THE ROLE OF SPECTRAL INFORMATION IN FOREIGN-ACCENTED SPEECH PERCEPTION

by

Michelle Rae Kapolowicz

APPROVED BY SUPERVISORY COMMITTEE:

_____

Dr. Peter F. Assmann, Chair

_____

Dr. John H. L. Hansen

_____

Dr. William F. Katz

_____

Dr. Raúl Rojas

Dedicated to Frances A. Kapolowicz

THE ROLE OF SPECTRAL INFORMATION IN FOREIGN-ACCENTED SPEECH PERCEPTION

by

MICHELLE RAE KAPOLOWICZ, BA, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

COGNITION AND NEUROSCIENCE

THE UNIVERSITY OF TEXAS AT DALLAS

December 2017

ACKNOWLEDGMENTS

First, I would like to thank Dr. L. Tres Thompson, without whom I would not be where I am today. I can never repay him for the amount of knowledge that he bestowed upon me both during the time that I worked in his laboratory and in each class that I took with him. He made me fall in love with neuroscience, and he never doubted that someone who had a background in philosophy would be able to learn the techniques and have the mindset needed to succeed in a neuroscience laboratory setting. He also taught me how to use everything from a bandsaw to a soldering iron; I truly attained the most varied skillset during my time in his lab.

I would also like to extend my gratitude to Dr. Van Miller, another important mentor to me at UT Dallas. Although I had other excellent experiences serving as a teaching assistant for various classes, most of my time in this role was spent under the supervision of Dr. Miller. I learned so much from him regarding effectively communicating difficult material to students with a wide variety of academic backgrounds. He also taught me the value of having an ideal work-life balance, such as enjoying fun nights cheering on FC Dallas (DTID)! Most importantly, however, he taught me to realize that having a goal of simply crossing the finish line is, perhaps, more important than having the goal to win. This advice is, single-handedly, why I am able to complete this dissertation.

I would also like to express my deep appreciation for my committee members. Each of them contributed uniquely with helping to steer my research in the right direction, and each provided me with insight that I will continue to apply in my future endeavors. No matter how busy he was, Dr. Hansen always made time for me. I always looked forward to my time with him because I knew that I would, certainly, learn something new. Perhaps most crucially, he taught me how to

see the value of my research beyond simply fulfilling my own scientific curiosity (a necessity when it comes to applying for grant funding). Dr. Katz inspired me with his work on foreign-accented speech production, and he encouraged me to continue challenging myself, which made the the final product of my work stronger. Dr. Rojas kept me calm more times than I can count throughout the most difficult stretches of my dissertation. Through his own research, he also influenced me to understand the perspective of the foreign-accented/bilingual talker. I also appreciated his careful consideration of my talker database, where he provided me with feedback to assure that I would be able to utilize the recordings for future experiments. I want to extend this appreciation to my lab members, past and present, for supporting me in terms of their contribution to my work here at UTD. Not only did I learn so much from each one of them, but I also established true bonds that I hope will last a lifetime. Additionally, I want to thank the many people who participated in this research.

I would also like to thank my friends and family for their continuous support for me throughout my (very long) academic journey. They always told me that they were proud of me for doing what I wanted to do, and they gave me the time that I needed to fulfill this important part of my life. I wouldn't be completing this dissertation with a clean conscience if it weren't for their patience and understanding.

Finally, I would like to thank the chair of my dissertation, Dr. Peter Assmann. Given my interest in speech perception, I took a class from him during my first semester of graduate school. His course was unlike any class I had ever taken. To say it was difficult would be an understatement. As the semester progressed, he continued to exude extreme tolerance for my (often naïve) questions. By the end of the semester, I was astounded by how much I had learned, and how

applicable the information has been throughout my research endeavors. His mentality in the classroom certainly extends to his laboratory. He taught me how to be an independent thinker while also realizing that often our greatest strengths come with teamwork and close collaborations. I cannot express my gratitude enough for all that he has done for me and all that he has taught me. Thankfully, Dr. Assmann is the kind of mentor who will always be there for you, ready to share his wisdom and advice; I know that he will always be just a phone call or an email away. For all who know him, to say that his passion and enthusiasm is contagious is an understatement. As I continue to learn and grow in my field, I think I can safely say that I doubt I will ever attain the amount of knowledge that he has, but I will do my best to make him proud, nonetheless. From the bottom of my heart, thank you for giving me the opportunity to pursue my dissertation under your direction, Dr. Assmann.

September 2017

THE ROLE OF SPECTRAL RESOLUTION IN FOREIGN-ACCENTED SPEECH PERCEPTION

Michelle Rae Kapolowicz, PhD
The University of Texas at Dallas, 2017

Supervising Professor: Dr. Peter F. Assmann

Source signals, vocal tract resonances and articulatory movements encode talker-specific spectral information that allows for appropriate adjustment of a listener's perceptual system to the acoustic characteristics of a particular talker. This implicit learning of talker-specific properties is known as talker normalization. Talker normalization requires prior experience and also structured knowledge about pronunciation variation across talkers that share the same native accent to guide perception. This process becomes difficult when the talker has an accent that is perceived as foreign. Although research suggests that listeners can adapt to foreign accents, the time-course and specificity of adaptation remain unclear, especially when listeners attend to speech produced by multiple alternating foreign-accented talkers. This dissertation focuses on the role of spectral cues in the perception of foreign-accented speech. While many factors contribute to the perception of foreign-accented speech, spectral cues are of particular interest because they play an important role in talker-specific phonetic recalibration in native speech to accommodate variations in vocal tract size across talkers. Through a series of experiments, we tested the hypothesis that listeners rely on talker-specific spectral cues when adapting to foreign-

accented speech. We assessed the contribution of spectral resolution to the intelligibility of foreign-accented speech by varying the number of spectral channels in a tone vocoder. We also tested listeners' abilities to discriminate between native- and foreign-accented speech to determine the effect of reduced spectral resolution on accent detection. Results showed a greater decrease in intelligibility when spectral resolution was reduced for foreign-accented speech compared to native-accented speech. Listeners also found it harder to detect a foreign accent with spectrally reduced speech. We extended these findings by investigating the effects of changing the talker from trial to trial, a manipulation that produces a reduction in intelligibility when compared to holding the talker constant within each block of trials. We hypothesized that limiting spectral resolution when listeners were exposed to multiple foreign-accented talkers would cause a further decrease in intelligibility. This prediction was confirmed, supporting the idea that detailed spectral resolution helps to maintain the intelligibility of foreign-accented speech when listeners are exposed to multiple interleaved talkers. Listeners were able to adapt with increased exposure if they heard a single foreign-accented talker, though not to the extent observed with unprocessed natural speech. Performance was higher for native-accented speech, with no difference between single- and multiple-talker conditions. Finally, we investigated how spectral shifting of foreign-accented speech would affect intelligibility by scaling the fundamental frequency and spectral envelope to simulate multiple talkers. Consistent with results for spectrally reduced speech, intelligibility was lower in the multiple-foreign-accented talker condition compared to the single-talker condition. Introducing frequency shifts produced a drop in intelligibility to levels observed in the multiple-talker condition. Results indicate that listeners depend on spectral cues when perceiving foreign-accented speech, and that spectral information

is especially important when listening to speech spoken by different foreign-accented talkers. The results support a model of foreign-accented speech perception that relies on spectral cues to adjust to the deviations between foreign-accented and native speech.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Foreign-accented speech is classified as non-pathological speech that noticeably differs from native talker pronunciation norms (Munro & Derwing, 1995). Foreign-accented speech often presents with more variability than native-accented speech, such acoustic variability stemming from interactions between a foreign-accented talker's native language and non-native language (Wade *et al*., 2007), as well as variability in speaking rate (Guion *et al*., 2000). Foreign-accented speech affects both segmental and suprasegmental aspects of the signal, which can result in increased processing effort, segmental/lexical ambiguity, and mapping failure on the part of the listener (Anderson-Hsieh *et al*., 1992). Despite such variability, there must be some relative consistency for production of foreign-accented speech that allows listeners to recalibrate their phonemic and/or prosodic categories within the course of a conversation, since it has been shown that native listeners are able to adapt with increased exposure (Clarke & Garrett, 2004; Bradlow & Bent, 2007; Baese-Berk *et al*., 2013). Although research suggests that listeners readily adapt to foreign accents after minimal exposure, listeners still often report difficulty understanding non-native accents, and the time-course and specificity of adaptation remain unclear (Trude *et al*., 2013), especially regarding speech produced by either a single or multiple talkers. This question is of particular importance in speech communication given that non-native talkers outnumber native talkers of English, and communication between these two groups are increasing (Jenkins, 2000; Graddol, 2006).

This dissertation focuses on the contribution of spectral information to the perception of foreign-accented speech. Much of the research on foreign-accent perception has focused on

temporal patterns (*e.g.*, Baese-Berk & Morrill, 2015) as well as lexically-guided information (*e.g.*, Cooper & Bradlow, 2016). Here, we address the role of spectral cues and the relationship between the perception of foreign-accented speech and talker variability. We compare conditions where listeners are hearing the same talker repeatedly as opposed to a succession of different talkers. It is well known that spectral properties contribute to the perception of a foreign accent (*e.g.,* Arslan & Hansen, 1997a); however, it is unclear how spectral cues contribute to the perceptual adjustments to different talkers with a foreign accent. Clarifying perceptual adaptation differences for these particular talker groups (single versus multiple talkers) may help elucidate underlying mechanisms involved in this adaptation process.

Foreign-accented talkers who share a similar linguistic background demonstrate relative consistency in their speech production. This consistency seems to aid native listeners with the perceptual learning process. However, as previously mentioned, non-native talkers tend to be more variable in their pronunciation of their non-native language than native talkers of that language. This entails that non-native talkers only sometimes succeed in producing canonical sounds, and this can vary from moment to moment (*e.g.*, Hanulikova & Weber, 2012). Given this propensity for such variability in pronunciation across foreign-accented talkers (even those sharing the same first language: L1), Bradlow and Bent (2007) investigated whether listeners would benefit more when adapting (as measured by intelligibility) to a particular foreign accent (such as Mandarin-accented English) if they trained multiple foreign-accented talkers. They found that adaptation was similar in conditions where listeners were exposed to the same Chinese-accented talker for training and testing (talker-dependent adaptation) as when listeners heard multiple Chinese-accented talkers for training (accent-dependent adaptation). Bradlow and

Bent also compared adaptation to a novel Chinese-accented talker in a condition when listeners were exposed to a single Chinese-accented talker for training with a condition when listeners trained on multiple Chinese-accented talkers. They revealed that training on a single foreign-accented talker did not generalize as well to a novel foreign-accented talker with the same native language as when training on multiple foreign-accented talkers who share the same L1.

Given the results of Bradlow and Bent (2007) and others, listeners seem readily able to adapt to nonnative talkers and are quite tolerant to their inconsistencies. For instance, Eisner and colleagues (2013) demonstrated that listeners relax their vowel categories more for non-native talkers than for native talkers. In order to better understand whether adaptation is aided simply by listeners being more lax regarding inconsistencies in foreign-accented speech, Witteman and colleagues (2014) tested native Dutch listeners either in a consistent-accent condition (German-accented items only) or in an inconsistent-accent condition (German-accented and native-like pronunciations intermixed). They found that listeners adapted more quickly in the consistent-accent condition compared to the inconsistent-accent condition, although after a short period of additional exposure, listeners were able to account for the variability in the inconsistent-accent condition. The general conclusion drawn from their study is that a single talker who inconsistently speaks with both native-like and non-native-like speech patterns is more difficult to understand, initially, than a single talker who speaks using non-native-like speech patterns consistently. Their results do not provide evidence that listeners, initially, become more lax in their target expectations when listening foreign-accented speech, at least not when perceiving speech from a single talker. Witteman and colleagues found that, with increased exposure, listeners were able to overcome the variability in the inconsistent-accent condition, implying that

listeners can eventually tolerate inconsistency in foreign-accented speech. However, it is important to mention that their study utilized the same talker for both conditions. This point leads to the possibility that listeners could have benefited during the adaptation process by other components unique to that specific talker which are independent of adapting to the accent itself.

To better understand this issue, it helps to consider how listeners readily adapt to variation in native-accented talkers. Speech varies across different talkers and even within the same talker. Much of this variation is spectral. Listeners need to be able to adjust to this varying spectral information, such as perceiving speech spoken by an adult compared to a child. To do this, it is hypothesized that listeners undergo a kind of perceptual calibration, whereby they take the incoming speech signal and map it onto their language-specific categories. These category boundaries are adapted with experience. A theoretical model for how this process might work could be similar to an adjustable template that accounts for spectral variation across talkers (Nearey & Assmann, 2007). This is a complicated process, yet listeners can calibrate relatively easily, depending on the condition. For example, when listening to the phoneme for the average female token for the vowel in /hood/, there is often overlap in F1-F2 space for the average male token for the vowel in /head/. Yet, listeners know that these sounds belong to distinct categories. Conversely, when listeners hear the vowel in /heed/ spoken by a male and a female, the formant patterns are quite different; however, listeners are aware that the two tokens belong to the same category.

