the SNR at which a listener can identify whole sentences 50 percent of trials) for the native English listeners and the native Mandarin listeners. The center line on each boxplot denotes the median score, the edges of the box denote the 25th and 75th percentiles, and the whiskers extend to data points that lie within 1.5 times the interquartile range. Points outside this range appear as outliers.

These results replicate previous findings showing that native listeners outperform non-native listeners on speech perception tasks in energetic masking conditions (Hazan and Simpson, 2000; Bradlow and Bent, 2002; Van Wijngaarden et al., 2002; von Hapsburg et al., 2004; Garcia Lecumberri and Cooke, 2006; Rogers et al., 2006, and many others).

An investigation of the relationships between HINT scores and other measures of English experience/proficiency showed no significant correlations: age at which English acquisition began (r = .037, p = .438), years studying English (r = -.137, p = .282), and TOEFL scores (r = -.277, p = .125).

Speech-in-babble results

The mean percentage of keywords identified by the L1 listeners (monolingual English listeners) and the L2 listeners (L1 Mandarin listeners) in each noise language and at each SNR are shown in Figure 3.2 and given in Table 3.2 below.

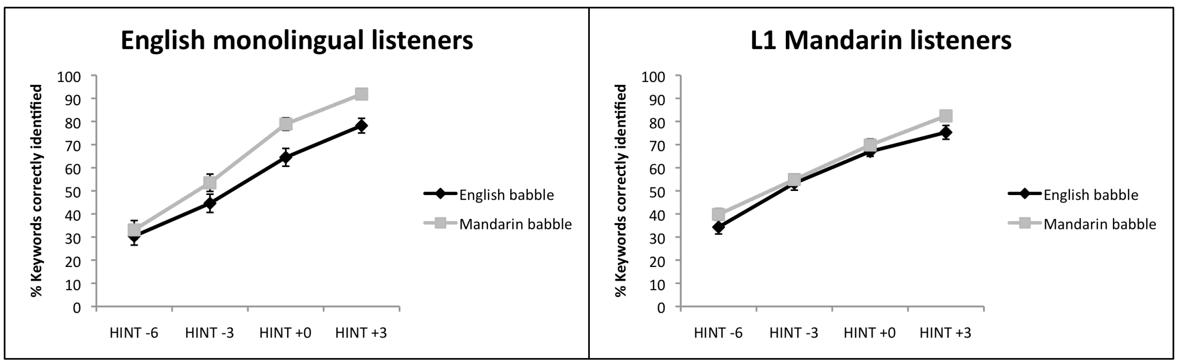

Figure 3.2. Mean intelligibility scores expressed as percentage of correct keyword identifications for native English listeners (left) and native Mandarin (L2 English) listeners (right). Error bars represent standard error.

Listeners		Eng+3	Man+3	Eng+0	Man+0	Eng-3	Man-3	Eng-6	Man-6
English	Mean	78.2	91.8	64.5	78.9	44.6	53.5	30.4	33.1
	(Std. Error)	(3.17)	(1.62)	(3.85)	(2.71)	(3.95)	(3.75)	(3.89)	(4.04)
Mandarin	Mean	75.3	82.4	67	69.8	53.3	54.9	34.3	39.8
	(Std. Error)	(2.98)	(1.89)	(2.08)	(2.62)	(3.00)	(2.31)	(2.97)	(2.51)

Table 3.2. Means and standard errors of English and Mandarin listeners' recognition scores in each noise condition (% keywords identified).

Keyword identification data were assessed statistically using mixed-effects logistic regression, with subjects as a random factor and native language, babble language, SNR, and all interactions among them as fixed effects. This analysis avoids spurious results that can arise when categorical data are analyzed as proportions using ANOVAs. It also has greater power than ANOVAs and does not assume homogeneity of variances (see Jaeger (2008) for discussion of the benefits of this analysis).[11] Analyses were performed using R, an open-source programming language/statistical analysis environment (R development core Team 2005). The results of the regression are shown in Table 3.3.

Predictor	Estimate (B)	Standard Error (SEB)	Odds ratio (e^B)
SNR	0.259***	0.011	1.295
Babble Language (Mandarin vs. English)	0.784***	0.065	2.190
Native Language (Mandarin vs. English)	0.048	0.177	1.049
Native Language * Babble Language	-0.544***	0.087	0.580
Native Language * SNR	-0.054***	0.015	0.947
Babble Language * SNR	0.115***	0.017	1.121
Native Language * Babble Language * SNR	-0.096***	0.023	0.908

Significance values: *p < .05; **p < .001; ***p < .0001

Table 3.3. Summary of logistic regression on probability of correct response including participant as random intercept (Overall intercept: 0.606; St.Dev. of participant intercepts: 0.529).

The results show that the overall probability of correct keyword identification is significantly higher as SNR increases ($z = 23.06$, $p < 0.0001$) and in Mandarin versus English babble ($z = 12.05$, $p < 0.0001$). The native language background of the listener was not a significant predictor of correct response ($z = 0.27$, $p = 0.79$), showing that the method of normalizing listeners according to their HINT scores succeeded in eliminating this factor as a predictor for performance on the speech-in-babble task.

[11] Note that the data were also converted to percentage correct scores, transformed using the rationalized arcsine transform (Studebaker, 1985), and analyzed using a traditional repeated measures ANOVA. The results were essentially the same.

Significant interactions with language background reveal that English listeners receive a greater release from masking in Mandarin babble than do Mandarin listeners. This is supported by the significant interaction between native language and babble language ($z = -6.29$, $p < 0.0001$), which shows that Mandarin listeners generally experienced more interference from Mandarin babble than did English listeners. This interaction is particularly strong at high SNRs, as revealed by the significant interaction between these three factors ($z = -4.18$, $p < 0.0001$).

To help visualize this three-way interaction, Table 3.4 reports the difference in accuracy across babble languages at each SNR for each listener group. These difference scores reveal the much larger noise language effect observed in the English listeners versus the Mandarin listeners (as shown by the significant two-way interaction) and show that this effect is considerably larger at the higher SNRs (as reflected in the three-way interaction). The confidence intervals also show English listeners performed better in Mandarin versus English babble in all SNRs except the most difficult (HINT -6 dB) and that Mandarin listeners performed significantly better in Mandarin versus English noise at the easiest and most difficult SNRs (as indicated by confidence intervals that do not extend beyond 0).

% keywords identified in Mandarin - English babble	HINT + 3dB	HINT + 0dB	HINT -3dB	HINT -6dB
English listeners	13.6 (7.3-20.9)	14.4 (9.1-19.7)	8.9 (2.8-15.0)	2.7 (-2.4-7.8)
Mandarin listeners	7.1 (1.4-12.8)	2.8 (-.5-6.1)	1.6 (-2.9-6.0)	5.5 (1.0-10.0)

Table 3.4. Mean differences with 95% confidence intervals for keywords identified in Mandarin - English babble (in percentage correct).

In order to further investigate the effects of noise language on the non-native listeners, a mixed-effects logistic regression was also performed on the data from the Mandarin listeners only. The regression showed a significant effect of SNR ($z = 29.94$, $p < .0001$) and showed that Mandarin listeners' performance was overall better in English versus Mandarin noise ($z = 11.26$, $p < .0001$). There was also a significant interaction between the two ($z = 5.29$, $p < .0001$).

The overall analysis also revealed two-way interactions of listener group with SNR and babble language with SNR. The steeper improvement for the English versus Mandarin listeners across SNRs is reflected by a significant two-way interaction of listener group and SNR ($z = -3.50$, $p < 0.0001$). The interaction of babble language and SNR ($z = 6.68$, $p < 0.0001$) reflects the overall greater difference between noise languages at easier SNRs.

In summary, the results show that performance for both listener groups increased on the speech-in-babble task as SNR increased, and performance was generally lower in English versus Mandarin babble. Interactions involving native language background reveal that native Mandarin listeners perform relatively worse in Mandarin noise as compared with monolingual English listeners.

Discussion

HINT thresholds

As expected, the HINT results showed that non-native listeners require a significantly more favorable SNR (by an average difference of about 8 dB) to identify English sentences in stationary, speech-shaped noise. This finding replicates previous findings that non-native listeners have more difficulty recognizing speech in noise than do native speakers (Hazan and Simpson, 2000; Bradlow and Bent, 2002; Van Wijngaarden et al., 2002; Garcia Lecumberri and

Cooke, 2006; Rogers et al., 2006 and many others). [12] Furthermore, this large and highly significant difference in scores on a standard clinical test points to the importance of taking language experience into account in the practice of audiology, and particularly in speech audiometry (von Hapsburg and Peña, 2002; von Hapsburg et al., 2004).

In addition to providing a test of our listener groups' tolerance for purely energetic masking in English sentence recognition, HINT thresholds also proved to be a useful tool for normalizing native and non-native listener performance on the speech-in-babble task. By selecting test SNRs relative to individual HINT scores, the two listener groups achieved similar performance levels on the task of English sentence recognition in 2-talker babble (as indicated by the lack of a significant effect for native language on the speech-in-speech test).

Sentence intelligibility in 2-talker babble

This study showed that, for native English speakers and L2 English speakers (L1 Mandarin), English babble was more disruptive overall to English sentence recognition than Mandarin babble. Crucially, however, it also showed that native English speakers receive a larger release from masking in Mandarin babble (a foreign language) relative to English babble than do native speakers of Mandarin. The greater overall interference from English versus Mandarin babble for both listener groups suggests that acoustic and/or linguistic similarity between the speech signal and the noise may be the most critical factor in driving noise language effects, and the greater relative interference from Mandarin babble for Mandarin-speaking listeners suggests that there is

[12] Note, of course, that the size of such differences are likely dependent (at least in part) on listeners' level of proficiency in English. The listeners in this study may be more or less proficient than listeners in other studies of non-native speech perception in noise.

also a component of informational masking that is specifically driven by the native-language status of the noise.

For the native listeners in this study, the speech-in-speech results replicate the previously observed effect of native-/same-as-target-language versus foreign-language 2-talker babble (Van Engen and Bradlow, 2007): English babble was found to be significantly more difficult than Mandarin babble for native English listeners. The replication of this finding with a new target talker shows that the effect cannot be attributed solely to the particular acoustic or stylistic characteristics of a single target talker's voice or its interaction with the babble tracks.

In this study, the release in masking experienced by native English listeners in Mandarin versus English babble was largest at the highest tested SNR and smallest at the lowest SNR. This pattern differs from Van Engen and Bradlow (2007), which found the language effect to be largest at the most difficult SNR tested (-5 dB). In the present study, however, the difficult SNR was significantly lower than -5 dB for most listeners (as low as -10 dB for the listeners with the lowest HINT thresholds). Therefore, it is likely that the higher amount of energetic masking at these lower SNRs eliminates observable linguistic informational masking effects.

While the difficulty in English versus Mandarin babble for native English listeners has primarily been considered in terms of linguistic informational masking effects, it must be noted (as pointed out by Mattys et al. 2009) that energetic masking differences between the noise languages may also exist. The English and Mandarin babble were controlled for SNR, but were not otherwise manipulated to equate, for example, long-term average speech spectrum or temporal modulation rates and depths[13]. This study avoided any signal processing that may have equated these signal-dependent factors in order to maintain the naturalness of the stimuli. This means that the two babble languages may possibly impose different amounts of energetic masking on the target sentences.

If energetic masking can account completely for the differences between English and Mandarin babble for native English listeners, then it is predicted that the effects of the two languages would be similar across listener groups. However, if the noise language effect is, indeed, driven at least in part by higher-level informational masking in the form of linguistic interference, then differential effects of noise languages on listener populations with different language experience are predicted. Furthermore, even if there are energetic masking differences across the two noise languages, differences in their relative effects on listener groups with different language experience could reveal linguistically-driven influences of informational masking. This was indeed what was observed in the present study: although their performance was lower in English babble than in Mandarin babble, native Mandarin listeners were more detrimentally affected by Mandarin babble relative to English babble than were monolingual English listeners.

With respect to non-native speech perception in noise, these results represent the first evidence that L2 babble may be more detrimental to L2 speech processing than L1 babble. As noted in the introduction, Garcia Lecumberri and Cooke (2006) did not find differential effects of L1 and L2 competing speech on L2 listeners, but several differences between these two studies may account

[13] In terms of the spectral properties of these particular maskers, running t-tests did reveal differences between the languages at some frequencies; in general, however, the long-term average spectra of the English and Mandarin babbles were highly similar.

for the different outcomes. First, the present study uses 2-talker babble, while Garcia Lecumberri and Cooke used single, competing talkers. 2-talker babble generally induces greater energetic masking, since the babble signal itself is more temporally dense than a single competing talker. It is possible that, by further reducing access to the signal, the additional energetic masking in 2-talker babble renders linguistic informational masking effects observable. It is possible that linguistic factors modulate the effects of speech noise on speech perception under relatively specific conditions. Van Engen and Bradlow (2007), for example, found no effect of babble language using a 6-talker babble.[14]

Noise types aside, another important distinction between these studies is that the speech perception tasks differed widely between them. This study measured L2 keyword identification in sentences, while Garcia Lecumberri and Cooke investigated L2 consonant identification. It is quite possible that the task of listening to sentence-length material is more susceptible to language-specific noise interference effects than is consonant identification. Sensitivity to linguistic interference from maskers may, for example, be greater when a fuller range of linguistic structures is being processed in the targets. For non-native listeners in particular, an accumulation of processing inefficiencies across levels of linguistic processing (Cutler et al.,

[14] In a detailed study of energetic and informational masking effects on speech segmentation biases, Mattys et al. (2009) also did not find differential effects of a single talker masker and an acoustically-matched modulated noise masker in a speech segmentation task (for native listeners). One of their suggestions for why language effects may emerge in 2-talker (but not in 1-talker babble) is that two talkers in an unintelligible language may cohere more readily for listeners, making segregation from the signal easier. This explanation may apply to the native listeners in this study, but for the non-native listeners, both maskers *were* intelligible. That said, their knowledge of the two languages is quite different (native and non-native), allowing, perhaps, for a tempered version of this explanation.

2004) may contribute to differential sensitivity to noise languages in sentence keyword identification versus consonant identification.

Along with the issue of L2 performance in L1 and L2 babble, one of the other open questions regarding the previous finding that native English listeners are more detrimentally affected by English versus Mandarin 2-talker babble was whether this effect is primarily driven by the native-language status of English or by its greater degree of acoustic and linguistic similarity to the English targets, which may lead to greater energetic and/or informational masking. The present results from the Mandarin listeners, for whom one of the two babble maskers is native (Mandarin) and the other matches the target (English) show that, at least for the non-native listeners, interference from a 2-talker masker in the target language (English) was greater than interference from the listeners' native language (Mandarin), at least at the easiest and most difficult SNRs that were tested. This finding suggests that signal similarity (a match between target and noise languages) is at least as important as native-language status (and perhaps more) in driving noise language effects in general.

While the finding that English babble induced more interference than Mandarin babble for both listener groups points to the importance of target-masker similarity in speech-in-speech masking, the interaction with native language status also crucially implicates a significant role for language experience in informational masking: while the native and non-native groups scored similarly in English babble, particularly at the easier SNRs, the native English listeners' performance was significantly better in Mandarin babble than the non-native listeners. That is, the native Mandarin listeners had relatively more trouble 'tuning out' Mandarin babble compared to the native English listeners.

In summary, this study of speech perception in noise by native and non-native listeners has shown that *both* similarity between the target and the noise (i.e., matched language) *and* the native-language status of noise for a particular listener group contribute significantly to the masking of sentences by 2-talker babble. Future studies comparing different types of noise (e.g., competing speech, non-speech noise that is filtered or modulated to match various speech maskers) will allow for further specification of the roles of energetic and informational masking in speech perception in noise by various listener groups. In addition, experiments using other target and noise languages and other listener groups will allow for further development of our understanding of the particular role of linguistic knowledge in speech-in-speech intelligibility. For example, the typological similarity between target and noise languages may modulate the degree of interference imposed by the babble, as may the availability of semantic content of the noise to listeners. Finally, studies to investigate the level of linguistic processing at which such effects emerge (phonetic, phonological, lexical, syntactic, semantic, prosodic, etc.) will allow for a fuller understanding of the processes involved in understanding speech in the presence of speech noise.

CHAPTER 4: SPEECH PERCEPTION IN SPEECH NOISE BY EARLY BILINGUAL LISTENERS

Introduction

The preceding study investigated speech intelligibility in the presence of 2-talker babble, focusing on the role of listeners' experience with the languages of the target and noise speech. Specifically, it compared native English listeners with native Mandarin listeners on an English speech recognition task in English and Mandarin 2-talker babble. Results showed a) that the non-native listeners required an easier SNR (on average, about 8 dB) in order to perform similarly to native speakers and b) that target language babble (English) interfered more with speech recognition than another language (Mandarin) for non-native listeners, even when the other language was the native language of the listener. Compared to the monolingual English listeners, however, the Mandarin speakers experienced a smaller release from masking when the noise was Mandarin babble as opposed to English babble.

This study, along with many others (Mayo et al., 1997; Hazan and Simpson, 2000; Bradlow and Bent, 2002; Van Wijngaarden et al., 2002; Cutler et al., 2004; Garcia Lecumberri and Cooke, 2006; Rogers et al., 2006), showed that *non-native* and/or *late bilingual* listeners have greater difficulty than native listeners on speech-in-noise tasks. The issue *of early bilingualism* and coping with speech in noise was examined recently by Rogers et al. (2006). In this study, the participants were Spanish/English bilinguals who had acquired English before age 6 and were judged to have no accent in English. On a monosyllabic word recognition task in speech-shaped noise and with reverberation, these bilinguals had more difficulty than the monolingual participants, even though their performance was the same in quiet. The current study follows up on this finding and extends the research presented in Chapter 3 by investigating early Spanish-English bilinguals' performance on the task of English sentence recognition in 2-talker babble in English and Mandarin. As shown in Van Engen and Bradlow (2007), Chapter 3 (Van Engen, 2010), and other studies of speech-in-speech perception, the 2-talker babble situation requires listeners to not only cope with signal degradation due to energetic masking, but also to cope with informational masking in the form of linguistic interference from the competing speech.

While Rogers et al. (2006) shows an apparent bilingual disadvantage for understanding speech in noise, there are also many cognitive abilities for which early bilinguals appear to have an advantage over monolinguals. These include problem solving and creativity (Kessler and Quinn, 1980; Kessler and Quinn, 1987) and tasks involving memory and/or inhibition of attention (Ransdell et al., 2001; Kormi-Nouri et al., 2003; Bialystok et al., 2004; Bialystok and Martin, 2004). Bialystok and Martin (2004), for example, showed in a card-sorting task that bilingual children had better inhibitory control for ignoring perceptual information than monolinguals. Similarly, Bialystok et al. (2004) showed that bilingual adults also carry out controlled processing more effectively than monolinguals, and further, that bilingualism appears to offset age-related losses in certain executive processes.

The findings of Rogers et al. (2006) and Bialystok and colleagues (2004), while not in contradiction to one another, do present competing hypotheses with respect to bilingual performance on speech perception in *speech* noise. According to Rogers et al., bilinguals are disadvantaged on speech-in-noise tasks as a result of increased demand for attentional resources and/or increased processing demand. This may be due to the bilinguals' need to deactivate the inactive language, to select target phonemes from a larger number of alternatives, or to match native speaker productions to perceptual categories that may be intermediate between the norms for their two languages. Such factors predict a bilingual disadvantage on speech-in-speech noise tasks as well. On the other hand, bilingual advantages in inhibitory and/or controlled processing as observed in the work of Bialystok et al. predict a bilingual advantage for speech understanding in speech noise. That is, speech in *speech* noise crucially requires ignoring irrelevant information while focusing on target information, an ability that appears to be enhanced in bilinguals. The importance of executive control for speech-in-speech processing has been shown, for example, by Tun et al. (2002), who found that cognitive abilities such as executive control contribute heavily to older adults' ability to process speech in the presence of speech noise.

This study follows the methodology used for the study of non-native speech-in-speech perception (Chapter 3) and addresses the competing predictions discussed above by investigating early bilingual speech perception in 2-talker babble. The same materials and methods were used as in Chapter 3, but the participants were early bilingual speakers of Spanish and English. Their performance is compared to that of the monolingual native English participants from Chapter 3.

Methods

Participants

19 listeners between the ages of 18 and 27 (average = 20.7) participated in this study. They were recruited using fliers posted on the Northwestern University campus and were paid for their participation. As in the studies above, all protocols and recruitment materials were approved by the IRB at Northwestern University. All listeners had begun speaking both Spanish and English

by age 8 and considered themselves bilingual speakers of the two languages. All speakers reported normal speech and hearing. 16 of the 19 listeners provided additional information regarding their learning and use of Spanish and English. The average age at which they started learning Spanish was 0.38 years (SD = 0.81), and the average age at which they started learning English was 2.88 years (SD = 2.45). Although 88% reported Spanish as the first language they had acquired, 94% reported English as their dominant language. In terms of percentage of current language usage, English was reported to be used an average of 76.23% of the time, and Spanish 20.31% of the time.

Materials

The same speech and noise materials from Chapter 3 were used for the bilingual listeners as were used for the native English and native Mandarin speakers in Chapter 3.

Procedure

The procedure was identical to the procedure in Chapter 3.

Results

HINT results

HINT SNR thresholds for the monolingual listeners and the bilingual listeners are shown in Figure 4.1 below. The average threshold for the monolingual listeners was -2.3 dB, and for the bilingual listeners was -1.9 dB. A one-tailed, unpaired t-test was conducted to determine whether the bilingual listeners indeed had significantly higher HINT thresholds than monolingual listeners, but the difference did not reach significance ($p = .0969$).

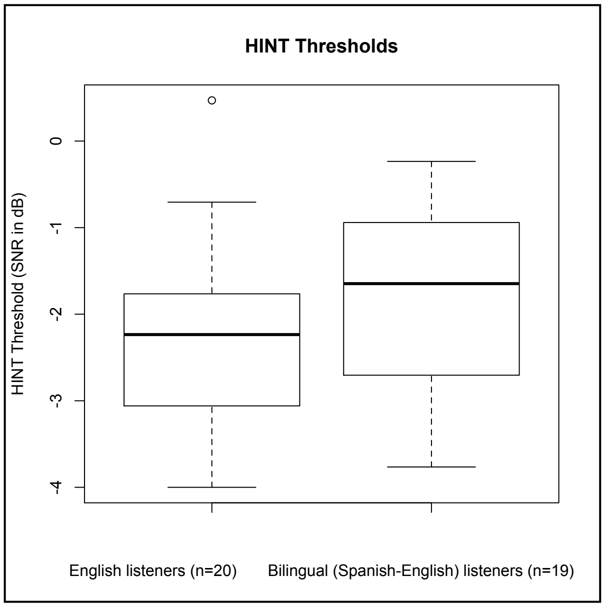

Figure 4.1. Boxplots showing HINT SNR-50 thresholds for monolingual English listeners and bilingual speakers of English and Spanish. In each boxplot, the black line represents the median value, the box extends to the 25th and 75th percentile, and whiskers extend to values within 2.5 times the interquartile range.

Speech-in-babble results

The bilingual listeners' performance on the speech-in-speech task is shown next to the monolinguals' performance in Figure 4.2 below, with scores presented as the percentage of keywords accurately identified in each condition.

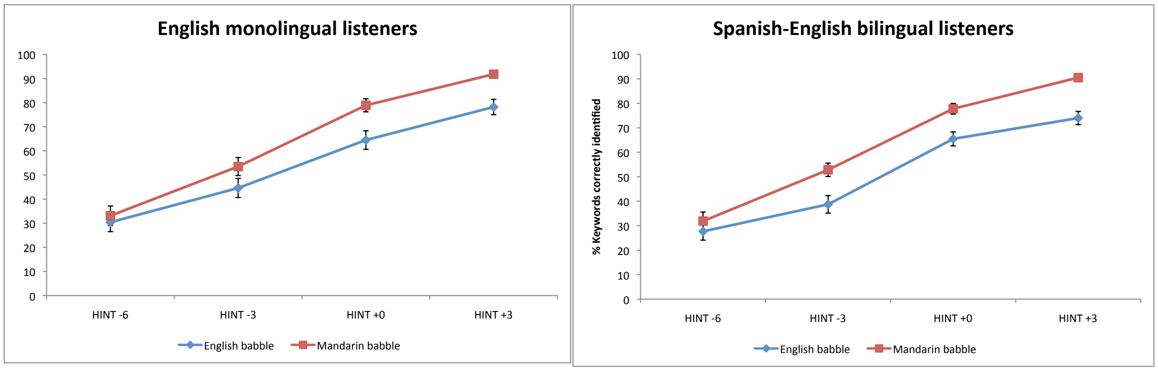

Figure 4.2. Percentage of keywords identified by Monolingual English listeners and Spanish-English bilingual listeners in English and Mandarin 2-talker babble at four SNRs. Error bars represent standard error.

The proportion of keywords identified by each subject in each condition (Mandarin and English babble at SNRs of HINT +3, HINT +0, HINT -3, and HINT -6) was converted to rationalized arcsine units (RAU) (Studebaker, 1985) for statistical analysis. This transformation "stretches" the upper and lower ends of the scale, thereby allowing for valid comparisons across the entire range of the scale. Scores on this scale range from -23 (0%) to 123 (100%) RAU.

Listeners' performance was analyzed using a three-way repeated measures analysis of variance (ANOVA) with language background (monolingual vs. bilingual) as a between-subjects factor and SNR and masker language (Mandarin vs. English) as within-subjects factors[15]. The ANOVA shows significant main effects for SNR [$F(3, 111) = 399.066$, $p < .0001$] and masker language [$F(1, 37) = 77.668$, $p < .0001$], as well as a significant interaction between SNR and masker language [$F(3,111) = 11.161$, $p < .0001$] such that the language effect was greater at higher

---

[15] The use of ANOVA in this study represents an early approach to data analysis within this project. Later analyses (as in Chapters 3 and 5) use mixed effects logistic regressions, which have several advantages over ANOVA for analyzing correct vs. incorrect keyword identification data. In future work, these data may be re-analyzed accordingly.

SNRs (as seen in Chapter 3). Crucially, there was no significant main effect or interaction involving listeners' language background.

Discussion

HINT thresholds

While the average bilingual HINT threshold was slightly higher than the average monolingual threshold, the difference in these groups was not significant. This finding differs from the results reported in von Hapsburg et al. (2004), in which bilingual participants did show poorer speech reception thresholds in noise in the HINT test. Those listeners, however, were born in Latin America and did not begin learning English until after age 10, whereas the participants in this study were born in the United States and started learning English no later than age 8 (and often, several years earlier). The combination of these results and those of von Hapsburg et al. suggest that HINT thresholds have the potential to contribute to a broad assessment of language proficiency in bilingual individuals. Future research in this area would involve a comparison of HINT thresholds with other measures of language proficiency, as well as comparing listeners' performance on the HINT in each of her/his languages.

Sentence intelligibility in 2-talker babble

The bilingual listeners did not differ from the monolingual listeners on the task of sentence recognition in English and Mandarin 2-talker babble. Their overall performance was similar, and the relative effects of the two noise languages were similar for the two listener groups. Given the lack of a significant result, there is little that can be concluded from the present experiment. It is possible that the combination of increased attentional and processing demand (which would suppress bilingual performance in 2-talker babble) and enhanced inhibitory/controlled processing (which would enhance bilingual performance in 2-talker babble) leads to similar bilingual and monolingual performance on this task. It is also possible that the task of sentence recognition in 2-talker babble is not, in fact, sensitive to either of these previously-observed differences between monolingual and bilingual individuals. Additional research is required to determine whether there are aspects of speech-in-speech processing that differ between monolingual and early bilingual listeners.

CHAPTER 5: SPEECH-IN-SPEECH RECOGNITION: A TRAINING STUDY

Abstract

This study aims to identify aspects of speech-in-noise recognition that are susceptible to training, focusing on whether listeners can learn to "tune in" to target talkers and "tune out" various maskers after short-term training. Listeners received training on English sentence recognition in speech-shaped noise, Mandarin babble, or English babble. The training signal-to-noise ratio (SNR) was customized for each trainee based on individual Hearing in Noise Test (HINT) thresholds (i.e., tolerance for speech-shaped noise). Results from a speech-in-babble post-test showed evidence of both tuning in and tuning out: listeners were able to take advantage of target talker familiarity; training with babble was more effective than speech-shaped noise training; and after babble training, listeners improved most in coping with the babble language in which they were trained. Additionally, working memory capacity and SNR were significant predictors of post-test success. Since SNRs for training and testing were selected relative to individual HINT thresholds, the SNR effect suggests a dissociation between individuals' tolerance for energetic and informational masking. In general, the results show that processes related both to tuning in to speech targets and tuning out speech maskers are involved in speech-in-speech recognition and can be improved with auditory training.

Introduction

Speech communication rarely takes place in quiet environments. Instead, listeners must extract linguistic information from speech signals that are degraded by noise and/or where the informational content of noise detracts from accurate perception of target speech. While such environments render the task of speech perception more difficult for all listeners, individuals with hearing loss experience particular difficulty in noisy environments (e.g., Plomp and Mimpen, 1979; Smoorenburg, 1992; Killion and Niquette, 2000), as do listeners who are communicating in a language other than their native language (e.g., Mayo et al., 1997; Van Wijngaarden et al., 2002; Rogers et al., 2006, Van Engen, 2010) and individuals with central auditory processing disorders (Bamiou et al., 2001), learning disabilities (Hugdahl et al., 1998; King et al., 2003), specific language impairment (Wright et al., 1997), and attention-deficit/hyperactivity disorder (Chermak et al., 1998). These populations, as well as others who would profit from honing their ability to understand speech in noise, could potentially benefit from auditory training. One commercially-available training program – Listening and Communication Enhancement (LACE) – has indeed shown improvements in speech-in-noise performance by hearing aid users after training (Sweetow and Henderson Sabes, 2006). This program employs a variety of training tasks, including degraded speech tasks (speech-in-babble, time-compressed speech, speech-in-competing speech), cognitive tasks, linguistic tasks, and overt instruction on communication strategies. The success of such a program is highly promising, but relatively little research has been conducted to identify the particular speech and noise conditions that can best be utilized in training to improve real-world speech-in-noise hearing.

The goal of the present study, therefore, is to identify aspects of speech-in-noise processing that are susceptible to improvement through training. In particular, this study examines whether short-term training can improve listeners' ability to cope with, or "tune out", particular types of noise (i.e., maskers) and whether they can learn to better perceive, or "tune in", to specific target voices in various noise conditions. Given the ubiquity of interfering speech noise in everyday communication, this study focuses on identifying training parameters that can help listeners perceive speech in the presence of interfering speech (specifically, 2-talker babble). By identifying the aspects of speech-in-noise processing that are most susceptible to training for adults with normal speech, language, and hearing abilities, this research will provide insight into the most efficacious approaches to speech-in-noise training for a wider range of listener populations.