Listeners' perception of speech from a single talker improves over time, and this process has been attributed to talker normalization. Talker normalization can be informed by properties of the source (F0, perceived as pitch), vocal tract resonances (formants, which are concentrations

of acoustic energy around particular frequencies in speech) and articulatory movements to encode talker-specific information. This process allows for appropriate adjustment of a listener's perceptual system to the acoustic characteristics of a particular talker.

Weatherholtz and Jaeger (2016) assess the role of talker normalization as an explanation for the many-to-many mapping between acoustic patterns and linguistic categories that forms a core theoretical problem for theories of speech perception. For example, there is a relationship between the length of the vocal tract and resonant frequencies, with men having longer vocal tracts, and therefore lower formants. Also, talkers with the same dialect maintain the same canonical relationships among speech sounds. Weatherholtz and Jaeger argue that talker normalization can explain how the speech perception system copes with acoustic variability by utilizing relational aspects of speech rather than the absolute value of category-relevant speech cues. Much of these relations reside within the spectral information of speech.

Evidence supporting talker normalization has been reported for the perception of vowels (Nearey, 1989), consonants (Summerfield, 1981; Johnson, 1991), whole words (Mullennix *et al*., 1989) and lexical tones (Wong & Diehl, 2003). Normalization has been further described as comprising two distinct mechanisms by Nearey (1989): Intrinsic normalization suggests that each utterance is self-normalizing by virtue of vocal characteristics encoded into the utterance itself. Extrinsic normalization involves using cues derived from preceding context that may help constrain interpretations of a target utterance. A key example demonstrating the need for talker normalization was reported in a landmark study by Peterson and Barney (1952). They recorded the speech of several American English vowels produced by men, women and children, and they found considerable overlap between certain vowel categories (entailing considerable variability

for several vowel categories) in production. Despite the acoustic overlap in the production of different phonetic categories, they found that listeners are quite accurate in maintaining phonetic constancy across talkers.

Although no detailed studies have extended the application of talker normalization to the process of adapting to foreign-accented speech, it is worth considering the potential overlap, as it may be helpful in addressing differences observed when adapting to multiple talkers with a foreign accent as opposed to a single foreign-accented talker. Considering the greater degree of variability in foreign-accented speech, spectral cues may prove to be a consistent source of information which allows adaptation to talker-specific characteristics that can aid listeners' perceptual calibration. This enables listeners to map incoming speech onto their language-specific categories and adapt the category boundaries accordingly with experience *(e.g.*, Nearey & Assmann, 2007).

Listeners are well-accustomed to variations within their native language, where this adaptation process occurs with minimal dependency on fine spectral detail in ideal speech conditions. Given that foreign-accented speech can present as an adverse listening condition, the purpose of this dissertation is to depict whether spectral cues may be critically involved in the perception of foreign-accented speech, both initially and after increased exposure. A second aim is to examine the contribution that spectral cues make when adapting to either the same or different interleaved foreign-accented talkers (who share the same native language) with increased exposure. There are four main components under investigation: The research first assesses the ability of listeners to perceive foreign-accented speech in conditions with limited spectral information, behaviorally testing if listeners show a drop in intelligibility scores when

exposed to foreign-accented speech with limited spectral resolution compared to full spectral resolution. Second, this research examines whether limiting spectral resolution makes it more difficult for listeners to accurately detect whether a talker has a foreign accent. Third, this research asks whether there is an additional cost (a further decrease in intelligibility) associated with perceiving multiple interleaved foreign-accented talkers with extended exposure over time (adaptation) when spectral resolution is limited. Fourth, this dissertation investigates perceptual adaptation to a single foreign-accented talker whose speech has been spectrally shifted (fundamental frequency and spectral envelope) to sound like multiple different talkers who share the same foreign accent. The aim is to compare whether intelligibility scores will align more closely with patterns observed when listeners are trained with a single foreign-accented talker or with multiple foreign-accented talkers.

Together, these experiments will provide a comprehensive investigation into how well the talker normalization approach applies to foreign-accented speech perception. If intelligibility scores are higher when listeners are exposed to a single foreign-accented talker (with full access to spectral resolution) compared to when they are exposed to multiple foreign-accented talkers who share the same L1, then this would be suggestive of talker normalization, since listeners are more easily able to rely on talker-dependent cues. Finally, if spectral shifts applied to a single foreign-accented talker produce a reduction in intelligibility that is comparable to when listeners are exposed over time to unprocessed speech from multiple foreign-accented talkers, that outcome would underscore the importance of spectral information. It would also indicate that listeners were unable to utilize the preserved temporal information to attain intelligibility scores comparable to the unprocessed single-talker condition.

A comprehensive speech corpus was collected in which native and non-native talkers of American English produced several low-context Harvard sentences (IEEE Subcommittee, 1969). The corpus recordings were used to test experimental hypotheses in perceptual listening tests. Test stimuli for all experiments were presented to normal hearing adults who were native talkers of American English. Results from a subset of listeners were used to classify talkers as either native or varying degrees of foreign-accented. Once production data was categorized, another subset of listeners were presented with the same data either as unprocessed speech or as vocoded (spectrally-degraded or spectrally-shifted) speech in order to examine the possible interactions between spectral degradation and foreign-accented speech and spectral shifting and foreign-accented speech. Listeners underwent training in either single-talker, multiple-talker or simulated-multiple-talker conditions to compare differences in perceptual learning across these groups. The overarching goal of this collective research is to better characterize the role of spectral information in adapting to foreign-accented speech over time. This central theme is of particular importance for yielding valuable information for cochlear-implant users whose devices provide reduced spectral resolution and for users of speech recognition devices and Voice over Internet Protocol (VoIP) systems, which can have varying degrees of spectral degradation/distortions.

This dissertation is divided into six chapters, consisting of an introduction (Chapter 1) followed by four chapters covering separate major phases of work (Chapters 2-5) and a final chapter summarizing overall findings and conclusions (Chapter 6). Chapter 2 describes the carefully controlled speech corpus consisting of native- and foreign-accented talkers who were classified in terms of accentedness and intelligibility. Chapter 3 details listening experiments

conducted to investigate the role of spectral resolution on initial intelligibility and accent detection across different spectral channels. Chapter 4 explains the perceptual experiments designed to test how well listeners can adapt to foreign-accented speech with limited spectral resolution in single- and multiple-talker conditions. Chapter 5 discusses research designed to test listeners exposed to frequency-shifted foreign-accented speech over time. The remainder of this introductory chapter is organized as follows: First, a brief literature review covering foreign-accented speech perception is presented; second, research concerning the role of spectral resolution in speech perception is discussed; third, a brief review of research spanning the effects of spectrally-shifted speech on perception is presented; fourth, the research plans for examining the role spectral information on perception of foreign-accented speech is introduced; section 1.4 also discusses the innovation and novelty that this research can offer to advance previous work in this area.

1.1 Perception of foreign-accented speech

Talker normalization is a theory originally advanced to explain how listeners achieve perceptual constancy across variation in age, sex, and size of talkers (Liberman *et al*., 1967; Strange *et al*., 1983). Listeners attempting to resolve such acoustic-phonetic ambiguities feel that a greater cognitive load (increased listening effort) is needed for perceptual accuracy (Eckert *et al*., 2008; Heinrick & Schneider, 2011; Mattys *et al*., 2012; Rönnberg *et al*., 2013; Cousins *et al*., 2014). Similar experiences of increased listening effort are reported for people listening to foreign-accented (non-native) speech (Lane, 1963; Munro & Derwing, 1995; Van Wijngaarden, 2001; Floccia *et al*., 2006; Van Engen & Peelle, 2014). Despite increased listening effort,

adaptation to foreign-accented speech has also been demonstrated *(e.g*., Clarke & Garrett, 2004; Bradlow & Bent, 2007; Baese-Berk *et al*., 2013; Witteman *et al*., 2013; Reinisch & Holt, 2014).

Prior to studying perceptual improvements occurring over time (adaptation), studies addressed the initial struggle that listeners face when perceiving foreign-accented speech. Lane's instrumental work (1963) compared intelligibility in noise of English words produced by a native English talker and three foreign-accented talkers and found that non-native talkers were about 35% less intelligible. Although he tested different signal-to-noise ratios (SNR), he found no interaction between factors of SNR and talker. He suggested that there was no interaction between linguistic (speaker-related) distortion and environmental distortion (noise). Following this finding, studies have continued to use varying levels of background noise when researching foreign-accented speech perception.

Later, Munro (1998) compared the effects of noise on the intelligibility of sentences produced by native English talkers and Mandarin-accented English talkers. He found that the intelligibility of Mandarin-accented talkers, who ranged in level of accentedness from moderate to strong, was lower than native English talkers in both quiet and noise, although there was a large degree of variability across talkers. A more specific examination of the effect of noise on foreign-accented speech found that sentences produced by non-native talkers required 3 decibels (dB) greater SNR for 50% intelligibility than sentences produced by native talkers (Van Wijngaarden, 2001). Another study conducted by Rogers and colleagues (2004) found that intelligibility in quiet for foreign-accented speech was only about 7% lower than intelligibility in quiet for native-accented speech; however, this high-proficiency foreign-accented speech was less robust than native-accented speech when presented to listeners under three different levels of

background noise. These findings illustrate how perceptually fragile foreign-accented speech can be for a listener in sub-optimal conditions unlike when perceiving native speech.

An interesting phenomenon has also been observed in listeners perceiving synthetic speech presented in competing background noise. Several studies reveal that synthetic speech-in-noise is more prone to perceptual degradation than natural speech-in-noise, even when performance in quiet for synthetic speech was within 10% of natural speech in quiet (Pisoni & Koen, 1981; Luce *et al*., 1983). Given the similarities in perceptual performance for foreign-accented speech and synthetic speech, it could be that listeners initially struggle with a mismatch in expected spectral information compared to what they actually hear (*e.g*., non-native or non-natural frequency shifts). This initial target mismatch would take time for perceptual adjustments to occur. In both cases of perceptual distortion (synthetic speech and foreign-accented speech), despite the listener's initial struggle, especially in the presence of competing noise, the listener can adapt (*e.g*., for foreign-accented speech: Clarke & Garrett, 2004; Bradlow & Bent, 2007; Baese-Berk *et al*., 2013; *e.g*., for synthetic speech: Hervais-Adelman *et al*., 2008; Bent *et al*., 2011).

Regarding adaptation to foreign-accented speech, Clarke and Garrett (2004) focused their study on the time-course underlying the adaptation process, showing how this is a rapid perceptual process, with the initial perceptual deficit taking about 1 minute with exposure to 2-4 sentences to diminish. Clarke and Garrett only measured listeners' exposure to a single foreign-accented talker, however, which implies adaptation to the talker more so than to the accent itself. As discussed previously, Bradlow and Bent (2007) took this potential limitation into account and looked at both talker-dependent (where listener's heard the same foreign-accented talker with

increased exposure) and talker-independent/accent-dependent (where listeners heard different talkers who shared the same foreign-accent over time) perceptual adaptation to foreign-accented speech presented with white noise at a +5 dB SNR. Their results revealed that listeners could adapt in both conditions, although the underlying mechanism could differ. Going a step further, Baese-Berk *et al*. (2013) compared accent-dependent and accent-independent adaptation to foreign-accented speech also embedded in white noise at a +5 dB SNR. They found that listeners were able to adapt to novel accents after increased exposure to talkers with different foreign-accents. This result suggests that generalization of foreign-accent adaptation is the result of exposure to systematic variability in accented speech that is similar across talkers from multiple language backgrounds, a finding that would be better explained by an examination of lexical features and word duration (*e.g*., Baker et al., 2011) or speaking rate consistency (Morrill *et al*., 2016) rather than talker-specific spectral cues.

Contrary results were presented by Bent and Holt (2013). They showed that there was a detriment to word recognition in a multiple-foreign-accented talker condition when talkers shared the same L1 compared to when listeners heard foreign-accented speech from the same talker over time. They also found that there was a further detriment when listeners heard speech from multiple talkers with different foreign accents compared to when listeners heard speech from multiple foreign-accented talkers who share the same L1. Their results demonstrate that the processing of foreign accent variation may influence word recognition in ways similar to other sources of variability, such as speaking rate or style, therefore including multiple foreign accents can result in a significant performance decrement beyond the multi-talker effect.

These results share similar aspects of the learning process proposed in the perceptual normalization hypothesis originally purported for adapting to varying characteristics across talkers *(i.e.*, males, females, and children). In the case of adapting to foreign-accented speech, listeners are able to perceptually adjust their expectations based on characteristics of a talker's identity and also for characteristics regarding the talker's deviation in pronunciation from native speech (adjusting to the accent, itself), thereby resulting in faster and more accurate speech recognition over time; although, adaptation time may be limited with conditions of multiple talkers and/or multiple accents. These results suggest that it may be easier for listeners to normalize to a single talker; however, listeners can still use similar spectral cues to a certain extent across talkers who share the same foreign accent. The mapping between the acoustic signal and phonemic categories are more consistent when listening to talkers who share the same foreign accent (Bent & Holt, 2013).