Speech noise (i.e., a single competing voice or multiple-talker babble) interferes with the perception of target speech both by physically degrading the target signal and by contributing acoustic and linguistic information that may distract a listener and/or impede separation of the target and noise signals. These two types of interference are typically understood as cases of energetic and informational masking, respectively. That is, noise imposes energetic masking on a target signal when spectral and temporal overlap between the target and noise causes mechanical interference in the auditory periphery such that components of the target signal are rendered inaudible to the listener. Energetic masking, therefore, results in a reduction of the available acoustic and linguistic cues relevant to speech understanding. Informational masking, by contrast, generally refers to any reduction in target signal intelligibility that cannot be explained by energetic masking. Although the precise definition of informational masking is still under discussion (see, for example, Durlach, 2006 and Kidd et al., 2007), the term is used here in this broad sense (i.e., non-energetic masking) to draw the important distinction between interference that occurs in the auditory periphery and interference that occurs at higher levels of auditory and cognitive processing during speech-in-speech listening. See Kidd et al. (2007) for an in-depth history and overview of masking terminology in hearing science.

Since noise can interfere with target speech intelligibility both in the auditory periphery and at higher levels of processing, a variety of linguistic and non-linguistic features of speech noise can modulate the intelligibility of target speech. Relevant non-linguistic features include the level of the noise relative to the target level (signal-to-noise ratio, or SNR) (e.g., Brungart et al., 2001), the spatial location of target and noise sources (e.g., Freyman et al., 2001), the number of competing talkers (e.g., Simpson and Cooke, 2005), and physiologically-driven characteristics of interfering voices such as fundamental frequency (e.g., Brungart et al., 2001). In addition to such non-linguistic factors, there is also evidence that the *linguistic* content of noise can modulate target speech intelligibility (Sperry et al., 1997; Arbogast et al., 2005; Simpson and Cooke, 2005). Sperry et al. (1997) found, for example, that listeners had more difficulty on a word recognition task in the presence of a multi-talker masker compared with a reversed version of that masker and with matched amplitude-modulated speech-shaped noise. Similarly, in a study of consonant identification in multi-talker babble and amplitude-modulated noise that was generated to match the spectral and temporal properties of the babble (i.e., “babble-modulated noise”), Simpson and Cooke (2005) showed lower performance in real babble compared to matched noise when there were more than two talkers in the babble.

In speech-in-speech scenarios, several studies have shown, further, that the *language* spoken in the noise significantly affects target speech intelligibility: Van Engen and Bradlow (2007) and Van Engen (2010) showed that native English listeners have greater difficulty understanding keywords in English target sentences in the presence of English 2-talker babble versus Mandarin 2-talker babble. In a similar task, Calandruccio et al. (in press) showed that native English listeners have more difficulty in English versus Croatian 2-talker babble, and Garcia Lecumberri and Cooke (2006) showed poorer consonant identification by native English listeners in the presence of a competing talker in English versus Spanish.

In addition to increased interference from native- versus foreign-language noise, a recent study has provided further evidence for particularly linguistic interference from noise by showing that the semantic content of speech noise can modulate the intelligibility of target speech (Brouwer et al., 2010). In a comparison of the effects of 2-talker babble composed of semantically anomalous and semantically normal sentences produced by the same two talkers, native English listeners had greater difficulty coping with semantically normal noise during sentence recognition.

Given that listeners' language experience and the linguistic content of speech targets and noise can affect target speech intelligibility in speech-in-speech perception, it is hypothesized that the ability to tune in to target speech and/or to tune out speech noise is subject to experience-related modification, and as such, that explicit training can improve speech-in-speech understanding. This general hypothesis is supported by a wide range of behavioral and neuroscientific research showing evidence of adult perceptual learning for speech signals. Several studies, for example, have shown evidence of talker-specific learning, in which the accuracy of spoken word recognition increases as listeners become familiar with individual talkers (Mullenix et al., 1989; Nygaard et al., 1994; Nygaard and Pisoni, 1998; Bradlow and Pisoni, 1999; Bradlow and Bent, 2008). Further, talker-independent adaptation has been documented for foreign-accented speech

(Weil, 2001; Clarke and Garrett, 2004; Bradlow and Bent, 2008; Sidaras et al., 2009), speech produced by individuals with hearing impairment (McGarr, 1983), and computer-synthesized speech (Schwab et al., 1985; Greenspan et al., 1988; Francis et al., 2007). Finally, adaptation has been shown for speech signals that have been distorted by time compression (Dupoux and Green, 1997; Pallier et al., 1998), noise-vocoding (as in cochlear implant simulations) (Davis et al., 2005; Hervais-Adelman et al., 2008; Loebach and Pisoni, 2008; Bent et al., 2009), and the presence of noise (Bent et al., 2009). In addition to such behavioral evidence for adaptation to variability in speech signals, neuroscientific studies of plasticity in the central auditory system have shown that short-term auditory training with speech stimuli can lead to changes in cortical responses (Kraus et al., 1995; Tremblay et al., 1997; Tremblay and Kraus, 2002; Wong et al., 2007; Wong and Perrachione, 2007) and subcortical responses (Song et al., 2008). With respect to speech-in-noise training in particular, a recent study has shown, further, that brainstem responses to speech signals *in noise* can be enhanced with short-term training (Song et al., in preparation). Taken together, this body of research shows remarkable flexibility in the speech perception system and provides support for the potential efficacy of training listeners to better understand speech in noise.

Preliminary evidence for a speech-in-speech training effect for normal-hearing listeners was shown in Van Engen and Bradlow (2007). In that study, listeners performed a sentence recognition task in English and Mandarin babble at two SNRs. Testing always began with the easier SNR and ended with the harder SNR. For half of the listeners, the testing took place at +5 dB and 0 dB; for the other half, testing was at 0 dB and -5 dB. A comparison of the two groups' performance in the 0 dB condition revealed significantly higher performance by the group tested at that SNR in the *second* half of the experiment, suggesting that improvement on a speech-in-speech task can take place after a relatively short amount of exposure. What is not clear, however, is whether these listeners learned to "tune in" to the target talker, to "tune out" the noise they were exposed to in the first half of the experiment, or both.

The present study, therefore, investigates talker and noise effects in a two-day speech-in-noise training experiment. Training involved presenting English target sentences in one of three different noise conditions: speech-shaped noise, Mandarin 2-talker babble, and English 2-talker babble. A control group received no training. The effectiveness of these training conditions on speech-in-speech recognition was then compared by assessing listeners' performance on a common post-test that presented English target sentences in English and Mandarin 2-talker babble. The post-test further addressed listener adaptation to a target talker in noise by including target sentences spoken by both a talker from the training and by a novel talker.

The three noise conditions—speech-shaped noise, Mandarin babble, and English babble—were selected for training in order to investigate listeners' ability to learn to cope with different types of masking and to compare the effects of any such learning on speech-in-speech performance in the common post-test. Speech-shaped noise, which is produced by filtering white noise using the long-term average spectrum of speech, is temporally and spectrally static, and therefore imposes only energetic masking on speech targets. Furthermore, it imposes *more* energetic masking on target speech than 2-talker babble, since babble contains temporal and spectral "dips" that give listeners the opportunity to hear components of the target speech at more favorable SNRs (Peters et al., 1998). For the purposes of comparing training maskers, the 2-talker babbles impose less energetic masking than speech-shaped noise masker, but *more* informational masking, since babble is a dynamic signal containing linguistic information. Studies that have directly compared babble to spectrally and temporally matched non-speech noise have indeed shown greater overall masking from real babble on speech identification tasks (e.g., Sperry et al., 1997; Simpson and Cooke, 2005), providing evidence for informational masking imposed by real speech signals. It is possible, further, that the addition of informational masking in 2-talker babble may even make babble a more effective masker overall than static speech-shaped noise for the task of sentence recognition, despite the reduced energetic masking imposed by babble versus static speech-shaped noise. Helfer and Freyman (2009), for example, showed better keyword identification in speech-shaped noise compared to two competing talkers.

With respect to the two babble languages, English is expected to be a more effective masker than Mandarin, presumably because it imposes more informational masking for native English listeners attending to English speech targets (Van Engen and Bradlow, 2007; Van Engen, 2010). Van Engen and Bradlow (2007) compared the effects of 2- and 6-talker babble in English and Mandarin at several SNRs. Sentence intelligibility was better in 2-talker babble overall, most likely because 6-talker babble, being spectrally and temporally more dense, imposes more energetic masking on the target speech. A significant effect of noise language was observed in the 2-talker babble only, where performance was lower in the English masker than the Mandarin masker. Presumably, the additional energetic masking in the 6-talker babble eliminated the differential language effects that were observable with 2-talker babble. These language effects are likely to be a form of informational masking, though energetic masking differences may also exist between the English and Mandarin 2-talker babble. Van Engen and Bradlow (2007) addressed the possibility of energetic differences first by analyzing the long-term average spectra of the babble tracks. There were several frequencies at which the two languages differed significantly, but these differences were small and not consistent across the frequency spectrum. (More generally, Byrne et al. (1994) showed that languages do not differ significantly in long-term average spectra averaged over talkers.) The overall spectral similarity between the two languages suggests that spectral differences, at least, are not the sole source of the language effect. More recently, the relative effects of English and Mandarin maskers have been shown to differ across listener populations with different experiences with the two languages (Van Engen, 2010): English babble was more disruptive than Mandarin babble for English sentence recognition for native speakers of English and native speakers of Mandarin, but there was a significantly greater difference between the effects of the two maskers for the English listeners. It is likely, therefore, that difference in the effects of these two maskers are due, at least in part, to differences in informational masking in the form of linguistic interference.

In addition to investigating whether listeners can learn to tune out particular noise types, the current training study is also concerned with whether they can improve their ability to tune in to target talkers in various noise conditions. As discussed above, listeners are able to adapt to the speech of individual talkers (e.g., Mullenix et al., 1989; Nygaard et al., 1994; Nygaard and Pisoni, 1998; Bradlow and Pisoni, 1999; Bradlow and Bent, 2008), and Mullenix et al. (1989) and Bradlow and Bent (2008) showed such adaptation following exposure to talkers in the presence of noise. Mullenix et al. found that listeners performed better on a word recognition task in (presumably white) noise when the target words were spoken by a single talker versus multiple talkers. Bradlow and Bent showed adaptation to an individual speaker of Mandarinaccented English following exposure to the talker's speech in the presence of speech-shaped noise. More recently, Bent et al. (2009) also showed adaptation to talkers in the presence of 6-talker babble. Such findings suggest that listeners are able to learn to better understand voices they are exposed to during speech-in-noise training. In the present study, we investigate whether such adaptation takes place in the presence of both speech-shaped noise and 2-talker babble, which imposes more informational masking than the types of noise employed in previous studies of talker adaptation. This study will show, further, whether adaptation to target talkers' speech in these noise conditions may help listeners segregate the talkers' speech from babble during the post-test, thereby improving their ability to tune in.

In addition to providing evidence for listener adaptation to a wide range of speech signal types, perceptual learning studies have also shown that the use of multiple talkers in training can be beneficial to learning. For example, Bradlow and Bent's (2008) comparison of various training conditions for promoting adaptation to foreign-accented English showed that training with multiple talkers of a given accent facilitated the intelligibility of a test talker just as much as training on only that talker. Multiple-talker training has also been shown to be particularly effective for generalized learning in studies of non-native phoneme contrast perception (e.g., Logan et al., 1991; Lively et al., 1993), lexical tone (Wang et al., 1999; Wang et al., 2003), and dialect classification (Clopper and Pisoni, 2007). In light of this research, the current training study utilizes multiple talkers (n = 4) during training. In the post-test, intelligibility is assessed for a familiar talker (one from the training) and a new talker.

It is hypothesized that training will improve performance on the speech-in-speech post-test in general, so that listeners who receive training will outperform those who do not. If the group trained in speech-shaped noise outperforms the babble-trained groups, it would suggest that training listeners to process speech with limited/partial acoustic cues (i.e., in high energetic masking environments) best prepares them to deal with environments with less energetic masking (even if there is greater informational masking in those situations). If the babble groups perform better than the speech-shaped noise group, then, presumably, training with informational masking is crucial for coping with informational masking environments.

It is expected that overall test performance in Mandarin will be higher than in English as observed in previous studies. Importantly, the comparison of performance in English and Mandarin babble by individuals trained in each language will provide insight into the specificity of listeners' ability to learn to tune out these types of noise, as well as whether differences in the maskers' effects on naïve performance can be mitigated by training. If such learning is language-specific, then it is predicted that listeners who are trained in English will perform better than other groups in English while listeners trained in Mandarin will perform better than others in Mandarin. The finding that one or the other training language is more efficacious overall would suggest that training with either higher (English) or lower (Mandarin) amounts of informational masking in the form of linguistic interference allows listeners to develop more generalizable listening strategies. An overall benefit of Mandarin training would suggest that such generalization proceeds from easier to harder informational masking environments, while an overall benefit of English would suggest that generalization proceeds from harder to easier informational masking conditions.

With respect to learning to tune in to particular talkers, it is expected that listeners who receive training will be able to take advantage of the familiarity of a training talker's voice, performing better on that voice in the post-test than on a novel voice. If baseline intelligibility of the two talkers is similar, then the untrained group should show no such difference. Further, it is possible that noise and talker effects will interact significantly, which may indicate that certain noise conditions provide listeners more opportunity to adapt to talkers than others. For example, in speech-shaped noise, where there is no competing speech, listeners may have greater opportunity to tune in to target talkers. If so, the speech-shaped noise-trained group would show a greater effect of talker familiarity than the other training groups.

Assessing performance *during* training (in addition to comparing post-test performance) will illuminate which noise conditions listeners are most able to learn to cope with through training. Since informational masking involves interference of processing at all levels beyond the auditory periphery, it is predicted that coping with informational masking will be more susceptible to improvement than coping with energetic masking only. If so, a greater effect of training will be observed between days 1 and 2 for the babble groups than for the speech-shaped noise group. This hypothesis is supported by data from LACE, which show speech intelligibility improvement from pre- to post-training on a speech-in-babble test (the QuickSin test by Etymotic Research, Inc.), but not on speech-in-speech-shaped noise (the HINT test by House Ear) (Sweetow and Henderson Sabes, 2006).

Experiment: multiple-talker training

This experiment investigates the effectiveness of training people to listen to English target sentences in different types of noise. Participants received two training sessions of approximately 30 minutes each, in which English target sentences spoken by four different target talkers were embedded in English 2-talker babble, Mandarin 2-talker babble, or speech-shaped noise. They were asked to repeat the target sentences orally to an experimenter, and were shown the text of each sentence after they responded. All participants were given a post-test in which they heard English sentence targets spoken by a talker from the training and by a new target talker. These targets were embedded in English 2-talker babble (half of the trials for each talker) and Mandarin 2-talker babble (half of the trials for each talker). A control group was given the post-test only. During the post-test, participants were also asked to repeat the target sentences to an experimenter, but they received no feedback.

Methods

Participants

Eighty-four young adults, ranging in age from 18 to 34, participated in the experiment and were either paid or received course credit for participation. Only monolingual, native speakers of American English who reported no speech or hearing problems were included in the analysis. Based on these criteria, 16 participants were excluded. (The Northwestern University Linguistics Department participant pool includes a large number of multilingual students.) 12 additional participants were excluded from analysis due to technical issues, illness, ear problems, or because they did not come to the second day of training. A total of 56 participants, then, were included in the analysis. Each participant was randomly assigned to one of four training groups: English babble training, Mandarin babble training, speech-shaped noise training, or no training.

Materials

Pre-testing

In addition to speech-in-noise training sessions and the speech-in-speech post-test, each participant also performed several preliminary tasks:

1. A speech, language and hearing background questionnaire, administered on a laboratory computer.
2. The letter-number sequencing task from the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV, 2008) – a measure of working memory. In this task, individuals hear a series of letters and numbers and must repeat them, re-ordering the items so that the numbers are in numerical order and the letters are in alphabetical order. Because working memory has been shown to correlate with speech perception in noise measures (e.g., Lunner, 2003; Parbery-Clark et al., 2009), this score was collected to assess the effect of working memory on the present speech-in-speech tasks.
3. The Hearing in Noise Test (HINT) (Nilsson et al., 1994). Using an adaptive presentation, the HINT test estimates the SNR at which a listener can achieve an accuracy score of 50% correct (whole sentences) in speech-shaped noise (SNR-50). Details regarding the test and its materials can be found in Nilsson et al. (1994). A 20-sentence version of the test was used (HINT lists 1 and 2).