As aforementioned, multiple studies have also explored adaptation to synthetic speech. A study by Hervais-Adelman and colleagues (2008) examined perceptual learning of noise-vocoded words, which removes spectral detail from speech, and they found a significantly greater improvement in comprehension if listeners were trained on clear speech prior to hearing distorted speech compared to listeners who heard distorted speech before clear speech. This perceptual learning generalized to untrained words as well, suggesting involvement of phonological short-term memory and top-down processes in the perceptual learning of noise-vocoded speech. The researchers proposed that a similar process aids comprehension of foreign-accented speech and speech perception following implantation of a cochlear implant.

Bent *et al*. (2011) aimed to test what level of information could be driving the process of adaptation to degraded speech and exposed listeners to sine-vocoded speech in German, English or Chinese, then tested them with sine-vocoded English. They found that training with vocoded German speech was as effective as training with English-vocoded speech when paired with visual stimuli, but training using Mandarin-vocoded speech paired with visual stimuli was not as beneficial. Results suggest that full lexical access is not necessary for adaptation to degraded speech, but training in a language that is phonetically similar to the testing language can facilitate adaptation.

Given the complexity of speech, several factors are evidently involved in adapting to foreign-accented speech, which makes narrowing down a single potential underlying mechanism difficult. For this reason, it is necessary to continue understanding how foreign-accented speech processing interacts with other sources of variations in speech. As aforementioned, studies have already addressed the interaction of foreign-accented speech and competing noise. This prior research offers the benefit that it resembles realistic listening conditions, as speech is rarely presented in quiet settings. Exposing listeners to foreign-accented speech presented in competing noise also allows the avoidance of ceiling effects in performance. Useful information has been gained from previous research, such as having a better understanding of the ability of a listener to parse out extrinsic perceptual distortions from target speech signals in foreign-accented speech. Further investigation is still needed, however, to better characterize the underlying perceptual learning process, itself. For example, it is still unclear if this adaptation process functions similarly to the explanation purported by talker normalization, namely talker

normalization for native speech occurs much more quickly when adapting to a single talker compared to multiple talkers (Assmann *et al*., 1982; Mullennix *et al*., 1989).

The present research aims to better understand this adaptation process when there are two sources of auditory perceptual distortions which are intrinsic to the signal itself, namely the presence of a foreign accent and altered spectral content. The spectral distortions examined here result from two types of manipulations: spectral contrast reduction and frequency shifting. Information gained from the results of these experiments help to elucidate the role of spectral cues in adapting to foreign-accented speech by normal-hearing listeners. It may help to provide insight into the added difficulties faced by cochlear implant (CI) users, who may have intact temporal processing but limited spectral resolution (Shannon, 1989; 1992) and frequency-place mismatch (Rosen *et al.*, 1999).

Upon characterizing the interaction of spectral information and foreign-accented speech perception, this research also compares adaptation to foreign-accented speech following exposure to either a single (talker-dependent adaptation) or multiple talkers who share the same L1 (accent-dependent adaptation). Uncovering the role of talker differences will further clarify how this adaptation process is related to talker normalization. The next section will briefly elaborate on some important findings in the literature regarding the role of spectral resolution in speech perception. The following section will briefly discuss how spectral shifting can affect the perceptual process in general, and how this may pertain to the perception of foreign-accented speech.

1.2 Perception of speech with limited spectral resolution

Spectral resolution in hearing, as defined by Winn and Litovsky (2015), is the perceptual ability of a listener to distinguish between sounds that differ in pitch or other qualities in the spectral (frequency) domain. Spectral resolution is important for speech perception in that several speech sounds are distinguished by spectral cues such as the frequency of formant peaks. It is, however, widely known that speech recognition of native speech in quiet can be understood almost perfectly with limited spectral resolution. Speech presented to normal-hearing listeners through CI simulations and speech presented to hearing-impaired listeners with CIs represent two situations where listeners are faced with reduced spectral information. Studies comparing these conditions have demonstrated high accuracy when speech sounds are presented in quiet, but lower accuracy when listeners are faced with adverse listening conditions, such as background noise (*e.g.*, for normal-hearing listeners exposed to CI simulations, see Shannon *et al*., 2004; for CI users, see Faulkner & Pisoni, 2013). Although spectral resolution is degraded in the population of CI users, there can be multiple contributing factors, such as the limited number of place-specific electrodes in the cochlea, electrical channel interaction, and history of deafness (Winn & Litovsky, 2015). Given the associated factors affecting performance in CI users, it is difficult to ascertain exactly how reduced spectral resolution alone affects speech perception under various listening conditions.

CI simulations provide a method to directly assess the role of spectral resolution in speech perception. These studies test normal-hearing listeners for whom spectral resolution is explicitly controlled using a vocoder to vary the number of physical channels available (Shannon *et al*., 1995; Rosen *et al*., 1999; Green *et al*., 2007). In order to test how limiting the spectral

information available to the listener interacts with competing background noise, Shannon *et al*. (2004) tested normal-hearing listeners on a series of conditions with varying degrees of available spectral information. They showed that, although simple sentence recognition in quiet can be achieved with minimal spectral information, more complex materials require a greater amount of spectral information to obtain similar levels of recognition in quiet. Researchers have also looked at processing of second language (L2) learners in conditions with reduced spectral resolution and found that even listeners who were completely fluent in their L2 required more channels for accurate speech recognition than native, monolingual listeners (Padilla & Shannon, 2002).

There is limited evidence regarding the interaction of reduced spectral information and foreign-accented speech, but two studies have shown that CI users struggle with foreign-accented speech perception (Ji *et al*., 2014; Tamati & Pisoni, 2015). Both studies reveal that CI users display a larger deficit in perception of foreign-accented speech than normal-hearing listeners. These studies suggest the importance of spectral information for accurate speech perception when listening to foreign-accented talkers. The research presented here examines the role of spectral information in perceiving foreign-accented speech by testing normal-hearing listeners using vocoders, thereby avoiding the variable factors associated with CI listeners. It was expected that normal-hearing listeners would initially struggle under conditions with limited spectral resolution given results from similar studies testing CI users. It was also predicted that some degree of perceptual adaptation to foreign-accented speech could occur, but it would be more limited in conditions where listeners heard more than one foreign-accented talker over time. This dissertation, therefore, not only examines how limiting the availability of spectral resolution in foreign-accented speech can impair initial speech perception but also how reduced

spectral information can delay or even prevent adaptation to foreign-accented speech in a multiple-talker condition.

The following section will offer an overview of the literature regarding how spectrally-shifted speech can affect the perceptual process in general, and how this might pertain to perceiving foreign-accented speech. The next section will also discuss how spectrally shifting the speech of a single foreign-accented talker could simulate the effects observed when listeners are exposed to multiple foreign-accented talkers, thereby further uncovering the role of spectral information in this adaptation process and better clarifying whether or not this process is akin to perceptual normalization.

1.3 Perception of spectrally-shifted speech

Perceptual adjustments to stable acoustic properties in various listening contexts aid the listener with understanding foreign-accented speech. Foreign-accented speech is less canonical than native speech, making this adaptation process more difficult. If the underlying mechanism related to this perceptual learning process is related to talker normalization, then it would seem that adapting to a foreign accent would be much easier if listeners were exposed to a single talker rather than multiple talkers of the same foreign accent. However, as previously discussed, adaptation can occur for listeners exposed to either a single or multiple talkers of the same foreign accent, as well when exposed to talkers with several different foreign accents. Also, in native speech, it has been shown that talker information is not needed to produce spectral contrast effects, as sine tones can also produce spectral contrast effects rather than having a speech context precede the target sound (Holt, 2006; Huang & Holt, 2012). Similarly, it was

shown that manipulating F1 or F3 in a speech context to induce the percept of different talkers could also produce spectral contrast effects; however, manipulating talker identity in the F1 region failed to produce this effect (Liang *et al*., 2012). It, therefore, seems plausible that talker normalization, like adaptation to foreign-accented speech, may not necessarily be 'talker specific.' To further address this question, however, it is worth considering other situations involving the role of spectral information in speech intelligibility, particularly in scenarios showing perceptual decline.

Speech intelligibility reduction occurs when spectral envelope scale factors are increased or decreased (spectrally/frequency-shifted) relative to the unshifted original. This can occur in CI users, a population with impaired access to fundamental frequency and formants. It is unclear how this impacts foreign-accented speech intelligibility. Men's vowels are more susceptible to downward shifts compared to vowels spoken by women and children, while children's vowels show a greater decline with upward shifts, suggesting that the performance decline is associated with the absolute ranges of the formant frequencies or related features of the spectral envelope across age/sex classes (Assmann & Nearey, 2008). These results are generalizable to the perception of connected speech as well (Assmann & Nearey, 2007). Studies of frequency-shifted speech have found that intelligibility improves with extended exposure (Rosen *et al*., 1999; Nogaki *et al*., 2007).

Fundamental frequency and average formant frequencies provide important cues for indexical properties such as the age, sex and size of the talker. Hillenbrand and Clark (2009) have shown that using vocoded adult voices with upward scaling of the fundamental frequency and formants increases the likelihood that a voice will be perceived as "female" while downward

scaling increases the probability that a voice will be heard as "male." These results were replicated for adults and also extended to children's voices. Swapping both the fundamental frequency and the average formant frequencies for children's voices did not consistently induce a change in perceived speaker sex, suggesting that other cues are involved (Assmann *et al*., 2014, 2015). The same experimental design was utilized in a follow-up study by Guest *et al*. (2016) investigating the perception of voice gender in CI simulations compared to unprocessed conditions, and results indicated that listeners in CI-simulated conditions had, overall, lower performance accuracy for both gender and age perception, and all older voices were more likely to be heard as male in these conditions. Also, age estimation error was much higher for CI-simulated conditions compared to unprocessed conditions. The results are directly relevant to CI users, as this population can encounter issues with frequency-place mismatch due to the insertion depth of the CI device's electrodes within the cochlea. These results also indicate that the normalization process may not be as direct as previously indicated.

In order to further investigate if talker normalization may be more directly involved when perceiving foreign-accented speech, the fifth chapter of this dissertation extends the experimental findings observed in Chapter 4 (comparing perception of foreign-accented speech in single- and multiple-talker conditions) to examine how spectrally shifting foreign-accented speech from a single talker to simulate five different talkers who share the same foreign accent interacts with intelligibility scores. Talker normalization indicates that listeners are sensitive to talker-specific spectral cues. Assuming the potential relevance of talker normalization to perception of foreign-accented speech, the expected result for this particular experiment was that listeners would be sensitive to changes in talker-specific spectral information. Listeners would, therefore, perform

worse in a simulated-multiple talker condition since spectral information from a single foreign-accented talker was manipulated despite temporal information remaining intact. Outcomes from this research will further uncover the role of spectral information in perceptual adaptation to foreign-accented speech and better elucidate the potential application of the theory of talker normalization in foreign-accented speech perception.

1.4 An overview of conducted research

Variability in speech is pervasive, especially regarding foreign-accented speech, where listeners incur additional processing costs. Although listeners can adapt to either a single foreign-accented talker or even multiple talkers with the same foreign accent, it becomes more problematic under adverse listening conditions, such as in the presence of background noise. Given the literature detailing the importance of spectral information on perceptual normalization (a process designed to explain the listener adaptation process that occurs across multiple talkers in a general sense), the experiments described in this dissertation examine the importance of spectral cues when listening to foreign-accented speech. Specifically, this dissertation first describes the methods used to build a carefully controlled talker database consisting of native- and foreign-accented talkers in order to test hypotheses regarding the importance of spectral information when perceiving foreign-accented speech. Experiments were conducted to examine how well listeners can initially understand foreign-accented speech and detect a foreign-accent under conditions with minimal spectral resolution available by using a tone vocoder to limit the number of spectral channels. Experiments followed to investigate the ability of listeners to adapt to a single and multiple foreign-accented talkers under conditions with reduced spectral

resolution. Finally, experiments were conducted to test the condition whereby speech from a single foreign-accented talker had been spectrally-shifted to sound like five different talkers who share the same foreign accent in order to gain additional insight regarding the influence of talker variability and spectral information. Combined results help to understand how closely the perceptual learning process for foreign-accented speech perception can be accommodated by the talker normalization theory, which was originally proposed to explain how listeners adapt to variations in native-accented speech (*e.g.*, due to age and sex). It was predicted that limiting spectral cues would further impair perception of foreign-accented speech, similarly to how listeners rely on spectral cues when native speech is presented in other adverse conditions (*e.g.*, background noise). It was also predicted that decreased spectral resolution would present an additional constraint on listeners' abilities to adapt to multiple foreign-accented talkers. However, it should also be considered that listeners could, additionally, rely on temporal cues in the speech signal; therefore, it would be predicted that when listeners heard foreign-accented speech spoken by a single-talker, but spectrally-shifted to sound like several different talkers, listeners could still adapt similarly to when hearing unprocessed speech from a single foreign-accented talker. In other words, spectrally shifting speech from a single foreign-accented talker to sound like multiple talkers may not be enough to simulate the expected detriment observed when listeners heard unprocessed speech from multiple foreign-accented talkers. If talker normalization is a process underlying foreign-accented speech perception, then an alternative prediction assumes that listeners would be sensitive to changes in the spectral domain; therefore, perceptual patterns would be more similar to patterns observed when listeners heard foreign-accented speech from multiple talkers.