Target sentences

Target sentences were taken from the Revised Bamford-Kowal-Bench (BKB) Standard Sentence Test (lists 1, 4 6, 8, 11, 14, 19, and 21 for training; lists 7, 9, 10, and 15 for the post-test). These particular lists were selected based on their approximately equivalent intelligibility scores for normal hearing children as reported in Bamford and Wilson (1979). Each list contains 16 simple, meaningful English sentences and a total of 50 keywords (3-4 per sentence) for intelligibility scoring. Since the HINT test was developed from the original set of BKB sentences, some sentences appear in both the HINT and the BKB lists. In order to eliminate any overlap between the two tests, sentences from BKB lists 2 and 20 were used to replace items in the training/test lists that also appeared in HINT lists 1 and 2. Replacement sentences were selected to match the number of keywords and the basic sentence structure of the items they replaced. This amounted to a total of 10 replacements.

Five adult female native speakers of American English produced the full set of BKB sentences. They were instructed to speak in a natural, conversational style, as if they were speaking to someone familiar with their voice and speech articulation patterns. Recording took place in a sound-attenuated booth in the Phonetics Laboratory at Northwestern University. Sentences appeared one at a time on a computer screen, and the speakers read them aloud, using a keystroke to advance from sentence to sentence. They spoke into a Shure SM81 Condenser microphone, and their speech was recorded directly to disk using a MOTU Ultralight external audio interface. The recordings were digitized at a sampling rate of 22050 Hz with 24 bit accuracy. The sentences were then separated into individual files using Trigger Wave Convertor, an automatic audio segmentation utility developed in the Department of Linguistics at

Northwestern University. The resultant files were trimmed to remove silence on the ends of the sentence recordings, and then the RMS amplitudes of all sentences were equalized using Praat (Broersma and Weenink, 2009).

Noise

Two-talker babble was created in both English and Mandarin using the following procedure: two female native speakers of each language were recorded in a sound-attenuated booth reading a set of 20 semantically anomalous sentences (e.g. *Your tedious beacon lifted our cab.*). These sentences were created in English by Smiljanic and Bradlow (2005) and translated into Mandarin by a native Mandarin speaker. The sentence recordings (initially prepared for an earlier study, Van Engen and Bradlow, 2007) were separated into individual files and silent portions were removed from the ends of each file. All sentence files were then equated for RMS amplitude. For each talker, the full set of sentences was concatenated to create a continuous string of speech. The sentences were concatenated in different orders for the two talkers so that no sentence would be spoken simultaneously at any point in the completed babble track. Any difference in duration between the two talkers' strings was eliminated by trimming the longer track. The completed speech strings were again RMS-equalized, and the two talkers' strings were mixed using Audacity (Audacity Team, 2006). Finally, the English babble, Mandarin babble, and speech-shaped noise track were RMS-equalized.

Mixing target sentences and noise

Targets and noise were mixed in real time and presented to listeners through custom software created using Max/MSP (Cycling '74, 2005) running on a Macintosh computer. Each stimulus began with 400ms of silence followed by 500ms of noise (English babble, Mandarin babble, or speech-shaped noise), the noise and the target together, and then a 500ms noise trailer. On each trial, a random portion of the noise track was selected. The RMS level of the target sentences was held constant at 65 dB SPL and the maskers varied in level relative to the target speech in order to produce the desired SNRs. Using a highly similar sentence recognition task, Van Engen (2007) compared this approach to SNR manipulation with a method in which the mixed signal and noise stimuli were re-leveled to equate overall stimulus output levels. These two methods produced similar results.

Procedure

Pre-testing

Upon entering the laboratory, participants provided written informed consent, completed a questionnaire about their language, speech, and hearing background, and performed the Letter-Number Sequencing task. They were then seated inside a sound-attenuated booth in the Phonetics Laboratory of the Linguistics Department at Northwestern University. After a brief otoscopic evaluation, the HINT test was administered. A MOTU Ultralight external audio interface was used for digital-to-analog conversion (24 bit), and signals were passed through a Behringer Powerplay Pro XL headphone amplifier. Signals were delivered to the ear canal using ER-1 insert earphones (Etymotic)—an assembly similar to what is used in clinical audiometry. Both the target sentences and the noise were presented diotically.

Training

All speech-in-noise training and testing took place in a sound-attenuated booth. Listeners were seated in front of a computer monitor. The experimenter was inside the booth, seated at a computer that was positioned so that the listener could not see the experimenter's monitor. The equipment described above for the HINT was also used for signal delivery during training and post-test. All signals and noise were presented diotically for training and testing. The experiment was run using custom software created using Max/MSP (Cycling '74, 2005).

Speech-in-noise training took place over two days. On each day, listeners listened to four lists of BKB sentences (64 sentences; 200 keywords) mixed with a single type of noise (either English babble, Mandarin babble, or speech-shaped noise). Within each BKB list, the sentences were produced by four different female talkers in a repeated sequence, such that listeners heard ¼ of the sentences by each of the four talkers but never heard two sentences in a row from any given talker. Female voices were used for all targets and noise in order to eliminate the variable of gender differences in speech-in-speech intelligibility (e.g., Brungart et al., 2001 showed that target speech intelligibility is better when target and masker voices differ in gender). Talker-sentence pairings and the order of BKB list presentation were counter-balanced across subjects. For each trial, one BKB sentence was played with a randomly selected portion of the relevant noise track. Listeners were instructed that the noise would begin a half a second before the target speaker. They were asked to repeat the target sentence aloud, to make their best guess if they were unsure, and to report isolated words if that was all they were able to pick out of the noise. The experimenter entered the number of correctly identified keywords (regardless of word order in the participant's response) for each trial into her computer, after which the participant saw the target sentence on his/her computer monitor. The disappearance of the text from the screen indicated that the next stimulus would play after a 500 ms delay. The listener's speech was recorded directly to computer using an AKG C420 microphone in order to maintain a record of listener responses.

Listeners were trained and tested at their individual HINT thresholds -3 dB. This level was chosen on the basis of listener performance in Van Engen (2010), which tested native-speaking listeners at HINT +3, HINT +0, HINT -3, and HINT -6 dB. Based on this study, HINT -3 dB was a reasonable level at which to avoid both floor and ceiling effects. Listeners were tested relative to their HINT thresholds in order to normalize them according to their tolerance for energetic masking. That is, listeners who were able to tolerate a higher level of noise in the HINT test were presented with a correspondingly higher level of noise in the training and test phases of this study. In Van Engen (2010), this method successfully equated the performance levels of native and non-native English listeners on a similar task.

Testing

The post-test was administered after the second training session on Day 2. (For the control group, which did not receive training, the post-test was administered after pre-testing on Day 1.) The basic procedure was identical to the procedure for the training sessions, except that listeners received no feedback. Instead, an icon (a red circle) on their computer monitor flashed 500 ms before each stimulus was played so that they had the same preparation for each stimulus as they had in the training sessions. The post-test was comprised of four BKB lists (i.e., 64 sentences, for a total of 200 keywords) that were not used during training. Two lists were spoken by one of the training talkers, and two by a novel talker. The same training talker and novel talker were used for all participants. For each target talker, one list was mixed with English 2-talker babble and the other with Mandarin 2-talker babble. Between subjects, the four BKB lists were counter-balanced across the four talker/masker combinations. For presentation, all 64 sentences were randomized separately for each listener.

Control group participants and listeners who had been trained in speech-shaped noise were instructed that the noise in the post-test would be other people speaking, and that the noise would begin one-half second before the target (as in their training and/or in the HINT). They were also told that the target voices would be softer than the noise. Participants who were trained in babble were instructed that they would encounter new voices in the targets and the noise during the test. (This instruction was given because the test included a new target talker and included both English and Mandarin babble, whereas these listeners had only received one type of babble during training.) All participants were told that there would be two different target talkers during the post-test and were reminded that they were listening for simple, meaningful English sentences.

Data analysis

After each trial during training and testing, the experimenter recorded whether the listener identified each of the keywords (3 or 4 per sentence). Words with added or deleted morphemes were considered incorrect, but homophones were counted as correct. All identified keywords were counted as correct, regardless of word order in the listener's response.

Results

Letter-number sequencing

The mean raw score on the letter-number sequencing task was 21.43 (SD = 1.86), with scores ranging from 16 to 26 out of a possible 30. To put these values into context, the mean scaled score on the letter-number sequencing task was 11, which corresponds to a WAIS IQ score of 105 (IQ range: 85 - 145).

HINT

The mean SNR-50 on the HINT test for the participants in this study was -2.90 dB (standard deviation = .86). This result closely matches the data obtained during the development of the HINT: Nilsson et al. (1994) report a mean score of -2.92 dB with a 95% confidence interval of +/- 2.41 dB for normal-hearing listeners when speech and noise were co-located (in front of the listener). A one-way ANOVA showed that the listeners in the four different conditions did not differ significantly from one another on this measure, $F(3, 52) = 1.822$, $p = 0.1547$.

Training

Performance on Days 1 and 2 by each of the training groups is shown in Figure 5.1.

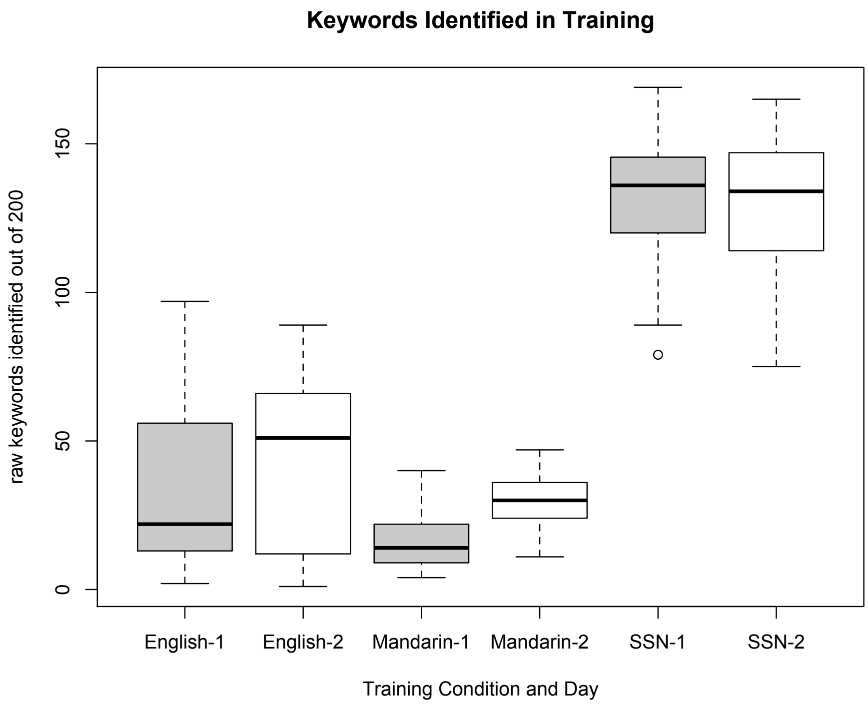

Figure 5.1. Keywords identified by each training group on Days 1 and 2 of training. Day 1 data is presented in grey boxplots; Day 2 in white boxplots. In each plot, the black line represents the median value, the box extends to the 25th and 75th percentile, and the whiskers extend to data points within 2.5 times the interquartile range.

The performance of the training groups during training was assessed statistically using a mixed-effects logistic regression model. This type of analysis avoids spurious results that can arise when categorical data are analyzed as proportions using ANOVAs. It also has greater power than ANOVAs and does not assume homogeneity of variances (see Jaeger (2008) for discussion of the benefits of this analysis). Furthermore, the mixed-effects regression allows for the inclusion of covariates. Analyses were performed using R, an open-source programming language/statistical analysis environment (R Development Core Team, 2005). Model comparison was performed to identify the random factors, fixed factors, and covariates that best fit the data. The resulting model includes subjects as a random factor and training language, training day, and their interaction as fixed factors. Note that the inclusion of SNR and working memory scores as covariates did not significantly improve the fit of the regression model, nor did inclusion of interactions between these measures and the other factors in the model. The results of the regression are shown in Table 1. The training condition factor was contrast-coded to investigate the following comparisons of interest:

Condition 1: Speech-shaped noise training vs. babble training

Condition 2: Mandarin training vs. English training

	Estimate	Standard Error	z-value	p-value
(Intercept)	-1.24074	0.12509	-9.919	< 2e-16
Condition 1 (SSN vs. babble)	-2.83950	0.26268	-10.810	< 2e-16
Condition 2 (Mandarin vs. English)	0.56802	0.30945	1.836	0.0664
Day 2 (vs. Day 1)	0.30648	0.04385	6.989	2.77e-12
Condition 1 x Day 2	0.49464	0.08372	5.909	3.45e-09
Condition 2 x Day 2	-0.30840	0.11719	-2.632	0.0085

Table 5.1. Coefficients of the mixed-effects model and associated Z-statistics for performance during speech-in-noise training.

The results for training condition show that the probability of correct keyword identification was significantly higher in the speech-shaped noise condition versus the babble conditions (Condition 1: $z = -10.810$, $p < 0.0001$).

The results for training day show that the overall probability of correct keyword identification was higher on day 2 than on day 1 ($z = 6.989$, $p < 0.0001$). There were also, however significant interactions between training day and training condition, both for the comparison of speech-shaped noise versus babble ($z = 5.909$, $p < 0.0001$) and the comparison of Mandarin versus English training ($z = -2.632$, $p = 0.0085$).

Follow-up regressions were performed to examine the significant interactions shown above. These regressions separated the training data for the three groups to investigate the relationship between performance on days 1 and 2 for each group. The results are shown in Tables 5.2-5.4 below.

Speech-shaped noise	Estimate	Standard Error	z-value	p-value
(Intercept)	0.65058	0.15097	4.309	1.64e-05
Day 2 (vs. Day 1)	-0.02329	0.05972	-0.390	0.696

Table 5.2. Coefficients of the mixed-effects models and associated Z-statistics for performance on Days 1 and 2 of training for the group trained in speech-shaped noise.

Mandarin	Estimate	Standard Error	z-value	p-value
(Intercept)	-2.45486	0.13550	-18.12	<2e-16
Day 2 (vs. Day 1)	0.62347	0.09063	6.88	6e-12

Table 5.3. Coefficients of the mixed-effects models and associated Z-statistics for performance on Days 1 and 2 of training for the group trained in Mandarin babble.

English	Estimate	Standard Error	z-value	p-value
(Intercept)	-1.93622	0.32607	-5.938	2.88e-09
Day 2 (vs. Day 1)	0.31867	0.07433	4.287	1.81e-05

Table 5.4. Coefficients of the mixed-effects models and associated Z-statistics for performance on Days 1 and 2 of training for the group trained in English babble.

The speech-shaped noise group showed no effect of training day ($z = -0.390$, $p = .696$), while both of the babble training groups did show significant effects (Mandarin: $z = 6.99$, $p < 0.0001$; English: $z = 4.287$, $p < 0.0001$). The interaction between Condition 1 and training day, therefore, arises because the babble-trained groups, but not the speech-shaped noise group, improved during training. The interaction between Condition 2 and training day arises because the Mandarin group showed greater improvement between days 1 and 2 than the English group (as indicated by the larger parameter estimate for the Mandarin group).

Test

The primary interest of this study is how the different types of training affected performance on the common speech-in-speech post-test. The post-test data was analyzed using mixed-effects logistic regression, and model comparison was performed to identify the random factors, fixed factors, and covariates that best fit the data. The resulting model includes the following factors:

<u>Random factors</u>

1) Subject

<u>Fixed factors</u>

1) Training condition

   This factor was contrast-coded to investigate the following comparisons of interest:

   Training 1: Trained vs. Control

   Training 2: Trained in speech-shaped noise vs. Trained in Babble

   Training 3: Trained in English babble vs. Trained in Mandarin babble

2) Masker (English vs. Mandarin)

3) Talker (familiar vs. new)

4) Interaction: Training Condition X Masker

5) Interaction: Training Condition X Talker

<u>Covariates</u>

1) SNR

2) Working memory

The results of the regression are shown in Table 5.5.

	Estimate	Standard Error	z-value	p-value
(Intercept)	-2.41588	1.08987	-2.217	0.026645
Training 1	1.09554	0.21031	5.209	1.90e-07
Training 2	0.62714	0.23081	2.717	0.006586
Training 3	0.64907	0.26109	2.486	0.012920
Masker (Mandarin vs. English)	1.37194	0.04687	29.272	< 2e-16
Talker (new vs. old)	-0.19376	0.04462	-4.343	1.41e-05
SNR	0.35500	0.10081	3.521	0.000429
Working Memory	0.14030	0.04598	3.052	0.002277
Training 1 X Masker	-0.69719	0.10805	-6.453	1.10e-10
Training 2 X Masker	-0.45754	0.11958	-3.826	0.000130
Training 3 X Masker	-1.06714	0.12658	-8.431	< 2e-16
Training 1 X Talker	-0.24251	0.09984	-2.429	0.015144
Training 2 X Talker	0.04515	0.11461	0.394	0.693624
Training 3 X Talker	0.11102	0.12336	0.900	0.368159

Table 5.5. Coefficients of the mixed-effects model and associated z-statistics for performance on the speech-in-speech post-test.

Talker

Mean keyword identification for each of the talkers is shown by training group in Figure 5.2.

Figure 5.2. Average percentage of keywords identified by listeners in each training condition in the post-test for target sentences spoken by the training talker and the new talker. Error bars indicate standard error.

The regression shows that Talker is a significant predictor of accuracy, with accuracy being better for the familiar talker than for the novel target talker ($z = -4.343$, $p < .0001$). There is also a significant interaction between Training 1 (training vs. no training) and Talker, showing that it is in *trained* participants that the familiar talker yielded higher accuracy than the novel talker ($z = -2.429$, $p = .015$). This interaction, with the talker effect driven by the performance of the trained participants, confirms the equivalence in the intelligibility of the two talkers for naïve listeners. The interaction was not significant for training in noise vs. babble (Training 2) or for training in English vs. Mandarin (Training 3), showing that the talker familiarity effect did not depend on the noise type used for training.

Masker

The regression also shows that overall performance in the Mandarin masker was significantly better than in the English masker ($z = 29.272$, $p < 0.0001$). This finding replicates previous studies that have shown better performance in a foreign language masker than in a native-language masker (Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Van Engen, 2010; Calandruccio et al., in press). Figure 5.3 shows mean performance by training group in the two masker languages.

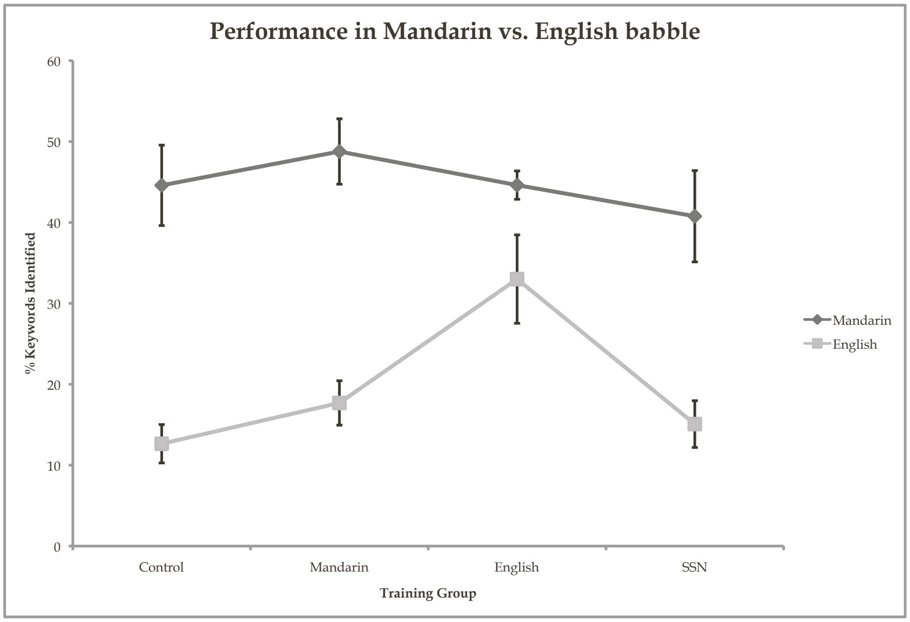

Figure 5.3. Average percentage of keywords identified by listeners in each training group in the post-test for target sentences embedded in Mandarin and English babble. Error bars indicate standard error.

Training condition

The regression shows a significant effect of training vs. no training (Training 1: $z = 5.209$, $p < 0.0001$), a significant effect of speech-shaped noise vs. babble training (Training 2: $z = 2.717$, $p = 0.006586$), and a significant effect of English vs. Mandarin training (Training 3: $z = 2.486$, $p = 0.012920$). Overall, these effects indicate that the likelihood of correct keyword identification in the post-test is higher in trained vs. untrained participants; in babble-trained vs. noise-trained participants, and in English-trained vs. Mandarin-trained participants.

In addition, however, there were significant interactions between all three training contrasts and the effect of Masker language. The effect of Masker language was reduced for the trained listeners relative to the untrained listeners ($z = -6.453$, $p < 0.0001$); for babble-trained vs. noise-trained subjects, ($z = -3.826$, $p = 0.000130$); and for English- vs. Mandarin-trained participants (z

= -8.431, $p < 0.0001$). That is, training of any type modulated the masker language effect, and babble training did so more than speech-shaped noise. Most importantly, the final interaction pattern arises because the English training group performed better than the Mandarin group in English (and, if anything, worse than the Mandarin group in Mandarin). Although performance in the Mandarin masker remains better than in English overall for all groups, the English training group shows a significantly smaller difference in performance between the two maskers. The interaction pattern, therefore, suggests that listeners are most able to learn to tune out the type of babble that they received in training.

Working memory and SNR

There was a significant effect of working memory, such that higher working memory scores predicted better performance on the speech-in-speech test ($z = 3.052$, $p = 0.002277$). There was also a significant effect of SNR, such that testing at higher SNRs predicted better performance on the speech-in-speech test ($z = 3.521$, $p = 0.000429$). This result suggests that using HINT scores to determine the SNR for the speech-in-speech test did not, in fact, equalize listeners' performance on keyword identification in 2-talker babble (in which case there would have been no significant effect of SNR). This result will be discussed in more detail below. Note that including interactions between SNR, working memory, and the other factors in the regression model did not improve the fit of the model. Crucially, this indicates that the SNR effect does not interact with the other factors of interest.

Discussion

This study has examined speech-in-noise training in several types of noise, showing that short-term training, in general, can provide a speech-in-speech intelligibility benefit and that listeners can both learn to tune in to familiar voices over short-term training and to tune out particular types of noise.

Performance during training

Listeners in the training conditions received two days of speech-in-noise training with feedback, listening to 64 target sentences per day. They were trained in English 2-talker babble, Mandarin 2-talker babble, or speech-shaped noise.

Although speech-shaped noise, which lacks the spectral and temporal “dips” that are present in 2-talker babble, imposes greater energetic masking than 2-talker babble on target speech, listeners who were trained in speech-shaped noise identified significantly more keywords during training than did the groups who were trained in babble. This result, which is similar to that of Helfer and Freyman (2009), shows that the informational masking imposed by 2-talker babble renders 2-talker babble an overall more effective masker than static, speech-shaped noise (at least for masking sentence-length stimuli), even though it imposes less energetic masking on the target speech.

This result contrasts with Simpson and Cooke (2005), who showed that babble with two talkers was a less effective masker than speech-shaped noise at an SNR of -6 dB (the average SNR in the current study was very similar: -5.9 dB). There were several differences between that study and the present one, however, that may account for the different results. First, Simpson and Cooke's task was closed-set consonant identification (16 VCVs produced by 5 talkers), whereas the training task in this study was an open-set sentence recognition task. The relative effects of multi-talker babble and speech-shaped noise on target speech intelligibility may, therefore, be dependent on the type of speech targets to be identified. In addition, the timing relations between the onsets of targets and noise differed across the two studies (the signal and noise were gated in Simpson and Cooke; noise led the signal by 500 ms in the current study). This difference may have contributed to the different relative effects of babble and speech-shaped noise. For example, it may be relatively more difficult to cope with speech-shaped noise when its onset coincides with the target onset, particularly when the targets are very short in duration as in the case of single syllables.

The finding that 2-talker babble was more detrimental to speech recognition than speech-shaped noise was also somewhat surprising in light of the results of Van Engen and Bradlow (2007), where performance was lower in 6-talker babble versus 2-talker babble. Van Engen and Bradlow attributed that difference to the increased energetic masking in 6-talker babble. Speech-shaped noise, which is an even denser signal than 6-talker babble, would then be expected to be a more, rather than less, effective masker than 2-talker babble. There are several important differences between Van Engen and Bradlow (2007) and the training task in this study that may affect overall levels of performance on the sentence-in-noise task, and may have contributed to the difficulty of the 2-talker babble conditions in training. First, Van Engen and Bradlow used static ("frozen") babble samples for speech-in-speech testing, whereas this study utilized randomly selected segments from longer babble tracks. Felty et al. (2009) compared these two types of babble in a word recognition task and showed a steeper learning curve in the frozen babble condition. The 2-talker babble in the current study, therefore, may have been more difficult to cope with overall, and more difficult to learn to tune out, than frozen babble samples. Second, participants in Van Engen and Bradlow (2007) had the benefit of performing the task first at a relatively easy SNR. Perhaps most importantly, performance in 2-talker babble in the training portion of this study may have been lowered by the fact that there were four target talkers as opposed to a single talker. As shown in Mullenix et al. (1989), word recognition in noise was higher when all words were spoken by a single talker versus by multiple talkers. Finally, it may simply be the case that 6-talker babble is a more effective masker than speech-shaped noise for sentence recognition due to the informational content that is still available in the babble. In Simpson and Cooke (2005), for example, 6-talker babble was indeed more effective than speech-shaped noise in masking consonants. A direct comparison of the effects of speech-shaped noise and multi-talker babbles (with various numbers of talkers) on sentence recognition is required to clearly delineate the relative effectiveness of these maskers for sentence recognition.

Surprisingly, the groups trained in English and Mandarin babble performed similarly to one another during the training sessions (though the effect of masker language was highly significant in the post-test, with higher performance in Mandarin than in English, as seen in Van Engen and Bradlow (2007) and Van Engen (2010)). Several of the differences between this study and the previous ones may also account for the lack of a noise language effect during training. First, the comparison of performance in the two maskers during training is *between* subjects, whereas it was *within* subjects in the previous studies and in the post-test. Second, as mentioned above, the babble type used for training (randomly-varying 2-talker babble) differed from the babble used in the previous studies (frozen 2-talker babble). It is doubtful that this explains the result, however, because the current post-test data, which also used randomly-varying noise, does show a noise language effect, and language effects have also been observed in other studies using randomly-varying babble (e.g., Brouwer et al., 2010; Calandruccio et al., in press). Another potentially relevant difference between the current training protocol and the previous studies is that the training sessions included feedback, and this additional information may have mitigated the noise language effect. For example, feedback to listeners in the English training condition might help them determine whether they were incorrectly attributing words from the maskers to the target speech, but this type of learning would not be available in Mandarin babble for listeners with no knowledge of that language. As mentioned above in the discussion of the effects of speech-shaped noise versus babble, perhaps the most important difference between this training and previous noise language studies is that the training used four talkers, whereas the previous studies used a single target talker throughout the speech-in-noise task. Preliminary analysis of data from a follow-up study using single-talker training suggests, indeed, that listeners may perform better in Mandarin than in English on the training task when there is only one target talker. It may be that the increased processing demands associated with listening for multiple target talkers led to reduced processing of the linguistic information in the interfering babble and, therefore, to a reduced effect of noise language.

In addition to comparing performance across the training groups, a comparison of performance on Days 1 and Day 2 within the three groups provides insight into whether listeners are able to improve their performance in the noise condition in which they receive training. This analysis showed there was no significant difference in listeners' performance in speech-shaped noise from Day 1 to Day 2. This finding suggests that the ability to cope with strictly energetic masking during speech recognition may not be particularly susceptible to training. This is not to say, however, that experience cannot affect listeners' ability to cope with energetic masking during speech recognition. Indeed, many studies have shown that even highly proficient non-native speakers of a target language perform worse than monolingual, native speakers of the language on speech recognition tasks in speech-shaped noise (e.g., Hazan and Simpson, 2000; Van Wijngaarden et al., 2002; Cooke et al., 2008; Van Engen, 2010), showing that experience with the target language does enhance processing of speech in conditions of energetic masking. Native listeners are better able to cope with such signal degradation by relying on useful alternative acoustic cues and contextual information at higher levels of linguistic structure. The current result suggests that, for native speakers of a language with normal hearing, coping with speech-shaped noise may not be particularly malleable. Performance in speech-shaped noise during training was still well below 100% (average = 65% of keywords identified), but it is possible that listeners were performing as best as is possible given the signal degradation in the speech-shaped noise condition. It should be noted, however, that these listeners did show significantly better performance on the training talker versus a new talker in the post-test, suggesting that they did adapt to the target talkers, even if this adaptation did not result in a significant improvement in performance during training.

In the English and Mandarin 2-talker babble training conditions, by contrast, listeners were able to make improvements in their performance between Days 1 and 2, indicating that people are more likely to make rapid improvements in their ability to cope with informational masking. These results suggest that auditory training programs that focus on improving listener performance in informational masking scenarios may be more successful than ones that aim to improve speech perception in primarily energetic masking environments.

Performance in the speech-in-speech post-test

Talker

The familiar post-test talker was one of four talkers used in the training sessions, meaning listeners who received training heard 32 training sentences (out of 128 total) spoken by that talker. Even with this relatively small amount of exposure, listeners were able to take advantage of talker familiarity in the post-test, showing that listeners are able to learn to tune in to a familiar voice after short exposure to that target voice in noise. The lack of significant interactions between training type and talker familiarity shows that no noise condition facilitated talker learning more than the others. That is, even though listeners who were trained in speech-shaped noise were able to understand a significantly greater number of target words than listeners trained in babble, this greater access to linguistic information during training either did not facilitate better learning of the target voice or any such learning did not generalize to the task of listening to target speech in 2-talker babble.

Masker

Overall performance in the post-test was significantly better in Mandarin 2-talker babble than in English 2-talker babble, replicating earlier findings in which native language babble was more detrimental than foreign language babble for target speech intelligibility (Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Brouwer et al., 2010; Van Engen, 2010; Calandruccio et al., in press). The sentence recordings from Van Engen and Bradlow (2007) and

Van Engen (2010) were also used to generate the 2-talker babbles for this study. It is noteworthy that different target voices were used in each of these studies, suggesting that the greater masking imposed by the English versus Mandarin babble is not due to the interactions of characteristics of the babble with characteristics of a particular target voice. As discussed above, this study also differed from the previous ones by utilizing randomly varying babble. The replication further suggests, therefore, that the noise language effect observed in earlier studies could not have been the result of listeners learning to tune out a small number of repeated frozen babble samples.