A corpus of speech recordings was collected to evaluate these hypotheses. Each talker produced 100 low-context Harvard sentences. The linguistic background of the two talker groups consists of American English talkers from Texas and Mandarin-accented English talkers from Taiwan. Low-context sentences were chosen to simulate a more accurate *listening* task (without using nonsense syllables), since it is more difficult for listeners to predict one word from another word in any given Harvard sentence. The recordings were produced using two elicitation methods: talkers heard each sentence being produced by a male native talker of American English, and they were also presented with each sentence on a computer monitor. They were asked to repeat the sentence that they heard/read. This corpus is valuable for controlling for variations in linguistic backgrounds (as foreign-accented talkers filled out a questionnaire to assure that they had predominantly been exposed to English in Texas and Taiwan, and their primary language is Mandarin; native talkers filled out a questionnaire to assure that they were all monolingual and had only ever resided in Texas). Listeners were recruited to judge the collected speech samples by typing the words that they heard as a measure of intelligibility (percent of keywords correctly identified) for each talker, and listeners judged the level of foreign-accentedness of each talker by using a 9-point Likert scale. Additional information regarding the speech corpus can be found in Chapter 2.

Following collection and analysis of the speech corpus, experiments were conducted to investigate the effects of limited spectral resolution on foreign-accented speech perception. We used a tone vocoder to test the contribution of spectral resolution across different channel conditions on intelligibility of native- and foreign-accented speech and on the ability of listeners to detect a native or foreign accent. The subset of native talkers included for the planned

experiments were the top six talkers, as rated by their intelligibility scores in quiet (for unprocessed speech), and these talkers were all rated as having no foreign accent. The subset of foreign-accented talkers included the bottom six talkers, as rated by their intelligibility scores in quiet (for unprocessed speech), and these talkers were all rated as having a medium-to-heavy foreign accent. It was predicted that intelligibility levels and ability to discriminate between native- and foreign-accented speech would decrease with less spectral resolution available to the listener. We established the number of spectral channels that would be suitable for comparison across conditions based on the number of channels needed for listeners to reliably detect a foreign-accent, yet still provide room for adaptation (as measured by an improvement in intelligibility scores) to occur. Separate experiments examined perceptual adaptation to foreign-accented speech in single- and multiple-talker conditions with limited spectral resolution. It was predicted that listeners would adapt to a single talker more easily than to multiple talkers in this condition. An alternative possibility was also considered whereby, with limited spectral resolution, listeners might find that adapting to multiple talkers is more beneficial because this condition could allow for more opportunities to adapt by depending on canonical temporal cues, which are less affected by reducing spectral resolution. Additional information regarding these experiments is given in Chapters 3 and 4.

Final experiments were conducted to further investigate the role of spectral information by testing listeners' abilities to adapt to spectrally-shifted foreign-accented speech. Speech stimuli were taken from the speech corpus, and listeners were tested in an adaptation condition designed to mimic a simulated-multiple talker situation by spectrally shifting a single foreign-accented talker's fundamental frequency and average formant frequencies to sound like varying

talkers (*e.g.*, Hillenbrand & Clark, 2009). Results from this condition were compared with conditions where listeners were either exposed to a single foreign-accented talker's unprocessed speech over time or to unprocessed speech from multiple foreign-accented talkers over time. Based on results presented by Assmann and Nearey (2008), it was predicted that, if the fundamental frequency and formants are scaled up or down in the same direction, thereby maintaining "natural-sounding" speech, perceptual accuracy for intelligibility scores would improve with increased exposure. Additional detail describing these experiments can be found in Chapter 5.

The research conducted represents a comprehensive study examining the role of spectral information when perceiving foreign-accented speech in single- and multiple-talker conditions, both initially and with increased exposure. Results from this work can not only better inform us of the mechanisms that listeners rely on when perceiving foreign-accented speech but also aid in our understanding of specific situations when spectral information can be limited/distorted, such as with cochlear implant devices, when using voice recognition devices, and when using VoIP systems.

CHAPTER 2

NATIVE AND NON-NATIVE SPEECH PRODUCTION

Abstract

A speech corpus is often employed for several speech perception experiments. Researchers should carefully consider the task of the planned experiments, and decide whether or not to utilize a previously existing speech database or to build their own. Building a speech database can be a timely endeavor, involving collection of several speech samples and often from several different talkers. These recordings are usually further categorized by a separate set of listening experiments and/or acoustic analyses. The advantage of building a speech database is that it allows for researchers to plan for a carefully controlled set of stimuli that can minimize potential confounds that can arise from using pre-existing databases. Here, a speech corpus of low context sentences recorded from native and non-native talkers of American English is presented. This corpus matches the dialect of the native talkers to the dialect of the listeners who participated in all experiments, and it also carefully controls for the demographic of the foreign-accented talkers.

2.1 Introduction

A speech corpus (consisting of recordings of 100 phonetically balanced and semantically meaningful low-context Harvard sentences (IEEE Subcommittee, 1969) produced from native and non-native talkers of American English) was created and utilized in all listening experiments comprising this dissertation. For the experiments presented in this dissertation work, Chinese-accented American English talkers were chosen to represent the non-native (foreign-accented)

talkers because this is the second-to-largest foreign-accented population in the United States (U.S. Census Bureau, 2013).

This speech corpus carefully controls for dialect differences in Mandarin Chinese and American English. Non-native speech recordings consisted of native talkers of Mandarin Chinese who have only ever resided in Taiwan prior to living in Texas. Native speech consisted of native talkers of American English who have only ever resided in Texas. The native talkers were chosen to match the linguistic background of our native listeners, namely monolingual American English talkers who have only ever resided in Texas. These talkers also shared a linguistic background with the majority of the native English talkers interacting with our non-native talkers.

2.2 Method and procedure

2.2.1 Speech materials

Audio recordings of Harvard sentences (IEEE Subcommittee, 1969) were obtained from 15 native, 18 Mandarin-accented talkers of American English and 4 Farsi-accented talkers of American English (age range: 18-47). All Mandarin-accented talkers were born in Taiwan and had only ever lived in Taiwan and Texas, and all were native talkers of Mandarin. All Farsi-accented talkers were born in Iran and have only ever lived in Iran and Texas. All non-native talkers were students at The University of Texas at Dallas with a range of 2 weeks to 22 years residency in Texas. They were paid a nominal fee for producing the recordings. Native talkers were college students who were recruited from The University of Texas at Dallas, School of Behavioral and Brain Sciences' undergraduate research participation pool through an online

research credit sign-up system and were awarded research credit for participation. Native talkers had only ever resided in Texas and were monolingual. Both groups (native and non-native) were given a brief hearing screening and further screened with a questionnaire to assure that they had no hearing impairments and that they fit the criteria for inclusion.

Talkers were instructed to repeat each sentence after listening to the sentence spoken by a male native talker of American English and viewing a transcript of the sentence on a computer monitor. Recordings were made in a sound-attenuated booth using a Shure SM-94 microphone, Symetrix SX202 dual-microphone pre-amplifier and Tucker-Davis Technologies data acquisition hardware (MA1, RP2.1). Digital waveforms were stored on a computer disk at a rate of 48 kHz and 16-bit resolution. Sentences were RMS-equalized across all talkers. Procedures for talkers were reviewed and approved by The University of Texas at Dallas Institutional Review Board.

2.2.2 Listeners

Listeners recruited for the talker group assignment task (see section 2.2.3 below) were monolingual, native English-speaking college students who had only ever resided in Texas (age range: 18-26). Listeners were awarded research credit for participation and were screened for hearing impairments and linguistic background. Sentences from recorded talkers were presented in quiet through Sennheiser HD 598 headphones at a comfortable level in a sound booth. Listeners only heard two sentences from each talker, and no sentence was be repeated. Presentation of talkers and sentences were randomized to prevent talker adaptation. Listeners were asked to listen to each sentence and type the words that they heard and rate the degree of foreign-accentedness using a 9-point Likert scale (with 1 being *no foreign accent* and 9 being

heavily foreign-accented). Procedures for listeners were reviewed and approved by The University of Texas at Dallas Institutional Review Board.

2.2.3 Talker group assignment

Intelligibility scores for each talker were based on the percent of keywords correctly heard based on what listeners typed. To minimize subjective scoring, an automated scoring program was implemented in Matlab version R2016b (The MathWorks USA) which removes the non-keywords from listeners' responses (such as article adjectives) and compares the remaining words to the corresponding transcripts for these sentences. This program also largely accounts for misspelled/mistyped words by including a comprehensive list (over 4000 cases) of commonly misspelled/mistyped American English words. To maintain consistency across experiments, this comprehensive list was not changed; however, it has the added advantage that it can be adapted/updated at any time to include newer renditions of misspelled words/typos. In addition, the scoring program counts homonyms, such as read/red or bare/bear correctly, since the listener would hear the same sound regardless of the case.

The average intelligibility score for each talker was used to provide intelligibility rankings. The bottom six Mandarin-accented talkers (in terms of each talker's average intelligibility score) were chosen to represent the foreign-accented talker group, and the top six native talkers of American English were chosen for the native talker group. Farsi-accented talkers were excluded from the subset of foreign-accented talkers utilized in experiments presented in the remainder of this dissertation because this talker group was only included to add additional accented talker stimuli to prevent listeners from adapting to Mandarin-accented English during the listening procedure described above. The experiments described in the

remainder of this dissertation were designed specifically to analyze perception of talkers who share the same L1, as a mixed-foreign-accented talker group presents additional considerations that are beyond the scope of this work.

2.3 Innovation

This speech corpus was designed to carefully control for linguistic variability in both talker groups and uses low-context sentences intended to make word prediction more difficult for the listener. Although several American English corpuses exist for both native talkers of American English and foreign-accented talkers whose native language is Mandarin, this corpus is the first based on the Harvard sentences to control more strictly for dialect differences that could introduce additional variability into the results collected during this research by minimizing variations in the linguistic backgrounds of our talkers. Most pre-existing speech databases do not account for language background and demographics of Chinese talkers. Specifically, there are various Chinese "dialects" such as Mandarin versus Cantonese. Moreover, there are variations within the Mandarin dialect: Standard Mandarin spoken in China is dialectically different from Mandarin spoken in other countries (*e.g.*, Taiwan). The standard Mandarin taught in China, known as Putonghua, varies from the standard Mandarin taught in Taiwan, known as Guoyu. These dialects vary in their mutual intelligibility, for example, Chang and Fox (2010) reported different durational patterns of tones emerging for these two groups. The result of such surface acoustic variations in linguistically identical categories can result in perceptually ambiguous tones. Such variations can influence production of L2 categories and should be controlled for in perception studies of foreign-accented speech. This speech database

overcomes the potential problems that can be introduced due to such linguistic variations by including only talkers who learned the standard Mandarin spoken in Taiwan (Guoyu). This corpus also controls for the influence of the dialect used by native English talkers by including only native-accented talkers who have only ever resided in the same geographical location with whom the non-native talkers are now immersed. Also, this speech corpus is versatile in that it can be utilized not only for the experiments presented in this dissertation work but also for future studies, such as experiments involving carefully controlled acoustical analyses of Mandarin-accented English or in various other auditory perceptual tasks concerning foreign-accented speech.

2.4 Results

Speech recordings were obtained from 15 native and 18 non-native talkers of American English, resulting in a total of 3,600 sentences from non-native talkers and 3,000 sentences from native talkers, with a total of 200 sentences per talker. Listeners provided intelligibility scores for the speech recordings from foreign- and native-accented talkers. The intelligibility scores consisted of the percent of keywords correctly heard, which was based on what the listeners typed. The average intelligibility score for each talker was used to provide intelligibility rankings. Listeners were also asked to rate the degree of foreign-accentedness using a 9-point Likert scale, with 1 being *no foreign accent* and 9 being *heavily foreign-accented*. The correlation between accentedness and intelligibility were analyzed using a simple linear regression model.