While English persisted as the more difficult masker in the post-test, there were also significant interactions between the language of the masker and the noise type in which listeners were trained. These will be discussed below.

Training condition

Most importantly, this study has shown that listeners can benefit from short-term speech-in-noise training: listeners who were trained performed better, overall, in the speech-in-speech post-test than listeners who did not receive training. Further, training in babble was more efficacious overall for coping with babble in the post-test than was training in speech-shaped noise, even though listeners in the speech-shaped noise condition actually performed better during the training sessions themselves.

Babble training effects appear to be primarily specific to the language of the training. Interactions between training conditions and performance in the two maskers show a reduction of the effect of native- versus foreign-language noise for all listeners who were trained. Most importantly, however, training in English versus Mandarin babble interacted with masker language such that English training led to a significantly smaller difference between the effects of the two maskers than did Mandarin training. As shown in Figure 5.4, the English group performed relatively well in English, thus reducing the difference between the two maskers' effects. Meanwhile, the group trained in Mandarin performed slightly better in Mandarin babble while English remained quite difficult, thus yielding a larger difference in performance between the two masker languages.

Even though English babble remained more difficult than Mandarin (even for English training group), the overall pattern of results supports the conclusion that English training can facilitate improvement in English babble. One potential explanation for such learning is that knowledge of the language spoken in the background noise may allow listeners to take advantage of indexical and linguistic cues that aid separation of the noise from the target speech signals. That is, some of the features of English noise that make it a relatively more difficult masker to cope with (i.e., the presence of acoustic and linguistic features that match those of the target speech) may also facilitate listeners' ability to learn to tune it out.

In sum, the training results showed that speech-in-speech learning was primarily context-specific: training in babble was more beneficial than training in speech-shaped noise, and benefits were primarily specific to the language of the noise in which listeners receive training. Furthermore, the results at least suggest that the benefits of such training may be greater for coping with native-language noise versus foreign-language noise. For the purposes of coping with speech-in-speech environments, auditory training that utilizes speech noise in the language of the trainee's primary communicative environment, therefore, will likely be most useful for improving everyday communication. Future studies are needed to identify the optimal number of babble talkers in training for making real-world speech-in-speech improvements.

Working memory

This study also showed that individuals' working memory capacity was significantly related to their ability to understand speech in the presence of 2-talker babble. Across listener groups (i.e., listeners in the control condition and all three training conditions) and test conditions (i.e., Mandarin and English babble, familiar and new talkers), working memory was a significant predictor of accurate keyword identification.

This finding is in line with previous studies that have also shown correlations between working memory measures and measures of speech perception in noise for both normal-hearing listeners and listeners with hearing impairment (e.g., Lunner, 2003; Parbery-Clark et al., 2009). Parbery-Clark et al. (2009) showed significant correlations between individuals' scores on the Woodcock-Johnson III Cognitive test (specifically working memory scores, which are composed of the Numbers Reversed and Auditory Working Memory subtests) and their performance both on the QuickSIN test (sentences in 12-talker babble) and the HINT-F (sentences in speech-shaped noise, with speech and noise presented from the same location). In a study of first-time hearing aid users, Lunner (2003) assessed working memory with a reading span measure and compared it to performance on a task of keyword identification in sentences presented in speech-shaped noise. Even after taking age and hearing impairment into account, a significant correlation was found between reading span and SNR threshold in both aided and unaided conditions, showing that good working memory is important for good performance in difficult listening conditions.

The current study extends such previous findings by showing a significant relationship between an auditory working memory task and sentence keyword intelligibility in 2-talker babble—a masking condition which involves higher levels of informational masking than speech-shaped noise or 12-talker babble. Working memory, therefore, appears to be an important predictor of successful speech recognition in a variety of noisy conditions.

SNR

The HINT estimates the SNR an individual requires in order to understand full sentences 50% of the time in speech-shaped noise. It therefore tests listeners' tolerance for energetic masking in the task of sentence recognition. Performance on the HINT was used as a basis for SNR selection for training and testing in this study in order to roughly equate listeners' performance levels on the speech-in-speech tasks according to their toleration for energetic masking. Based on data from Van Engen (2010), training and testing was conducted at HINT thresholds -3 dB. Using this approach, the average training/test SNR for the current set of listeners was -5.9 dB.

In Van Engen (2010), this method proved useful for normalizing the performance levels of native and non-native English speakers on a task of English sentence recognition in 2-talker babble. In the current study, however, SNR remained a significant predictor of correct keyword identification in the post-test, meaning that the SNR adjustment based on the HINT did not perfectly normalize performance levels on the speech-in-babble task within this group of normal-hearing, native-English speaking listeners. Listeners who had higher (i.e., worse) HINT thresholds received easier SNRs, and SNR, in turn, predicted correct keyword identification in the speech-in-babble test. For cross-population studies where significant differences are expected between groups, normalization of performance levels is often required, and this HINT method represents a useful approach. The current result suggests that, within a population of native-speaking, normal-hearing listeners, SNR normalization for speech-in-speech tasks is probably not necessary. It may be effective to use HINT scores instead as a covariate in analyses of speech-in-speech performance in order to determine more straightforwardly whether tolerance for energetic masking significantly predicts performance on an informational masking task.

While the HINT normalization scheme may have been unnecessary with the current population of listeners, it did not prevent meaningful analysis of the other factors of interest in this study because SNR did not significantly interact with any of them. Furthermore, the finding that good performance on the HINT does not mean an individual can tolerate 2-talker babble at more difficult SNRs than an individual with poorer performance on the HINT suggests that there is a dissociation between the ability to cope with purely energetic masking and the ability to cope with maskers that also impose informational masking on target speech.

With respect to the development of auditory training programs for speech perception in noise, the present SNR result further supports the use of masking that is maximally similar to the noise environment in which a listener needs to improve. That is, if coping with strictly energetic masking (such as the speech-shaped noise in the HINT) versus a combination of energetic and informational masking (as with 2-talker babble) requires different auditory and cognitive processes, then targeting the most relevant processes in training may lead to more successful training outcomes. If coping with noisy environments where speech is the primary noise source is the goal for a listener, then training that utilizes speech noise rather than non-speech noise will likely be more helpful. (This recommendation was also supported by the training condition results, since learning appeared to be greatest in the babble type used in training.) Furthermore, as noted above, listeners' performance during the training sessions also suggested that coping with informational masking may be more trainable than coping with energetic masking.

Future directions

This study has shown that two relatively short speech-in-noise training sessions over two days can produce significant post-test performance differences between trained and untrained listeners and between listeners who received different training conditions. While two sessions produced measureable results and can illuminate the types of learning that take place in speech-in-noise training, it is unlikely that this small amount of training would produce long-term changes in a person's ability to cope with noise. Additional research is required to determine how much training would be required to produce long-term, persistent changes in a person's ability to cope with noise.

Along with training duration, additional research is needed to determine the most efficacious SNRs to use in training. While this study did show that listeners can learn both to tune into a target talker and to tune out interfering babble, performance levels (at HINT -3 dB) were quite low overall – the average percentage of keywords identified over all of the groups in the post-test was just 32%. In order to develop maximally useful speech-in-noise training programs, further research on the role of SNR in speech-in-noise learning is needed. Training listeners by starting at an easy SNR and moving to more difficult SNRs, for example, may be a successful training strategy (see, e.g., Liu et al., 2008 for a discussion of easy-to-hard effects in auditory training), both for improving listeners ability to tune in to a target voice (by making that voice more salient in early exposure) and their ability to tune out (by making the noise easy to segregate from the target in early exposures). The training effect observed in Van Engen and Bradlow (2007), for example, occurred in a situation where performance at a given SNR was better after exposure to the task at an easier SNR first. Similarly, average performance for native-speaking listeners at HINT -3 dB in Van Engen (2010) was higher (49%) than in the present test (32%) after listeners first performed the task at HINT +3 dB and HINT +0 dB.[16]

This study showed that listeners were able to benefit from talker familiarity in the post-test, even when the familiar talker was just one of four that was presented in noise during training. Multiple talkers were used for training because previous studies have shown that high variability enhances generalization of perceptual learning for various aspects of speech perception. It is possible, however, that more exposure to a target talker would provide listeners with an even greater intelligibility benefit and/or would interact with listeners' ability to adapt to the various types of noise. In order to assess whether multiple-talker training is indeed, more efficacious than single-talker training for speech-in-speech recognition in particular, a follow-up study was performed with one talker in the training sessions. It was hypothesized that, by making the training task

[16] Note, however, that additional differences between Van Engen (2010) and the present study may also have led to better performance at HINT -3 dB in the previous study. In particular, Van Engen (2010) used frozen babble and a single target talker, whereas this study uses randomly varying babble and two talkers for the post-test.

easier (only one target talker), listeners may be able to benefit from more opportunity to tune in to the talker during training and be better able to allocate processing resources to learning to tune noise out. The preliminary analysis of the post-test data for this single-talker training study revealed significant complex interactions between training condition, masker, SNR, and talker. Crucially, although training performance was higher in the single-talker conditions, there was no significant simple effect or interaction effect involving the number of talkers used in training.

Finally, this study showed that SNR persisted as a predictor of performance on a speech-in-speech task after training, even after it was adjusted based on listeners' tolerance for energetic masking. Future research is needed, therefore, to examine the relationship (or lack thereof) between individuals' ability to cope with energetic and informational masking. A better understanding of this relationship will also provide important information for the future development of auditory training by suggesting ways to customize training for the needs of particular listeners.

Conclusion

This study of speech-in-noise training has shown that listeners are able to improve speech-in-speech recognition after short-term training. More specifically, listeners were able to improve their ability both to understand a particular talker in noise and to cope with speech noise that is employed during training. These results support the hypothesis that processes related both to tuning in and tuning out are involved in speech-in-speech recognition and are susceptible to auditory training.

CHAPTER 6: CONCLUSION

In the preceding chapters, I have presented data from several experiments that investigated the intelligibility of English sentences in the presence of noise. The results of these studies are summarized below.

Word vs. Non-word noise (Chapter 2):

Native English listeners identifying keywords in sentences in 2-talker babble were not differentially affected by babble that was composed of sentences whose content words were real words versus non-words.

SNR manipulation (Chapter 2):

Native English listeners identifying keywords in sentences at two SNRs in Mandarin and English 2-talker babble showed no differences in performance based on whether or not the mixed speech + noise stimuli were equated for RMS amplitude.

Language experience (Chapters 3 and 4):

Monolingual native English listeners and non-native English listeners whose native language is Mandarin both identified fewer keywords in English sentences in the presence of English 2-talker babble versus Mandarin 2-talker babble. The native English listeners, however, showed a greater release from masking in Mandarin babble relative to English babble (i.e., the difference in their performance in the two languages was greater than it was for the Mandarin listeners). The non-native listeners required, on average, an

8 dB SNR benefit in order to perform at the level of the native English listeners on the speech-in-speech task. In a follow-up study, early Spanish-English bilinguals did not differ significantly from monolingual English listeners on the task of keyword identification in sentences in English and Mandarin babble.

Speech-in-noise training (Chapter 5):

Native-speaking, normal-hearing listeners who received two days of speech-in-noise training outperformed untrained listeners on a sentence recognition task in English and Mandarin 2-talker babble. Interactions between training condition and performance in the two maskers indicated that babble training in a given language particularly benefited post-test performance in that babble language. Listeners were trained and tested at SNRs selected based on their performance on the HINT test, and SNR remained a significant predictor of success on the speech-in-speech post-test. Working memory, as measured by the letter-number sequencing task in the Wechsler Adult Intelligence Scale, was also a predictor of correct keyword identification in the post-test.

I have provided general conclusions for each of these findings in the preceding chapters. In this final chapter, I discuss some outstanding issues raised by the experiments presented in this dissertation and suggest directions for future research.

HINT normalization

The primary studies in this dissertation (Chapters 3-5) implemented what is, to my knowledge, a novel approach to normalizing listener performance on a speech-in-speech task. This approach –

to first administer the HINT (a standardized speech audiometric test) and use listeners' HINT thresholds as the basis for determining SNRs for speech-in-speech testing – was highly successful for the purpose of comparing native and non-native sentence recognition in two babble languages (Chapter 3). After selecting SNRs relative to individual HINT scores, the two populations performed at approximately the same level on the speech-in-speech test. This normalization, therefore, allowed for a more valid comparison of the relative effects of the two noise languages on the two populations, despite their differences in baseline speech-in-noise recognition. The HINT, which presents sentences in speech-shaped noise, provides a measure of listeners' tolerance for energetic masking in full-sentence recognition. Adjusting SNR based on this measure allowed for a well-controlled observation of relative informational masking effects in two distinct listener populations.

This same approach to performance normalization, used within a group of native-speaking, normal-hearing listeners in the training study (Chapter 5), however, did not eliminate SNR as a significant predictor of success on the speech-in-speech test. That is, listeners with poorer HINT thresholds were presented with higher (easier) SNRs for the speech-in-speech test, and these higher SNRs predicted success on that task. If the HINT-based testing had straightforwardly equated listeners' performance in speech noise, there should have been no significant effect of SNR on test performance.

In Chapter 3, the primary goal of HINT-based testing was to normalize two distinct listener populations – populations who were expected to perform significantly differently on a speech-in-noise task without any such adjustment. For this purpose, the method was successful. This approach is therefore recommended for future studies in which the comparison of interest lies in the relative effects of multiple maskers on multiple populations where significant baseline differences in coping with noise are expected. Since HINT-based testing entails that listeners are tested at different SNRs from one another, it of course cannot be used when performance in a given level of noise on a single task type is of interest.

While HINT normalization can be a useful method for comparing performance across populations, maskers, and SNRs, Chapter 5 showed that, when there is no group-wise performance level to be equated, the HINT normalization introduced a perhaps unnecessary complication. Simply put, normalization is probably not needed for native-speaking, normal-hearing listeners on a speech-in-noise task. For future within-population studies of speech perception in speech noise, then, listeners should be tested at the same SNR. The HINT (or a similar test) should still also be conducted for two primary reasons: first, to screen listeners and ensure that their speech-in-noise performance is within normal limits; and second, so that these scores (a proxy for energetic masking tolerance) can be entered into an overall analysis. The single-talker training data, mentioned at the end of Chapter 5, may have been more interpretable had this approach been taken instead of testing listeners relative to HINT scores.

Given that there are several important differences between the HINT and the speech-in-speech task, it is perhaps unsurprising that testing at levels selected relative to HINT thresholds did not perfectly normalize the performance of listeners sampled from the normal-hearing, native-speaking population on the speech-in-speech task. Aside from the crucial difference of noise type (speech-shaped noise versus babble), the HINT also uses full sentence scoring and an adaptive

SNR presentation method, whereas the speech-in-speech experiments used keyword identification (3-4 per sentence) scoring at static SNRs. Furthermore, the two tests use different target talkers, which may vary in intelligibility and susceptibility to distortion by different types of maskers.

Another approach to investigating different masker effects on different populations (and one that would avoid normalization procedures) is to use adaptive SNR presentation (as in the HINT) to estimate SNR thresholds for a given performance level. These thresholds can then be straightforwardly compared across noise conditions and listener groups. One difficulty with this approach when studying speech maskers (especially with a small number of talkers), is that, with randomly-varying samples of babble, the amount of energetic and informational masking present in a given trial can vary significantly from trial-to-trial. The methods employed in this dissertation provided many trials at each SNR, so that such variability could be averaged out. With adaptive testing, where performance on each trial affects the SNR of the next, many threshold measures would be required to obtain a good estimate of actual listener thresholds for various masker types. In addition, since adaptive testing yields a single threshold SNR for each listening condition, this method also has the disadvantage of not allowing for cross-SNR analyses. This is potentially problematic since the studies presented in Chapters 2-4 of this dissertation (and elsewhere) have shown that the relative effects of speech maskers differ across SNR conditions. Thus, studies of the effects of various speech maskers on different populations will continue to require the use of varied and complementary approaches to measuring performance.

It is important to emphasize that the persistence of an SNR effect in the speech-in-speech test in Chapter 5 did not compromise the investigation of different training conditions in that study, because SNR did not interact significantly with any of the other factors of interest; it simply emerged as a straightforward predictor of success on the task, regardless of training condition, target talker, masker language, or working memory capacity. The fact that SNR remained a significant predictor of success for the listeners in the training study, however, does represent a valuable result in and of itself. The HINT is a general measure of a person's ability to cope with energetic masking in sentence recognition while the speech-in-speech tasks measure a person's ability to cope with a combination of energetic and (crucially) informational masking. The finding that testing normally-hearing, monolingual listeners at noise levels relative to their HINT thresholds did not equate performance for the speech-in-speech task, therefore, suggests that distinct processes are involved when coping with strictly energetic masking versus coping with both energetic and informational masking. Additional research is warranted, therefore, on the relationship between individuals' tolerance for energetic and informational masking.

Linguistic masking effects in speech-in-speech recognition

Another critical issue is the extent to which we can understand how the linguistic content of speech maskers contributes to the overall masking of speech targets. This question is, of course, central to understanding the asymmetrical effects of different noise languages on different listener populations (e.g., Garcia Lecumberri and Cooke, 2006; Van Engen and Bradlow, 2007; Van Engen, 2010; Calandruccio et al., in press). In one approach to characterizing the role of language in masking, Cooke et al. (2008) include "interference from a known language" as one element of informational masking, along with misallocation of masker components to the target;

competing attention from the masker; and higher cognitive load imposed by a masker. Their framework, which will be considered further below, is as follows:

	→ energetic masking	→ partial information
NOISE		
	→informational masking	→ misallocation of audible masker components to target
		→ competing attention of masker
		→ higher cognitive load
		→ *interference from "known language" masker*

As in the Cooke et al. framework, this dissertation argues that the linguistic content of speech maskers contributes to informational masking. However, the possible role of energetic masking in driving differential language effects must also be considered. A recent study by Mattys et al. (2009) suggests, for example, that energetic masking may largely account for linguistic interference effects. In a study of the effects of various types of perceptual and cognitive loads on listeners' reliance on acoustic detail versus lexical-semantic information in word segmentation, they found that masker intelligibility (one competing talker as compared to modulated speech-spectrum noise) did not affect listeners' ratings of word boundaries. They conclude that this experiment showed no trace of linguistic interference beyond what could be explained by energetic considerations. At some levels of target speech processing, therefore, the linguistic content of maskers may not contribute significantly to informational masking. Of course, judging the location of word boundaries in ambiguous tokens under various conditions (the task in Mattys et al. (2009)), is very different from sentence identification in noise, which is where the studies in this dissertation showed different noise effects based on noise language.

For such comparisons, babbles in different languages may impose similar amounts of energetic masking, but it is unlikely that energetic masking is truly equal across babble languages. This means that energetic considerations may account for at least some of the differences in the effects of maskers in different languages. In Van Engen and Bradlow (2007), for example, there were significant differences between the long-term average spectra of the English and Mandarin babble at some frequencies, which may at least have contributed to their differential effects on listeners' sentence identification.

In comparing the effects of two noise languages (rather than a speech masker versus a non-speech masker), it is difficult, if not impossible, to completely control the energetic masking imposed by noise in two languages. First of all, spectral and temporal differences may arise from differences in the voices used to produce particular babble tracks. Such differences could be reduced in future studies by generating babble using recordings from bilingual speakers. However, energetic differences across babble languages may also result from the various linguistic features of the different languages, such as rhythmic structures, prosodic structures, or even the acoustics of particular phonetic segments. For example, two languages with different rhythmic structures may differ in terms of temporal modulation rates, and this could lead to energetic masking differences that would persist even within speech produced by a single talker in the two languages.

Although it is difficult to completely isolate energetic and informational masking when using real speech signals as maskers, the finding, in Chapter 3 (and in Calandruccio et al., in press) that the *relative* effects of different noise languages can differ across listener populations with different language backgrounds suggests that there are, indeed, informational masking differences across noise languages. That is, the relative energetic masking between two maskers is necessarily equal for all listeners, but the informational status of speech signals can vary across different listener groups. In addition, the different relative effects of maskers across different SNRs, where again the relative energetic masking is constant, suggests that informational masking varies. Additional experiments will be needed, therefore, to continue to delineate the specific roles of energetic and informational masking in speech-in-speech scenarios.

One approach to addressing the issue of energetic masking differences across masker languages is to manipulate babble signals to equate additional acoustic parameters (beyond RMS amplitude). While such methods have the disadvantage of moving away from natural speech signals, there may be some additional acoustic controls that would minimally interfere with the naturalness of the speech signals. One such approach has been implemented in follow-up work conducted in collaboration with Susanne Brouwer, Lauren Calandruccio, and Ann Bradlow. In this study (Brouwer et al., 2010), we have equated the long-term average spectra of 2-talker babble tracks in English and Dutch.

It is argued here, therefore, that the linguistic content of a masker indeed contributes to informational masking. More specifically, speech maskers can impose variable amounts of informational masking, depending on listeners' language background and on the relationship between the language(s) spoken in speech targets and speech noise. In all the studies presented here (Chapters 2-5), native English speakers had more difficulty identifying English keywords in sentences in the presence of English babble versus Mandarin babble. Since these listeners are native speakers of the more problematic noise language, this result suggests that the processing of linguistic information (e.g., prosodic, phonetic, lexical, semantic information) in the noise detracts from accurate identification of target speech. The basic result, however, could also be attributed to the greater similarity between English targets and English noise versus English targets and Mandarin noise. Targets and noise that share many acoustic and linguistic features are likely to be more difficult to segregate, regardless of listeners' knowledge of the language(s) involved. Linguistically-driven masking, therefore, may result both from the diversion of processing resources to the linguistic information in the noise and/or to the greater difficulty in segregating signals with shared linguistic features. The performance of Mandarin listeners compared to English listeners in Chapter 3 suggests that both of these processes are involved in speech-in-speech perception. These listeners had greater difficulty coping with English babble than Mandarin while identifying English targets, suggesting that segregation of matched language noise is more difficult than "tuning out" native language noise. However, they also showed a smaller release from masking in Mandarin noise than the English listeners, suggesting that the processing of native language information in the noise also contributes to informational masking.

To return to Cooke et al. (2008), that framework includes knowledge of the masker language as a separate component of informational masking, along with misallocation, competing attention, and cognitive load. However, it seems unnecessary at this point to separate knowledge of the masker language from these other three elements of informational masking. Instead, it may be more appropriate to view the linguistic content of speech maskers as a driving force behind these other mechanisms of informational masking. Knowledge of the language spoken in the noise may lead to a greater attentional and/or cognitive load, since the signal carries meaningful information for the listener. Known-language maskers may also increase the number of misallocations a listener makes, especially where the masker language matches the target language, in which case the two signals share linguistic features at all levels of linguistic structure.

To summarize, speech maskers in different languages may certainly differ in terms of the energetic masking they impose on speech targets, although the studies in this dissertation and elsewhere suggest that there are particularly informational masking effects of speech maskers, at least in the task of sentence recognition in noise. Speech maskers likely impose informational masking by increasing misallocations of noise components to signals, as well as by imposing higher attentional and cognitive loads on listeners. To date, linguistic factors cannot be easily separated from these other mechanisms of informational masking.

In addition to characterizing the role of linguistic information in masking, another issue that remains to be resolved is the identification of the level of linguistic processing at which linguistic interference occurs. Linguistic masking effects may arise as a result of listeners processing and/or misallocating any of a number of linguistic features, including prosodic (see Reel, 2009 regarding rhythm), phonetic, phonological, lexical, syntactic, or semantic features. The word/non-word study presented here rules out the strictest lexical explanation – that the presence versus absence of real content words in the noise leads to differential masking effects. As noted in Chapter 2, this result does not rule out lexical processing of the noise as a source of linguistic interference, since these non-words were so word-like as to induce real-word errors in listener responses. We have begun to explore other relevant levels of linguistic processing by investigating the effects of noise languages (e.g., Dutch, Croatian, Mandarin) that share varying amounts of linguistic features with English (Brouwer et al., 2010; Calandruccio et al., in press). In addition, we have compared the effects of noise comprised of semantically-normal versus semantically-anomalous sentences, finding that listeners do, indeed, have greater difficulty coping with semantically-normal noise (Brouwer et al., 2010).

Auditory training

The training study presented in Chapter 5 has shown that the effects of speech noise on target speech intelligibilty are subject not only to listeners' language experience as it interacts with the linguistic content of the noise, but also to auditory training experience. This study showed that two days of speech-in-noise training can modulate noise language effects. After training, English 2-talker babble was more detrimental than Mandarin to English keyword identification than Mandarin babble for monolingual English listeners overall. Crucially, this was true even after training in English babble. However, training in English versus Mandarin significantly reduced the noise language effect, suggesting that listeners were able to learn to better inhibit the processing of the native-language masker and/or they became better at segregating speech noise from the target as they gained experience with the noise. The Mandarin-trained group had no knowledge of that language but did show slight improvement in Mandarin babble, suggesting that improvement took place in the domain of target and noise segregation. Training did not appear to transfer from one training language to the other, suggesting that learning to "tune out" was specific to the noise type in which the listeners were trained.

The training study presented here, further, provides insight into speech-in-noise training parameters that may be utilized beneficially in auditory training programs. For example, the comparison of training conditions suggests that listeners will likely benefit more from training in speech noise in the language of their communicative environment than from training in non-speech noise or foreign-language noise: improvements were most significant after babble training and particularly in the language of the training. Speech-in-noise training has been implemented in Listening and Communication Enhancement (LACE), which is a commercially available auditory training program. LACE uses adapative SNR presentation and both a single competing talker and 6-talker babble. The results of the current study support LACE's use of competing speech and babble in English. There are many more aspects of speech-in-noise training, however, that still need to be explored. In particular, further research is needed to determine how SNR can best be utilized in speech-in-speech training, how many competing talkers should be used, and what is an appropriate training duration. As effective methods for auditory training are specified, their efficacy must be tested for relevant populations, such as hearing-impaired listeners, elderly listeners, and second-language users.

Individual differences

Finally, the research presented here revealed much variability in performance on speech-in-noise tasks across normal-hearing individuals. In order to understand this variability and how it may relate to speech-in-noise processing for other populations, further research is needed to investigate the individual differences that underlie such variability. Factors to consider more deeply include the roles of working memory and executive control in speech-in-speech recognition, as well as the relationship between an individual’s tolerance for energetic versus informational masking.

Conclusion

The fact that most everyday speech perception occurs in the presence of noise represents both a clinical challenge and a challenge to theories of speech perception. This dissertation has unmasked a component of this complex problem, showing that listeners’ experience with target and noise languages and training-related experience can modulate the intelligibilty of speech in the presence of speech noise.

REFERENCES

Arbogast, T. L., Mason, C. R. and Kidd, G., 2005. "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners." Journal of the Acoustical Society of America 117(4), 2169-2180.

Audacity Team (2006). Audacity.

Bamford, J. and Wilson, I., 1979. Methodological considerations and practical aspects of the BKB sentence lists, in: Bench, J. and Bamford, J., Speech-hearing tests and the spoken language of hearing-impaired children, Academic Press, London, pp. 148-187.

Bamiou, D.-E., Musiek, F. E. and Luxon, L. M., 2001. "Aetiology and clinical presentations of auditory processing disorder: A review." Archives of Disease in Childhood 85, 1-9.

Bent, T. and Bradlow, A. R., 2003. "The interlanguage speech intelligibility benefit." Journal of the Acoustical Society of America 114(3), 1600-1610.

Bent, T., Buchwald, A. and Pisoni, D. B., 2009. "Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech." Journal of the Acoustical Society of America 126(5), 2660-2669.

Bialystok, E., Craik, F. I. M., Klein, R. and Viswanathan, M., 2004. "Bilingualism, aging, and cognitive control: evidence from the Simon Task." Bilingualism, aging, and cognitive control: evidence from the Simon Task 19(2), 290-303.

Bialystok, E. and Martin, M. M., 2004. "Attention and inhibition in bilingual children: evidence from the dimensional change card sort task." Developmental Science 7(3), 325-339.

Bradlow, A. R. and Alexander, J. A., 2007. "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners." Journal of the Acoustical Society of America 121(4), 2339-2349.

Bradlow, A. R. and Bent, T., 2002. "The clear speech effect for non-native listeners." Journal of the Acoustical Society of America 112(1), 272-284.

Bradlow, A. R. and Bent, T., 2008. "Perceptual adaptation to non-native speech." Cognition 106(2), 707-729.

Bradlow, A. R. and Pisoni, D. B., 1999. "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors." Journal of the Acoustical Society of America 106(4), 2074-2085.

Broersma, P. and Weenink, D. (2009). Praat: doing phonetics by computer.

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

Bronkhorst, A. W. and Plomp, R., 1992. "Effect of multiple speechlike maskers on binaural speech recognition in normal and imparied hearing." Journal of the Acoustical Society of America 92, 3132-3138.

Brouwer, S., Calandruccio, L., Van Engen, K. J. and Bradlow, A. R., 2010. "How native and non-native listeners cope with semantic content in L1 and L2 noise."

Brungart, D. S., Simpson, B. D., Ericson, M. A. and Scott, K. R., 2001. "Informational and energetic masking effects in the perception of multiple simultaneous talkers." Journal of the Acoustical Society of America 110(5), 2527-2538.

Byrne, D., Dillon, H., Tran, K., Arlinger, S. and Wilbraham, K., 1994. "An international comparison of long-term average speech spectra." Journal of the Acoustical Society of America 96, 2108-2120.

Calandruccio, L., Van Engen, K. J., Dhar, S. and Bradlow, A. R., in press. "The effects of clear speech as a masker." Journal of Speech, Language, and Hearing Research.

Callan, D. E., Jones, J. A., Callan, A., M. and Akahane-Yamada, R., 2004. "Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory – auditory/orosensory internal models " NeuroImage 22, 1182-1194.

Chermak, G. D., Somers, E. K. and Seikel, J. A., 1998. "Behavioral signs of central auditory processing disorder and attention deficit disorder." Journal of the American Academy of Audiology 9, 78-84.

Cherry, E. C., 1953. "Some experiments on the recognition of speech, with one and two ears." Journal of the Acoustical Society of America 25(5), 975-979.

Clahsen, H. and Felser, C., 2006. "How native-like is non-native language processing?" Trends in Cognitive Sciences 10, 564-570.

Clarke, C. M. and Garrett, M. F., 2004. "Rapid adaptation to foreign-accented English." Journal of the Acoustical Society of America 101(4), 2299-2310.

Clopper, C. G. and Pisoni, D. B., 2007. "Free classification of regional dialects of American English." Journal of Phonetics 35, 421-438.

Cooke, M., 2006. "A glimpsing model of speech perception in noise." Journal of the Acoustical Society of America 119(3), 1562-1573.

Cooke, M., Garcia Lecumberri, M. L. and Barker, J., 2008. "The foreign language cocktail party problem: energetic and informational masking effects on non-native speech perception." Journal of the Acoustical Society of America 123(1), 414-427.

Cutler, A., Garcia Lecumberri, M. L. and Cooke, M., 2008. "Consonant identification in noise by native and non-native listeners: effects of local context." Journal of the Acoustical Society of America 124(2), 1264-1268.