Analyses were performed on these recordings in order to categorize the talkers in terms of their average intelligibility scores and their average degree of foreign-accentedness across listeners. Statistical analyses were performed using the R programming language (R Core Team, 2016). Talkers that were utilized in experiments for this dissertation were chosen from the bottom six non-native talkers (4 females, 2 males), and the top six native talker (4 females, 2 males) for comparisons between groups. The subset of non-native talkers chosen for inclusion had a mean intelligibility score of 66% in quiet. The subset of native talkers chosen for inclusion had a mean intelligibility score of 96% in quiet. For all non-native talkers, intelligibility scores and foreign accent ratings were correlated, $r = -0.82$, $p < 0.01$. Results displayed in Fig. 2.1. As expected, for all native talkers, there was a floor effect for foreign accent ratings and a ceiling effect for intelligibility, which indicates that these talkers were highly intelligible with no foreign accent detected, $r = -0.48$, $p = 0.07$. Results displayed in Fig. 2.2. Experiments in Chapter 3 utilized six talkers per accent condition, and experiments for Chapters 4 and 5 consisted of five talkers per accent condition (with one additional native-accented talker for practice sessions).

Additional analyses were performed to investigate whether accent ratings and intelligibility scores for Mandarin-accented talkers were correlated with biographical data obtained from the talkers. Appendix B includes a sample of the questionnaire used to obtain biographical data on all talkers. A multiple regression analysis was performed using the lm() function in R to predict the intelligibility scores (percent keywords correctly typed) from the talker's gender, weeks resided in the United States of America (USA) and their interactions with native English talkers. The analyses resulted in the following non-significant factors: gender, weeks resided in the USA and interactions with native English talkers. In addition, the overall R^2

value was found to be 0.10, indicating that the model was able to account for only 10% of the total variance in intelligibility scores. Likewise, a multiple regression analysis was performed to predict accent ratings (9-point Likert scale with 1 indicating no foreign accent and 9 indicating a heavy foreign accent) from the talker's gender, weeks resided in the USA and their interactions with native English talkers. The analyses resulted in the following non-significant factors: gender, weeks resided in the USA and interactions with native English talkers. Furthermore, the overall R^2 value was found to be 0.04, indicating that the model was only able to account for 4% of the total variance in accent ratings. These results indicate that intelligibility scores and accent ratings of the speech stimuli recorded from the foreign-accented talkers included in this database were independent of gender, length of residence in the USA and interactions with native talkers. It is, however, important to note that there was a trend towards significance for one factor: the frequency with which foreign-accented talkers interacted with native talkers. Specifically, there was a trend towards these talkers being rated as having a heavier foreign accent if they interacted less with native English talkers. Inclusion of additional talkers in the database could have further impacted these results.

2.5 Conclusions

These results offered a carefully controlled speech database, where sentences from a subset of talkers (6 native- and 6 foreign-accented) were utilized for the remaining experiments presented in this dissertation. The subset of talkers chosen for the remaining experiments represent a homogeneous talker demographic for each accent condition: Native talkers were rated as having no foreign accent with high intelligibility scores in quiet. Non-native talkers (also

known as foreign-accented talkers) were rated as having a heavy foreign accent with intelligibility scores that were in an ideal range to allow for perceptual improvement to occur with increased exposure without the likelihood of listeners achieving ceiling performance levels.

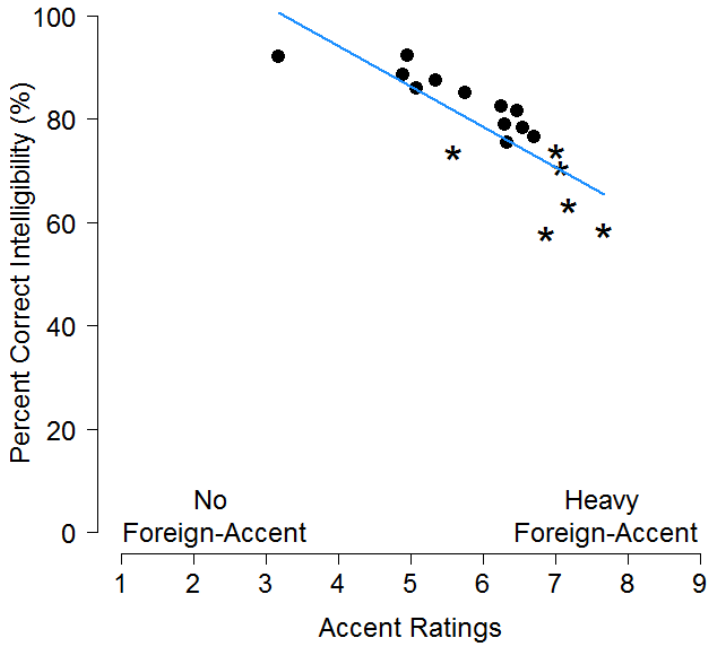

Fig. 2.1. Non-Native (Foreign-Accented) Talkers. The relationship between accent rating and intelligibility for foreign-accented speech presented in quiet, r = -0.82, $p < 0.01$. Stars represent the subset of non-native talkers included in proposed experiments.

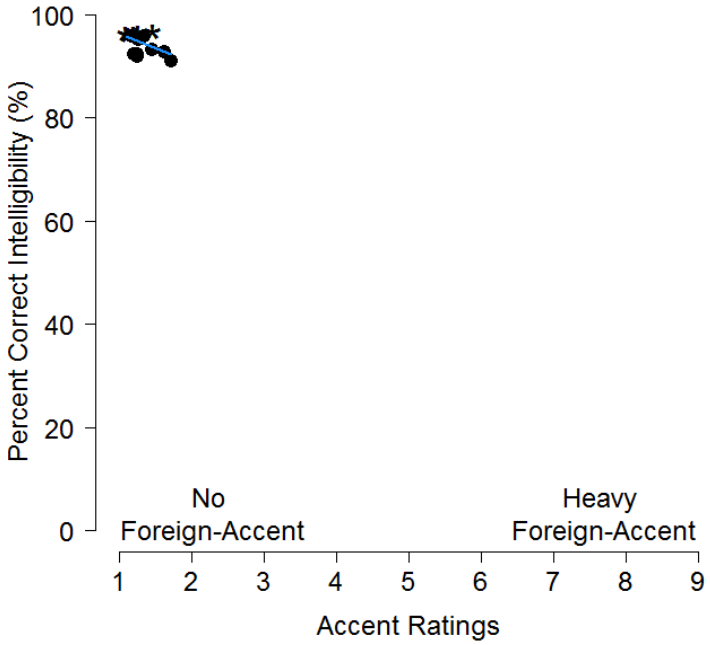

Fig. 2.2. Native Talkers. The relationship between accent rating and intelligibility for native speech presented in quiet, r = -0.48, p = 0.07. Stars represent the subset of native talkers included in proposed experiments.

CHAPTER 3

THE ROLE OF SPECTRAL RESOLUTION IN FOREIGN-ACCENTED SPEECH PERCEPTION[1]

Abstract

Several studies have shown that diminished spectral resolution leads to poorer speech recognition in adverse listening conditions such as competing background noise or in cochlear implants. Although intelligibility is also reduced when the talker has a foreign accent, it is unknown how limited spectral resolution interacts with foreign-accent perception. It is hypothesized that limited spectral resolution will further impair perception of foreign-accented speech. To test this, we assessed the contribution of spectral resolution to the intelligibility of foreign-accented speech by varying the number of spectral channels in a tone vocoder. We also examined listeners' abilities to discriminate between native- and foreign-accented speech in each condition to determine the effect of reduced spectral resolution on accent detection. Results showed that increasing the spectral resolution improves intelligibility for foreign-accented speech while also improving listeners' abilities to detect a foreign accent but not to the level of accuracy for unprocessed speech. Results also revealed a correlation between intelligibility and accent detection. Overall, results suggest that greater spectral resolution is needed for perception of foreign-accented speech compared to native-accented speech.

---

[1] This work was previously published in the Interspeech 2016 Proceedings and is reprinted in this dissertation with permission from the International Speech Communication Association (ISCA): Kapolowicz, M.R., Montazeri, V., & Assmann, P.F. (2016). The role of spectral resolution in foreign-accented speech perception. *Proc. Interspeech 2016*, 3289-3293, doi: 10.21437/Interspeech.2016-1585.

3.1 Introduction

Foreign-accented speech (FAS) is considered to carry an auditory perceptual distortion which requires more time and cognitive effort to understand (Van Engen & Peelle, 2014). Despite the initial difficulty with understanding FAS, listeners are generally able to adapt (Clarke & Garrett; 2004; Bradlow & Bent, 2007; Floccia *et al*., 2009; Trude *et al*., 2013; Baese-Berk *et al*., 2013; Witteman *et al*., 2013; Witteman *et al*., 2014). Here, we examine the relationship between perceived accentedness and intelligibility during listeners' initial exposure to non-native talkers as a function of spectral resolution (SR). It is predicted that decreased SR will result in decreased intelligibility and accent detection for FAS.

Compared to cochlear implant (CI) users, normal-hearing listeners have less difficulty perceiving native speech (NS) in quiet (Faulkner & Pisoni, 2013). CI users also have more difficulty with talker variability, such as talkers with varying linguistic backgrounds (Cleary & Pisoni, 2002; Clopper & Pisoni, 2002; Cleary *et al*., 2005). An attributable difference between CI users and normal-hearing listeners is decreased SR in CI users. CI users have access to reduced SR mainly due to a limited number of physical channels. The importance of SR in speech perception is further implicated by studies showing that intelligibility is reduced in adverse listening conditions, such as in the presence of competing background noise due limited spectral resolution (Loizou *et al*., 2003; Shannon *et al*., 2004).

Evidence from CI users showed lower intelligibility for FAS compared to normal-hearing listeners (Ji *et al*., 2014). CI users have more difficulty detecting foreign accents than normal-hearing listeners, which may limit their ability to make rapid perceptual adjustments required to adapt to the deviation from the expected target speech signal (Tamati & Pisoni, 2015). This

evidence suggests the hypothesis that there should be a systematic relationship between intelligibility and SR, as well as a relationship between accent detection and SR. A correlation between intelligibility and accent detection as a function of spectral resolution is also predicted. As aforementioned, it is expected that FAS will undergo a further reduction in intelligibility and accent detection when SR is reduced.

To parse the effect of reduced SR from other potential confounding factors found in CI users, we tested normal-hearing listeners using speech processed through a tone vocoder, where the number of channels can be varied to limit SR cues available to the listener (Shannon *et al*., 1995). Given that there is a general advantage for higher SR in difficult listening situations, it is expected that perceiving FAS will also require greater SR compared to NS. This hypothesis was tested by examining the effect of SR on speech intelligibility.

By increasing SR, it is expected that the "foreign-accentedness" of the speech would be more obvious to listeners than when SR is reduced. On the other hand, greater SR generally improves speech perception. However, there is a potential conflict if increasing the SR also increases the distortion stemming from the foreign accent. This conflict is addressed in this study by investigating the relationship between perceived accentedness and intelligibility. It may also be argued that decreasing SR is, itself, a source of perceptual distortion. Although this may be the case, for normal-hearing listeners, only minimal SR cues are needed to reach near perfect intelligibility in quiet (Shannon *et al*., 1995; Dorman *et al*., 1997). To control for this, outcomes from the perception of vocoded native-accented speech are also investigated in this study.

3.2 Method and procedure

3.2.1 Speech materials

Audio recordings of low-context Harvard sentences (IEEE, 1969) were obtained from 15 native (5 males, 10 females; age range: 18-38 years) and 18 non-native (9 males, 9 females; age range: 18-47 years) talkers of American English. All talkers were students at The University of Texas at Dallas. Non-native talkers, with a range of 2 weeks to 22 years of residency in Texas, were born and raised in Taiwan and reported using Mandarin as their native language. Non-native talkers were paid a nominal fee for producing the recordings. Native talkers have only ever resided in Texas and were monolingual. Native talkers were awarded research credits for participation. Both groups were given a brief hearing screening and reported no hearing impairments.

Talkers were instructed to repeat each sentence after listening to the sentence spoken by a male native-American English talker and viewing a transcript of the sentence on a computer monitor. Recordings were made in a sound-attenuated booth using a Shure SM-94 microphone, Symetrix SX202 dual-microphone pre-amplifier and Tucker-Davis Technologies data acquisition hardware (MA1, RP2.1). Digital waveforms were stored on a computer disk at a rate of 48 kHz and 16-bit resolution. Sentences were RMS-equalized across all talkers. All procedures for talkers were reviewed and approved by The University of Texas at Dallas Institutional Review Board. Based on our unpublished work classifying talkers from this database in terms of accentedness and intelligibility, stimuli included in these experiments were obtained from a subset of these talkers: six talkers for the NS condition (4 F; no foreign accent; group mean

intelligibility score: 66% in quiet) and six talkers for the FAS condition (4 F; medium-to-heavily foreign-accented; group mean intelligibility score: 96% in quiet).

3.2.2 Speech processing

A sine-wave processor was implemented replicating the specifications of Dorman *et al*. (1997). Speech stimuli were first passed through a pre-emphasis filter (low-pass below 1200 Hz, -6 dB per octave). The filtered signals were then band-passed using 6th-order Butterworth filters into *N* logarithmically-spaced frequency bands (where *N* was either 3, 4, 5, or 9, based on the expectation reported in Dorman *et al*. (1997), showing that performance would reach a plateau within this range). The envelopes of the band-passed signals were then extracted with full-wave rectification followed by low-pass filtering using a 2nd-order Butterworth filter with a cutoff frequency set to 160 Hz. *N* sinusoids were then generated with amplitudes equal to the RMS energy of the envelopes (computed every 4 ms) and frequencies equal to the center frequencies of the bandpass filters. The sinusoids generated for each band were then multiplied by the envelopes, filtered using the same bandpass filters, and finally summed across channels. For additional information, see (Dorman *et al*., 1997).

3.2.3 Experimental Procedure

To test the effect of spectral resolution on FAS perception, eleven students at The University of Texas at Dallas (who did not participate in our procedure for talker group assignment) with an age range of 18 to 25 were recruited for the listening experiment. Listeners (monolingual, native-English talkers from Texas) were screened for normal hearing, and were awarded research credits for participation. All procedures for listeners were reviewed and approved by The University of Texas at Dallas Institutional Review Board.

In each trial, the target signal was randomly selected (without replacement) from the previously recorded sentences. Participants were asked to type the words they heard. They were also required to specify whether or not the talkers had a foreign accent. Intelligibility scores were calculated as the ratio of the number of correctly identified keywords to the total number of presented keywords.

The experimental design was a 4 x 2 repeated measure design: 4 SR configurations (3, 4, 5, and 9 channels) x 2 accent conditions (native- and foreign-accented). The experiment was conducted in a double-walled sound booth. In each trial, participants were presented with stimuli through a Tucker-Davis sound system and Sennheiser HD 598 headphones. The stimuli were presented to the listeners at a comfortable level.

3.3 Results

3.3.1 Intelligibility

Fig. 3.1 shows the results of intelligibility scores as a function of number of vocoder channels and talkers' accentedness. A repeated measures analysis of variance indicated a significant main effect of number of channels ($F(3,30) = 96.89$, $p < 0.01$), as well as a significant main effect of foreign accentedness ($F(1,10) = 186.2$, $p < 0.01$), and also a significant interaction of accentedness by number of channels ($F(3,30) = 3.49$, $p < 0.05$) on speech intelligibility scores.

Post-hoc comparisons using Bonferroni corrections revealed significant differences between native- versus foreign-accented intelligibility scores for 3, 4, 5, and 9 channels (all $p < 0.01$). Additional post-hoc analyses were performed to compare intelligibility scores between channels for NS. Analyses showed that there was a significant difference between intelligibility

scores for 3 and 4 channels, for 3 and 5 channels, for 3 and 9 channels, for 4 and 9 channels, and for 5 and 9 (all $p < 0.01$). Analyses also indicated that the difference between intelligibility scores for 4 and 5 channels was not significant, ($p = 1.00$). Post-hoc analyses were also performed to compare intelligibility scores between channels for FAS talkers. Analyses showed that there was a significant difference between intelligibility scores for 3 and 5 channels, for 3 and 9 channels, for 4 and 9 channels, and for 5 and 9 channels (all $p < 0.01$). Differences between intelligibility scores were not significant for 3 and 4 channels, ($p = 0.18$) nor for 4 and 5 channels, ($p = 1.00$).

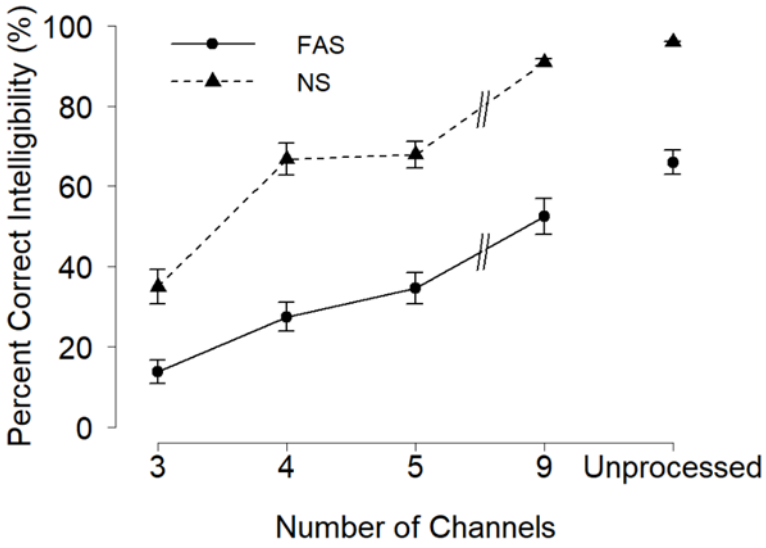

Fig. 3.1. Intelligibility scores across channels for native-accented speech (NS) and foreign-accented speech (FAS). Error bars represent the standard error of the means (*SEM*). Unprocessed scores were collected from the talker database described in Chapter 2 and are provided here for comparative purposes; the talkers are the same, but the listeners differ across studies.

3.3.2 Foreign accent detection

Fig. 3.2 summarizes the perceived accent judgments (where 0 = unaccented and 100 = foreign-accented, averaged across talkers and listeners) as a function of number of vocoder

channels and talkers' accentedness (NS, FAS). A mixed-effects logistic regression model on judgments of perceived accentedness indicated a significant effect of number of channels (χ^2 = 35.32, $p < 0.01$) and a significant effect of talker accentedness ($\chi^2 = 24.27$, $p < 0.01$). For native-accented talkers, a significant improvement was observed for 4 or more channels compared to 3 channels ($p < 0.01$ for each comparison). For foreign-accented talkers, only the 9-channel condition produced significantly better accent detection compared to 3 channels ($p < 0.01$).

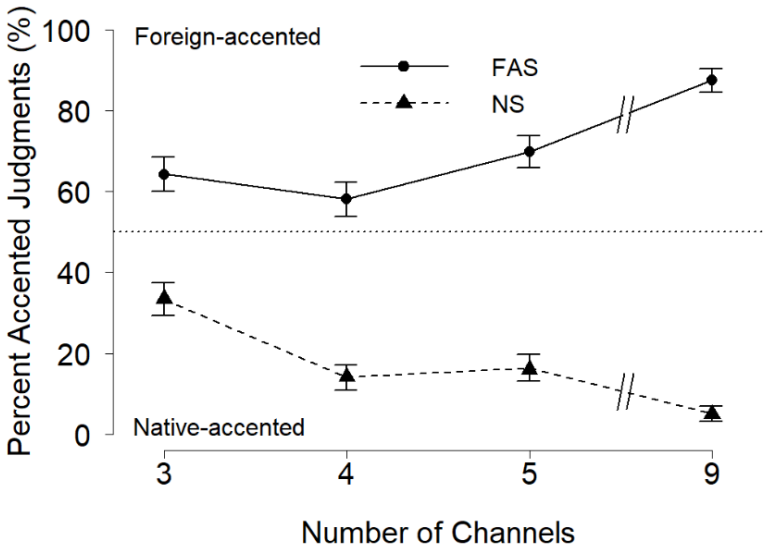

Fig. 3.2. Perceived accentedness across channels for NS and FAS. Error bars represent *SEM*. Scores approaching 0 correspond to participants' judgments as NS; scores close to 100 correspond to participants' judgments as FAS. Note the increase in accuracy of accent judgments for both FAS and NS conditions with increasing number of channels. Dotted horizontal line indicates chance.

3.3.3 Relationship between intelligibility and foreign accent detection in vocoded speech

Fig. 3.3 presents a scatterplot of mean intelligibility and foreign accent detection scores for individual listeners in each channel condition. The plot shows a systematic relationship between accent detection, intelligibility and number of channels: Intelligibility and accent detection are both higher when the number of channels increases. However, the benefit

associated with increasing the number of channels is larger for NS than for FAS. A significant linear relationship between accent detection and intelligibility was found for each talker group (r = 0.48, $p < 0.01$ for FAS and r = -0.67, $p < 0.01$ for NS). It should be noted that the signs of the correlations are reversed for the two talker groups because the abscissa shows accentedness judgments rather than proportion correct.

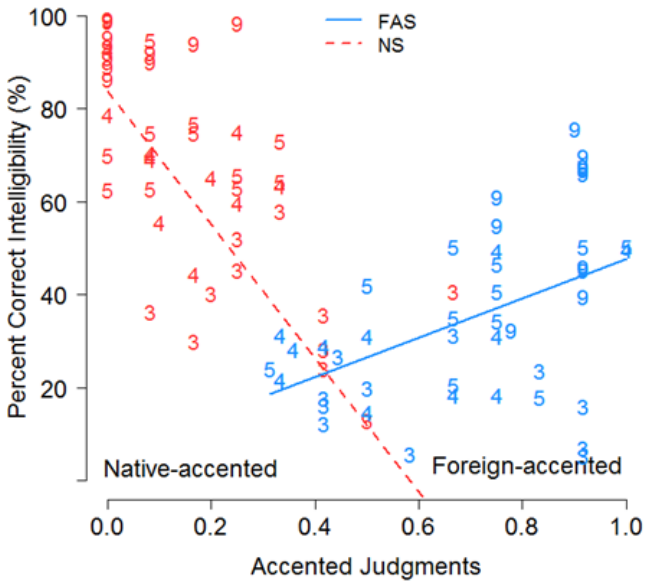

Fig. 3.3. The relationship between accent detection and intelligibility for NS (red dashed line) and FAS (blue solid line). Numbers represent spectral resolution condition (number of channels).

3.3.4 Relationship between intelligibility and foreign accent detection in unprocessed speech

Fig. 3.4 shows the relationship between intelligibility scores and participants' ratings of foreign-accentedness for unprocessed FAS (in contrast to Fig. 3.3 which shows accent detection for vocoded speech). A strong negative relationship was found (r = -0.82, $p < 0.01$). The comparison shows that listeners had more difficulty understanding talkers who were rated as having a heavier foreign accent.

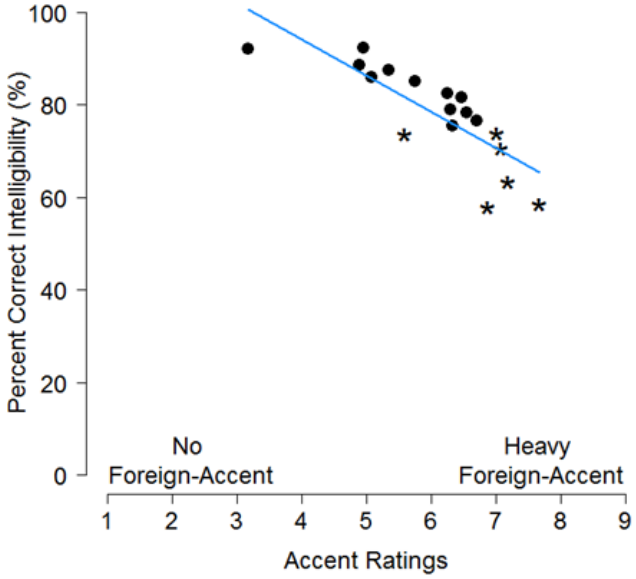

Fig. 3.4. The relationship between accent rating and intelligibility for foreign-accented unprocessed speech. Stars represent subset of FAS used in the vocoder experiment.

3.4 Discussion

FAS introduces a type of auditory perceptual distortion that is intrinsic to the signal, itself, whereas other distortions, such as competing background noise, are external to the source of the signal (Lane, 1963). The aim of the present study was to examine the effects of FAS and reducing the number of channels in a vocoder (two forms of intrinsic distortion) on speech intelligibility and accent detection. Our results show that greater SR is required when listening to FAS compared to NS.

Dorman *et al*. (1997) showed that for NS in quiet, sentence intelligibility approaches ceiling with 4 channels. They found that adding more SR did not further benefit the listener. In the present study, there were no differences in intelligibility between 4 and 5 channels for NS in

quiet, but there was a dramatic improvement approaching ceiling for 9 channels. The present study did not test SR using 6, 7, or 8 channels, (since Dorman *et al*. reported no perceptual benefit gained by increasing the number of channels from 5 to 9 for sentences spoken in quiet) so the precise point at which performance reaches a plateau is uncertain. Given the similarity in signal processing, the discrepancy between the two studies may be attributable to differences in speech materials. The present study used the more complex Harvard sentences, whereas the previous study used the more predictable HINT sentences (presented without competing noise).

Overall, intelligibility for FAS remained lower than for NS, as expected. Interestingly, unlike for NS with limited SR, listeners did not reach the same level of intelligibility for 9 channels as they did when they had access to unprocessed FAS. This leads to the conclusion that decreasing SR produces an additional deficit for understanding FAS. Also the benefit gained from increasing SR is limited for perceiving FAS when compared to NS. The question remains as to whether increasing SR further would benefit FAS perception as suggested by the results reported by Shannon *et al*., 2004. Increasing the low-pass cutoff frequency and the number of channels can directly test this possibility.