Cutler, A., Webber, A., Smits, R. and Cooper, N., 2004. "Patterns of English phoneme confusions by native and non-native listeners." Journal of the Acoustical Society of America 116(6), 3668-3678.

Cycling '74 (2005). Max/MSP 4.5.

Darwin, C. J., 2008. "Listening to speech in the presence of other sounds." Philosophical transactions of the Royal Society 363, 1011-1021.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K. and Carolyn, M., 2005. "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences." Journal of Experimental Psychology: General 134(2), 222-241.

Dupoux, E. and Green, K. P., 1997. "Perceptual adjustment to highly compressed speech: effects of talker and rate changes." Journal of Experimental Psychology: Human Perception and Performance 23, 914-927.

Durlach, N., 2006. "Auditory masking: Need for improved conceptual structure." Journal of the Acoustical Society of America 120(4), 1787-1790.

ETS (2005). TOEFL Internet-based Test: Score Comparison Tables, Educational Testing Services.

Felty, R. A., Buchwald, A. and Pisoni, D. B., 2009. "Adaptation to frozen babble in spoken word recognition." Journal of the Acoustical Society of America 125(3), EL93-EL97.

Francis, A. L., Nusbaum, H. C. and Fenn, K., 2007. "Effects of training on the acoustic phonetic representation of synthetic speech." Journal of Speech Language and Hearing Research 50, 1445-1465.

Freyman, R. L., Balakrishnan, U. and Helfer, K. S., 2001. "Spatial release from informational masking in speech recognition." Journal of the Acoustical Society of America 109(5), 2112-2122.

Freyman, R. L., Balakrishnan, U. and Helfer, K. S., 2004. "Effect of number of masking talkers and auditory priming on informational masking in speech recognition." Journal of the Acoustical Society of America 115(5), 2246-2256.

Garcia Lecumberri, M. L. and Cooke, M., 2006. "Effect of masker type on native and non-native consonant perception in noise." Journal of the Acoustical Society of America 119(4), 2445-2454.

Greenspan, S. L., Nusbaum, H. C. and Pisoni, D. B., 1988. "Perceptual learning of synthetic speech produced by rule." Journal of Experimental Psychology: Learning, Memory, and Cognition 14, 421-433.

Hazan, V. and Simpson, A., 2000. "The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects." Language and Speech 43(3), 273-294.

Helfer, K. S. and Freyman, R. L., 2009. "Lexical and indexical cues in masking by competing speech." Journal of the Acoustical Society of America 125(1), 447-456.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S. and Carlyon, R. P., 2008. "Perceptual learning of noise vocoded words: effects of feedback and lexicality." Journal of Experimental Psychology: Human Perception and Performance 34(2), 460-474.

Hugdahl, K., Heiervang, E., Nordby, H., Smievoll, A. I., Steinmetz, H., Stevenson, J. and Lund, A., 1998. "Central auditory processing, MRI morphometry, and brain laterality: Applications to dyslexia." Scandinavian Audiology 27(Supplement), 26-34.

Jaeger, T. F., 2008. "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models." Journal of Memory and Language 59, 434-446.

Jiang, J., Chen, M. and Alwan, A., 2006. "On the perception of voicing in syllable-inital plosives in noise." Journal of the Acoustical Society of America 119(2), 1092-1105.

Kessler, C. and Quinn, M. E., 1980. Positive effects of bilingualism on science problem-solving abilities, in: Alatis, J. E., Current issues in bilingual education: Proceedings of the Georgetown Roundtable on Languages and Linguistics, Georgetown University Press, Washington, DC, pp. 295-308.

Kessler, C. and Quinn, M. E., 1987. "Language minority children's linguistic and cognitive creativity." Journal of Multilingual and Multicultural Development 8, 173-186.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J. and Durlach, N. I., 2007. Informational Masking, in: Yost, W. A., Popper, A. N. and Fay, R. R., Auditory Perception of Sound Sources, Springer US, pp. 143-189.

Killion, M. C. and Niquette, P. A., 2000. "What can the pure-tone audiogram tell us about a patient's SNR loss?" The Hearing Journal 53, 46-53.

King, W. M., Lombardino, L. J., Crandell, C. C. and Leonard, C. M., 2003. "Comorbid auditory processing disorder in developmental dyslexia." Ear and Hearing 24, 448-456.

Kormi-Nouri, R., Moniri, S. and Nilsson, L.-G., 2003. "Episodic and semantic memory in bilingual and monolingual children." Scandinavian Journal of Psychology 44(1), 47-54.

Kraus, N., McGee, T., Carrell, T. D., C., K., Tremblay, K. and Nicol, T., 1995. "Central auditory system plasticity associated with speech descrimination training." Journal of Cognitive Neuroscience 7, 25-32.

Liu, E. H., Mercado, E. I., Church, B. A. and Orduña, I., 2008. "The easy-to-hard effect in human (homo sapiens) and rat (rattus norvegicus) auditory identification." Journal of Comparative Psychology 122(2), 132-145.

Lively, S. E., Logan, J. S. and Pisoni, D. B., 1993. "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories." Journal of the Acoustical Society of America 94(3), 1242-1255.

Loebach, J. L. and Pisoni, D. B., 2008. "Perceptual learning of spectrally degraded speech and environmental sounds." Journal of the Acoustical Society of America 123(2), 1126-1139.

Logan, J. S., Lively, S. E. and Pisoni, D. B., 1991. "Training Japanese listeners to identify English /r/ and /l/: A first report." Journal of the Acoustical Society of America 89(2), 874-886.

Lunner, T., 2003. "Cognitive function in relation to hearing aid use." International Journal of Audiology 42, S49-S58.

Marian, V., Blumenfeld, H. K. and Kaushanskaya, M., 2007. "The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals." J Speech Lang Hear Res 50(4), 940-967.

Mattys, S. L., Brooks, J. and Cooke, M., 2009. "Recognizing speech under a processing load: Dissociating energetic from informational factors." Cognitive Psychology 59, 203-243.

Mattys, S. L. and Liss, J. M., 2008. "On building models of spoken-word recognition: when there is as much to learn from natural "oddities" as articificial normality." Perception and Psychophysics 70(7), 1235-1242.

Mattys, S. L., White, L. and Melhorn, J. F., 2005. "Integration of multiple speech segmentation cues: a hierarchical framework." Journal of Experimental Psychology 134(4), 477-500.

Mayo, L. H., Florentine, M. and Buus, S., 1997. "Age of second-language acquisition and perception of speech in noise." Journal of Speech, Language and Hearing Research 40, 686-693.

McGarr, N. S., 1983. "The intelligibility of deaf speech to experienced and inexperienced listeners." Journal of Speech and Hearing Research 26, 451-458.

Mueller, J. L., 2005. "Electrophysiological correlates of second language processing." Second Language Research 21(2), 152-174.

Mullenix, J., Pisoni, D. B. and Martin, C. S., 1989. "Some effects of talker variability on spoken word recognition." Journal of the Acoustical Society of America 85, 365-378.

Nilsson, M., Soli, S. D. and Sullivan, J. A., 1994. "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise." Journal of the Acoustical Society of America 95(2), 1085-1099.

Nygaard, L. C. and Pisoni, D. B., 1998. "Talker-specific learning in speech perception." Perception and Psychophysics 60, 355-376.

Nygaard, L. C., Sommers, M. S. and Pisoni, D. B., 1994. "Speech perception as a talker contingent process." Psychological Science 5, 42-46.

Pallier, C., Sebastian Galles, N., Dupoux, E., Christophe, A. and Mehler, J., 1998. "Perceptual adjustment to time-compressed speech: a cross-linguistic study." Memory and Cognition 26(4), 844-851.

Parbery-Clark, A., Skoe, E., Lam, C. and Kraus, N., 2009. "Musician enhancement for speech-in-noise." Ear and Hearing 30(6), 653-661.

Parikh, G. and Loizou, P. C., 2005. "The influence of noise on vowel and consonant cues." Journal of the Acoustical Society of America 118(6), 3874-3888.

Pearson Education Inc., 2008. Wechsler Adult Intelligence Scale - Fourth Edition, San Antonio, TX.

Peters, R. W., Moore, B. C. and Baer, T., 1998. "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-imparied and normally-hearing people." Journal of the Acoustical Society of America 103, 577-587.

Plomp, R. and Mimpen, A. M., 1979. "Speech-reception threshold for sentences as a function of age and noise level." Journal of the Acoustical Society of America 66(5), 1333-1342.

R Development Core Team (2005). R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing.

Ransdell, S., Arecco, M. R. and Levy, C. M., 2001. "Bilingual long-term working memory: the effects of working memory loads on writing quality and fluency." Applied Psycholinguistics 22, 113-128.

Reel, L. A. (2009). Selective auditory attention in adults: effects of rhythmic structure of the competing language. Communication Sciences and Disorders, Texas Tech University Health Science Center. **Doctor of Philosophy:** 151.

Rhebergen, K. S., Versfeld, N. J. and Dreschler, W. A., 2005. "Release from informational masking by time reversal of native and non-native interfering speech (L)." Journal of the Acoustical Society of America 118(3), 1-4.

Rogers, C. L., Dalby, J. and Nishi, K., 2004. "Effects of noise and proficiency on intelligibility of Chinese-accented English." Language and Speech 47(2), 139-154.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M. and Abrams, H. B., 2006. "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing." Applied Psycholinguistics 27, 465-485.

Schwab, E. C., Nusbaum, H. C. and Pisoni, D. B., 1985. "Effects of training on the perception of synthetic speech." Human Factors 27, 395-408.

Sidaras, S. K., Alexander, J. E. D. and Nygaard, L. C., 2009. "Perceptual learning of systematic variation in Spanish-accented speech." Journal of the Acoustical Society of America 125(5), 3306-3316.

Simpson, S. A. and Cooke, M., 2005. "Consonant identification in N-talker babble is a nonmonotonic function of N." Journal of the Acoustical Society of America 118(5), 2775-2778.

Smiljanic, R. and Bradlow, A. R., 2005. "Production and perception of clear speech in Croatian and English." Journal of the Acoustical Society of America 118(3), 1677-1688.

Smoorenburg, G. F., 1992. "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram." Journal of the Acoustical Society of America 91(1), 421-437.

Song, J. H., Skoe, E., Banai, K. and Kraus, N., in preparation. "Training to improve speech in noise perception: Biological mechanisms."

Song, J. H., Skoe, E., Wong, P. C. M. and Kraus, N., 2008. "Plasiticity in the adult human auditory brainstem following short-term linguistic training." Journal of Cognitive Neuroscience 20(10), 1892-1902.

Sperry, J. L., Wiley, T. L. and Chial, M. R., 1997. "Word recognition performance in various background competitors." Journal of the American Academy of Audiology 8(2), 71-80.

Studebaker, G. A., 1985. "A "rationalized" arcsine transform." Journal of Speech and Hearing Research 28, 455-462.

Sweetow, R. W. and Henderson Sabes, J., 2006. "The need for and development of an adaptive listening and communication enhancement (LACE™) program." Journal of the American Academy of Audiology 17, 538-558.

Tice, R. and Carrell, T. (1998). Level16. Lincoln, Nebraska, University of Nebraska.

Tremblay, K., Kraus, N., Carrell, T. D. and McGee, T., 1997. "Central auditory system plasticity: generalization to novel stimuli following listening training." Journal of the Acoustical Society of America 102(6), 3762-3773.

Tremblay, K. L. and Kraus, N., 2002. "Auditory training induces asymmetrical changes in cortical neural activity." Journal of Speech, Language and Hearing Research 45(3), 564-572.

Tun, P. A., O'Kane, G. and Wingfield, A., 2002. "Disctraction by competing speech in older adult listeners." Psychology and Aging 17(3), 453-467.

Van Engen, K. J., 2007. "A methodological note on signal-to-noise ratios in speech research." Journal of the Acoustical Society of America 122(5), 2994.

Van Engen, K. J., 2010. "Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble." Speech Communication 52, 942-953.

Van Engen, K. J. and Bradlow, A. R., 2007. "Sentence recognition in native- and foreign-language multi-talker background noise." Journal of the Acoustical Society of America 121(1), 519-526.

Van Wijngaarden, S., Steeneken, H. and Houtgast, T., 2002. "Quantifying the intelligibility of speech in noise for non-native listeners." Journal of the Acoustical Society of America 111, 1906-1916.

von Hapsburg, D., Champlin, C. A. and Shetty, S. R., 2004. "Reception thresholds for sentences in bilingual (Spanish/English) and monolingual (English) listeners." Journal of the American Academy of Audiology 14, 559-569.

von Hapsburg, D. and Peña, E. D., 2002. "Understanding bilingualism and its impact on speech audiometry." Journal of Speech, Language, and Hearing Research 45, 202-213.

Wang, Y., Jongman, A. and Sereno, J. A., 2003. "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training." Journal of the Acoustical Society of America 113(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A. and Sereno, J. A., 1999. "Training American listeners to perceive Mandarin tones." Journal of the Acoustical Society of America 106(6), 3649-3658.

Weil, S. A. (2001). Foreign-accented speech: Adaptation and generalization, The Ohio State University.

Wong, L. L. N., Soli, S. D., Liu, S., Han, N. and Huang, M.-W., 2007. "Development of the Mandarin Hearing in Noise Test (MHINT)" Ear and Hearing 28(2), 70S-74S.

Wong, P. C. M. and Perrachione, T. K., 2007. "Learning pitch patterns in lexical identification by native English-speaking adults." Applied Psycholinguistics 28, 565-585.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M. and Merzenich, M. M., 1997. "Deficits in auditory temporal resolution and spectral resolution in language-impaired children." Nature 387, 176-178.

APPENDIX A

Semantically anomalous sentences (Smiljanic and Bradlow, 2005):

1. Your tedious beacon lifted our cab.
2. A cabbage would sink his tired Tuesday.
3. Their stew was digging a curious bet.
4. My puppy may stress their fundamental gallon.
5. The fast bucket was pecking her twin.
6. The distant budget is baking the sleepy cap.
7. Betty will consist of a tepid token and a pig.
8. Her duplex would tutor a dubious truck.
9. Peter and his chief ticket were hooded by their bed.
10. His kind pudding was taping a decade over my pick.
11. Her dense writer would fork their toga and clerks.
12. The routine body was keeping our wood.
13. The ultimate captain will creak the bottle tomorrow.
14. His grilled cookie derived the baby through a clause.
15. The braided habit twisted her pigeon into the segments.
16. Her abundant pocket circles to his marble.
17. My grill would milk her plump topic with the facts.
18. The ground baggage missed the soda briefly.
19. The theory should drag her home into the ocean.
20. Our rare future submitted a jump to the judges.

APPENDIX B

Semantically anomalous sentences with content words replaced by non-words:

1. Your bedious reakon loofted our bab.
2. A stabbage would bink his vired foozdoy.
3. Their smew was kigging a cunious rett.
4. My ruppy may strace their fustamental gaylin.
5. The fust backet was pessing her twip.
6. The diltant beedgit is yaking the sleeny cass.
7. Fetty will consaust of a lepid roken and a sig.
8. Her suplex would tyter a dunious treek.
9. Piter and his chaif tizzit were hodded by their ged.
10. His kibe pugging was tading a breckade over my pid.
11. Her doonce hiter would ferk their tola and herks.
12. The mootine koddy was kaiping our woob.
13. The iltimate caltin will crike the sottle tosorrow.
14. His grolled kaiky deplied the nabey through a trause.
15. The broaded zabbit twaysted her digeon into the selments.
16. Her aglundant vocket berkles to his marssel.
17. My brill would misk her plamp copic with the foocts.
18. The stound daggage minned the sidda broofly.
19. The geery should slag her heem into the osheff.
20. Our rame duture submodded a jamp to the dudges.