In this study, we also examined listeners' abilities to determine whether talkers were native- or foreign-accented. We found that accent detection increases with SR. The database used in this study consists of recorded speech samples from native- and foreign-accented talkers; all foreign-accented talkers were rated as having a foreign accent. This ensures that, under unprocessed conditions (where listeners had full access to spectral resolution), listeners were able to detect a foreign accent with 100% accuracy, unlike with reduced SR, where this task was more difficult.

The results of the present study also show a correlation between accent detection and intelligibility: Listeners are better able to detect differences between NS and FAS when the talkers are more intelligible. Somewhat unexpectedly, this implies that listeners can perceive FAS more accurately despite an increase in distortion from the accent. This result further strengthens the claim that degraded SR presents an added difficulty for the perception of FAS. Although we found a strong correlation between accent ratings and intelligibility in data from unprocessed FAS, previous researchers have reported a weak correlation. Munro and Derwing (1995) presented their FAS talkers with a cartoon story, and asked them to describe the story in their own words. In comparison, the present study elicited the low-context, low redundancy Harvard sentences. Although both studies used accented talkers with similar demographics, the difference in speech elicitation methods might explain this discrepancy.

Studies have shown the benefits of training on listeners' abilities to perceive vocoded speech (Hervais-Adelman *et al*., 2008; Bent *et al*., 2011) as well as lexically challenging words (Burk & Humes, 2007). Studies have also shown that listeners have the ability to adapt to unprocessed FAS with increased exposure (Clarke & Garrett; 2004; Bradlow & Bent, 2007; Floccia *et al*., 2009; Trude *et al*., 2013; Baese-Berk *et al*., 2013; Witteman *et al*., 2013; Witteman *et al*., 2014). As such, our future studies aim to focus on the effects of exposure and short-term training for the perception of FAS with limited SR to see if adaptation can occur without further increasing SR. This would allow us to determine the importance of SR for adaptation to FAS over time. This question is especially important for CI users whose devices provide reduced SR.

3.5 Conclusions

Our results, which tested normal-hearing listeners on their abilities to perceive FAS with decreased SR, corroborate evidence reported by Ji *et al*. (2014) and Tamati & Pisoni (2015) (both of which tested FAS perception in CI users). Taken together, these studies show the importance of spectral information in FAS perception. The data presented here reveals that listeners struggled more with accurately identifying whether or not a talker was foreign-accented when the SR was reduced. Listeners also showed a decrease in intelligibility with lower SR. There was a direct relationship between accent detection (both foreign and native) and intelligibility: More accurate detection of the presence or absence of a foreign accent was associated with higher intelligibility scores. Also, across all channel conditions, listeners were less accurate when detecting FAS compared to NS, and intelligibility scores were lower for FAS than for NS. This evidence strongly suggests that more SR is needed to perceive FAS than NS.

CHAPTER 4

PERCEIVING FOREIGN-ACCENTED SPEECH WITH DECREASED SPECTRAL RESOLUTION IN SINGLE- AND MULTIPLE-TALKER CONDITIONS[2]

Abstract

Under ideal conditions, speech can remain intelligible when spectral resolution is decreased. However, listeners may experience difficulties if the talker has a foreign accent. Effects of reducing spectral resolution on the intelligibility of foreign-accented speech were evaluated in single- and multiple-talker presentations. Intelligibility was similar for single- and multi-talker conditions following initial exposure. Performance improved with extended exposure, but only for the single-talker condition. For unprocessed speech, intelligibility was higher for both initial and extended exposure in single-talker compared to multi-talker conditions. Together, these results indicate that reduced spectral resolution can impair perception and inhibit adaptation to foreign-accented speech.

4.1 Introduction

Foreign-accented speech is classified as non-pathological speech that noticeably differs from native talker pronunciation norms (Munro & Derwing, 1995). Foreign-accented speech affects both segmental and suprasegmental aspects of the signal and can result in increased processing effort, segmental/lexical ambiguity, and mapping failure (Anderson-Hsieh *et al*., 1992). Due to the relative consistency of a talker's productions, however, listeners are able to

[2] This work has been submitted for publication to The Journal of the Acoustical Society: Express Letters and is currently under review.

recalibrate their phonemic and/or prosodic categories within the course of a conversation through a perceptual adaptation (learning) process (Clarke & Garrett, 2004; Bradlow & Bent, 2007; Baese-Berk *et al*., 2013). The precise mechanism(s) underlying this adaptation process is still uncertain and is further complicated by divergent findings reported in the literature. For example, conflicting evidence has been reported concerning the relative benefits of exposure to foreign-accented speech produced by either a single or multiple talkers. Bradlow and Bent (2008) demonstrated that exposure to sentences spoken by multiple Chinese-accented English talkers (talker-independent adaptation) was as effective as exposure to the same Chinese-accented talker over time (talker-dependent adaptation). In contrast, Bent and Holt (2013) showed that performance was higher for listeners presented with speech spoken by a single Japanese-accented talker compared to speech spoken by multiple Japanese-accented talkers. Although the role of talker variability in the perception of foreign-accented speech remains uncertain, its role in the perception of native speech is well established, where several experiments have shown that there is a processing cost associated with listening to speech spoken by different talkers as opposed to listening to speech spoken by the same talker (*e.g*., Mullennix *et al*., 1989).

Talker variability presents an even greater challenge for cochlear implant (CI) users compared to normal-hearing listeners. CI users are unable to utilize acoustic-phonetic information to make judgments about the talker to the same extent as normal-hearing listeners (Kirk *et al*., 2000; Cleary & Pisoni, 2002; Cleary *et al*., 2005). It is probable that the increased difficulties associated with talker variability for CI users are largely due to the poor spectral resolution of the CI device. Speech from non-native talkers presents even more acoustic variability for the listener. There is limited evidence regarding perception of foreign-accented

speech by CI users, but reported findings suggest that CI users struggle more with perception of foreign-accented speech perception (Ji *et al*., 2014; Tamati & Pisoni, 2015). Ji and colleagues measured performance of sentence recognition in noise from CI users and normal-hearing listeners. They found that the deficit in speech recognition thresholds with nonnative talkers relative to native talkers was approximately 3 dB greater for CI users compared to normal-hearing listeners. Tamati and Pisoni asked CI users and normal-hearing listeners to rate the intelligibility of sentences spoken by native and foreign-accented talkers using a 7 point Likert scale. They found that, compared to normal-hearing listeners, CI users perceived smaller differences in intelligibility between native and foreign-accented sentences. Tamati and Pisoni postulated that their results indicate that, compared to normal-hearing listeners, CI listeners are less sensitive to foreign accents. Given the degree of variability in the responses from the CI listeners who participated in their experiment, however, they speculated that the results from the CI users might be due to development and use of basic speech and language processing skills.

Although it is possible that the difficulty observed in CI users regarding talker variability and perception of foreign-accented speech is due to the aforementioned reasons suggested by Tamati and Pisoni, reduced spectral resolution, as well as additional factors such as frequency-to-place mismatch, electrical channel interaction, and history of deafness also constrain speech intelligibility in CI recipients (Winn & Litovsky, 2015). Given these additional factors present in CI users, it difficult to determine the extent to which each factor, alone, is responsible for their struggle when listening to multiple talkers or when listening to foreign-accented speech. Here, we explore the possibility that the foreign accent information conveyed by fine spectral detail, which is not well encoded in CI users, may aid listeners when perceiving foreign-accented

speech. The approach adopted here is to test normal-hearing listeners using a vocoder that controls spectral resolution explicitly by varying the number of filter channels (*e.g.*, Shannon *et al.*, 1995).

Our previous research has shown that decreased spectral resolution can result in inaccurate judgments of accentedness and a drop in intelligibility in perception of foreign-accented speech (Kapolowicz *et al.*, 2016). The present research aims to extend these findings by investigating the role of spectral resolution in perceiving either a single or multiple interleaved native- and foreign-accented talkers using vocoded speech with limited spectral resolution. The present study also investigates whether adaptation to foreign-accented speech can occur when spectral resolution is limited to 9 channels. Nine channels has been shown to allow listeners to achieve near-perfect sentence recognition for native speech and high identification rates for vowels in /bVt/ syllables (Dorman *et al.*, 1997). It is hypothesized that limiting spectral resolution in foreign-accented speech can reduce intelligibility and limit adaptation, with further deterioration when listeners are exposed to multiple interleaved foreign-accented talkers.

4.2 Method and procedure

4.2.1 Listeners

128 monolingual, native talkers of English (age range: 18-35 years, mean: 21.5 years) with normal hearing were recruited for participation. Participants had only ever resided in Texas and reported variable exposure to foreign-accented speech. All participants reported normal hearing and passed a hearing screening at 20 dB hearing level at octave frequencies from 250 to 8000 Hz in both ears. Participants were students at The University of Texas at Dallas who were

compensated with course credit. All procedures were reviewed and approved by The University of Texas at Dallas Institutional Review Board.

4.2.2 Stimuli

Harvard sentences (IEEE, 1969) were recorded in a sound booth from native and non-native talkers of American English. Sentences for the native-accented conditions were obtained from a subset of 15 monolingual, native talkers of American English who had only resided in Texas (4 females, 2 males; age: 18-38 years, mean age: 23 years; 96% mean intelligibility score in quiet). Sentences for the foreign-accented conditions were obtained from a subset of 18 Chinese-accented (L1 = Mandarin) talkers of American English (3 females, 2 males; age: 18-47 years, mean age: 30.6 years; 66% group mean intelligibility score in quiet) who had only resided in Taiwan and Texas (range of 1 month – 22 years of residency in Texas). Talkers repeated sentences after listening to the sentences spoken by a male native talker of American English and viewing a transcript on a screen. Digital waveforms were stored at a rate of 48 kHz and 16-bit resolution and RMS-equalized across talkers and sentences. Talkers were students of The University of Texas at Dallas and were paid $20 for participation.

Spectral resolution was manipulated using a sine-wave processor (Dorman *et al*., 1997). Speech stimuli were passed through a pre-emphasis filter (low-pass below 1200 Hz, -6 dB per octave) and band-passed using 6th-order Butterworth filters into 9 logarithmically-spaced frequency bands. (9 channels were chosen based on results reported by Kapolowicz *et al*., 2016.) The envelopes of the band-passed signals were extracted with full-wave rectification followed by low-pass filtering using a 2nd-order Butterworth filter with a cutoff frequency of 160 Hz. Nine sinusoids were generated with amplitudes equal to the RMS energy of the envelopes (computed

every 4 ms) and frequencies equal to the center frequencies of the bandpass filters. The sinusoids generated for each band were multiplied by the envelopes, filtered using the same bandpass filters, and summed across channels.

4.2.3 Design and procedure:

A 4 x 4 mixed design was used: 4 group assignment conditions [native, control, experimental, unprocessed] (between) by 4 blocks (within). Stimuli in all conditions excluding the unprocessed conditions were processed with a 9-channel vocoder. For the native condition, listeners heard the same native-accented talker for the duration of the experimental phase (40 sentences). For the control condition, listeners heard 30 sentences spoken by the same native-accented talker followed by 10 sentences spoken by the same foreign-accented talker; this was to control for listeners potentially adapting to vocoded speech over time rather than specifically to vocoded *foreign-accented* speech. For the experimental condition, listeners heard the same foreign-accented talker for 40 sentences. For the unprocessed condition (full spectral resolution when listening to foreign-accented speech), listeners heard the same foreign-accented talker. This allowed for a comparison of spectrally reduced versus unprocessed speech when a foreign accent is present. The multiple-talker conditions followed the same descriptions for the single-talker conditions, except rather than hearing a single talker, listeners heard five interleaved talkers. Although talker presentation in the multiple-talker conditions was random, listeners always heard two sentences from each talker in every 10-sentence block.

Experiments were conducted in a double-walled sound-attenuating booth. Stimuli were presented at a comfortable level through headphones using a Tucker-Davis sound system. Talkers presented in native- or foreign-accented single-talker conditions were randomly chosen

from the subset of 5 talkers used in the multiple-talker conditions except when training on the task (block 1), where all listeners heard the same female native-accented talker for 10 sentences. After training, this talker was not heard again for the duration of the procedure (blocks 2-5, 10 sentences per block). Listeners' responses for Block 1 were not included in statistical analyses. The procedure was blocked using 10-sentence increments to consider if adaptation to stimuli occurred earlier than in the final exposure block. Talker presentation in the multiple-talker conditions was also randomized. In each trial, the sentence was randomly selected (without replacement) from the previously recorded sentences from either the same native- or foreign-accented talker or five interleaved native- or foreign-accented talkers, with condition assignment being randomly selected. Participants responded by typing the words they heard. Intelligibility scores were calculated as the ratio of correctly identified keywords to the total number of presented keywords.

4.3 Results

4.3.1 Single-talker conditions

Fig. 4.1 shows the results of intelligibility scores as a function of condition (native, control, unprocessed, and experimental) across blocks. Each block consists of exposure to 10 sentences. A mixed design analysis of variance indicated a significant main effect of condition [$F(3,60) = 41.36$, $p < 0.001$], as well as a significant main effect of block [$F(3,180) = 11.20$, $p < 0.001$] and also a significant interaction of block by condition [$F(9,180) = 22.78$, $p < 0.001$] on speech intelligibility scores. To test if adaptation (defined as an improvement in intelligibility from block 2 to block 5) occurred in either the experimental condition or the unprocessed

condition, post-hoc comparisons using Bonferroni corrections revealed a significant difference between intelligibility scores of blocks 2 and 5 for the experimental condition [$t(30) = 2.62$, $p < 0.05$] and a difference which was approaching significance for the unprocessed condition [$t(30) = 1.76$, $p = 0.09$, ns]. To test whether adaptation occurred within the first 10 sentences that listeners were exposed to single-talker foreign-accented speech in the unprocessed condition, a paired-sample *t*-test revealed no significant difference between the scores for the first and last trial of block 2 [$t(15) = 0.61$, $p = 0.55$, ns]. These results indicate that, compared to listening to a single native-accented talker, intelligibility scores are lower when listeners are exposed to a single foreign-accented talker, with a further detriment when spectral resolution is limited. Results also reveal that intelligibility scores improved in the experimental condition from block 2 to block 5, but not for the unprocessed condition (*i.e.*, adaptation to foreign-accented speech only occurred when listeners were exposed over time to the same foreign-accented talker with decreased spectral resolution).

4.3.2 Multiple-talker conditions

Fig. 4.2 shows the results of intelligibility scores as a function of condition across blocks. Each block consists of exposure to 10 sentences. A mixed design analysis of variance indicated a significant main effect of condition [$F(3,60) = 131.13$, $p < 0.001$], as well as a significant main effect of block [$F(3,180) = 12.63$, $p < 0.001$] and also a significant interaction of block by condition [$F(9,180) = 3.80$, $p < 0.001$] on speech intelligibility scores. To test if adaptation occurred in either the experimental condition or the unprocessed condition, post-hoc comparisons using Bonferroni corrections revealed no significant differences between intelligibility scores of blocks 2 and 5 for the experimental condition [$t(30) = 0.393$, $p = 0.70$, ns]

nor for the unprocessed condition [$t(30) = 0.77$, $p = 0.45$, ns]. Results show that, compared to listening to multiple native-accented talkers, intelligibility scores are lower when listeners are exposed to multiple foreign-accented talkers, with a further detriment when spectral resolution is limited. Results also indicate that intelligibility scores for foreign-accented speech do not improve with increased exposure (from block 2 to block 5) in either condition.

4.3.3 Single-versus multiple-talker conditions

When comparing the results of listeners' exposure to either a single talker or 5 interleaved talkers in blocks 2 and 5 of the experimental conditions (Fig. 4.1 and Fig. 4.2), a *t*-test for independent groups did not indicate a significant intelligibility difference between multiple- and single-talker conditions in block 2 [$t(30) = 0.17$, $p = 0.87$, ns]. A *t*-test for independent groups indicated a significant difference in intelligibility between multiple- and single-talker conditions in block 5 [$t(30) = 2.61$, $p < 0.05$]. Results show that when listeners initially hear vocoded foreign-accented speech limited to 9 channels of spectral resolution, there is no difference in intelligibility scores when listeners are exposed to either a single or multiple foreign-accented talkers. With increased exposure, intelligibility scores become higher for the single-talker condition compared to the multiple-talker condition.

When comparing the results of listeners' exposure to either a single talker or five interleaved talkers in blocks 2 and 5 of the unprocessed conditions (Fig. 4.1 and Fig. 4.2), a *t*-test for independent groups revealed a significant intelligibility difference between multiple- and single-talker conditions in block 2 [$t(30) = 2.91$, $p < 0.01$]. A *t*-test for independent groups also indicated a significant difference in intelligibility between multiple- and single-talker conditions in block 5 [$t(30) = 4.81$, $p < 0.001$]. Results show that when listeners initially hear foreign-

accented speech with full access to spectral resolution, intelligibility scores are higher when listeners are exposed to a single foreign-accented talker compared to multiple foreign-accented talkers. With increased exposure, intelligibility scores remain higher for the single-talker condition compared to when listeners are exposed to multiple foreign-accented talkers.

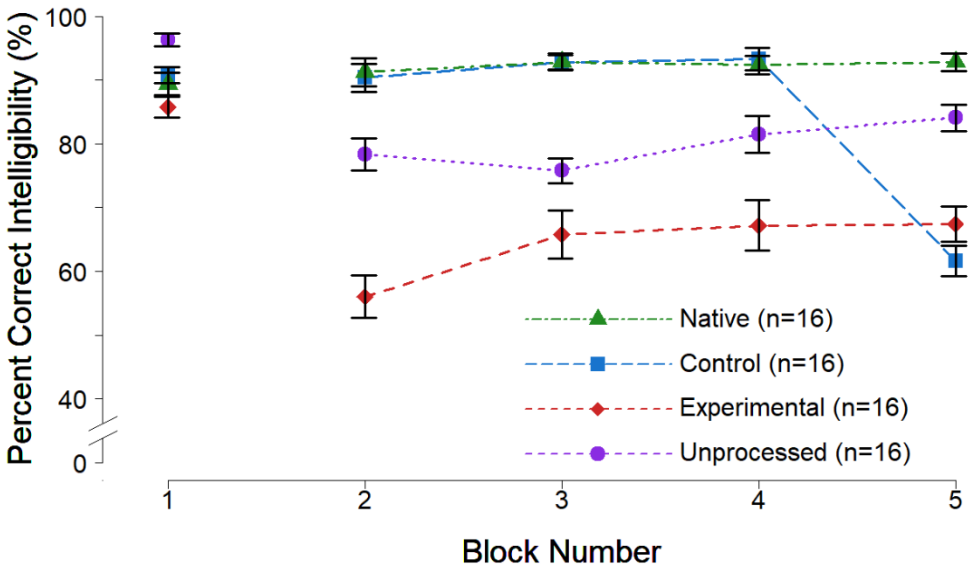

Fig. 4.1. Mean intelligibility scores expressed as percent correct across blocks for single-talker conditions. Standard error of the means (*SEM*) bars are also shown. Block 1 (practice session) was not included in statistical analyses.

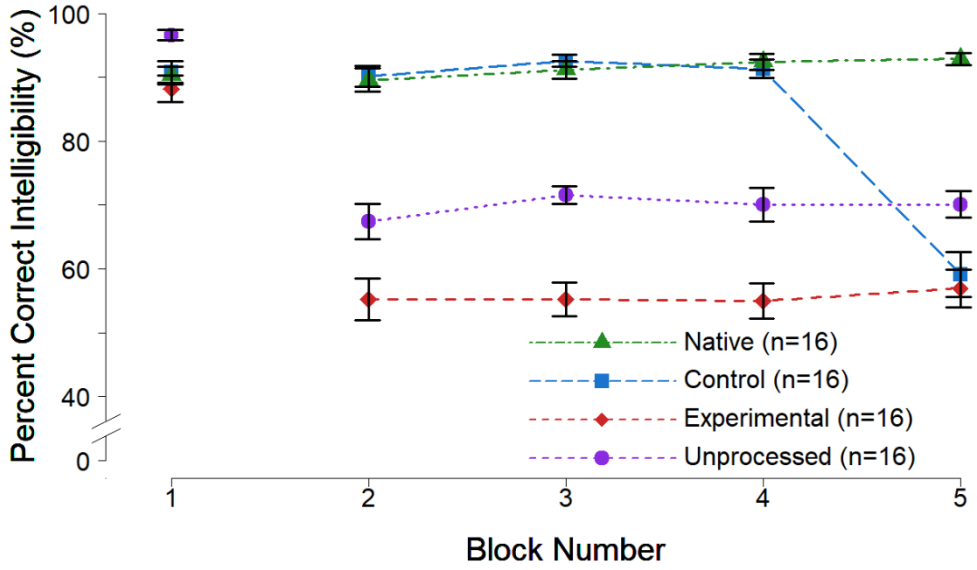

Fig. 4.2. Mean intelligibility scores expressed as percent correct across blocks for the multiple-talker conditions. *SEM* bars are also shown. Block 1 (practice session) was not included in statistical analyses.

4.4 Discussion

Talker variability presents a perceptual problem that listeners must resolve in order to interpret meaningful speech. Reduced spectral resolution may further hinder this process. In native speech, the effects of reduced spectral resolution can be mitigated with increased exposure (Chang & Fu, 2006). Since foreign-accented speech introduces additional sources of variability, we hypothesized that there would be an added detriment when perceiving foreign-accented speech with reduced spectral resolution, particularly when listeners are exposed to multiple interleaved talkers.

Kapolowicz *et al*. (2016) showed that foreign-accented speech perception is affected by a reduction in spectral resolution: When limited to 9 channels, intelligibility scores and accent detection were lower than when listeners had full access to spectral resolution, whereas scores

reached near ceiling performance for native-accented speech. The current research extended these previous findings by demonstrating that, even with increased exposure (comparing blocks 2 and 5, Fig. 4.1 and Fig. 4.2), intelligibility remained lower for foreign-accented speech with reduced spectral resolution compared to conditions with full spectral information. Additionally, intelligibility degradation was more substantial when listeners were exposed to multiple foreign-accented talkers throughout the time-course of the experiment. Results also revealed that when listeners were exposed to stimuli with full spectral resolution, intelligibility for foreign-accented speech was higher initially and after increased exposure for the single-talker condition compared to the multiple-talker condition. For native-accented speech, no further detriment was observed in the multiple-talker condition compared to the single-talker condition with reduced spectral resolution; intelligibility remained high in both conditions.

Previous research has reported rapid adaptation of foreign-accented speech when listeners had full access to spectral resolution. Clarke and Garrett (2004) used low probability Revised Speech Perception in Noise (SPIN-R) sentences and relied on reaction time rather than intelligibility to measure adaptation. They found an initial increase in reaction time for foreign-accented speech that decreased within 1 minute of exposure. In addition to using a different response measure, Clarke and Garrett only exposed listeners to a single foreign-accented talker. The present study also reports adaptation when listeners were exposed to a single foreign-accented talker with limited access to spectral resolution, but not with full spectral resolution. Using an intelligibility measure, Bradlow and Bent (2008) and Baese-Berk, Bradlow and Wright (2013) demonstrated adaptation when listeners were exposed to the same foreign-accented talker as well as to a novel foreign-accented talker if listeners were trained on multiple foreign-

accented talkers. They used high-context Bamford-Kowal-Bench (BKB) sentences, compared to the low-context Harvard sentences used in the present study, and their training was conducted over two separate days. The variability in experimental design, stimuli and training may account for the differences reported in adaptation performance for these studies when compared to our results.

CHAPTER 5

PERCEPTION OF SPECTRALLY-SHIFTED FOREIGN-ACCENTED SPEECH

Abstract

There is a natural covariation of the fundamental frequency and the spectral envelope across men, women and children in natural speech. The present study examined how shifting the spectral envelope and the fundamental frequency using a high-quality vocoder affected listeners' judgments of talker identity and perceived naturalness when listening to native English compared to Mandarin-accented English. The present study also investigated how increased exposure to spectrally-shifted speech from a single Mandarin-accented talker affects intelligibility scores. The spectral envelope and the fundamental frequency from previously recorded sentences were shifted up or down in the same direction by 8% and 30%, respectively. Results from the talker identity experiment indicated that listeners perceive spectrally-shifted speech from the same talker as a different talker regardless of accent condition. Results from the perceived naturalness study demonstrated that listeners were less tolerant to downward shifts from male talkers regardless of accent. Also, overall, listeners rated foreign-accented speech as sounding less natural even when the speech was unshifted. Finally, when stimuli from the same foreign-accented talker was shifted to simulate five different talkers, adaptation did not occur, and intelligibility patterns were similar to when listeners actually heard five different foreign-accented talkers. Results suggest that listeners are unable to adapt to foreign-accented speech in a talker-dependent manner when only the fundamental frequency and spectral envelope in speech stimuli from a single talker are manipulated.

5.1 Introduction

Intelligibility declines when listening to frequency-shifted speech compared to unprocessed speech. Large frequency shifts can impair intelligibility, but speech perception is relatively well preserved when small multiplicative shifts in the spectral envelope are introduced using a vocoder (*e.g.*, Assmann & Nearey 2008). This suggests that listeners with normal hearing are able to utilize information from the spectral domain to enable rapid adaptation to talker variability. On one hand, spectral shifts preserve intelligibility, even when speech goes outside the natural range (Smith *et al*., 2005); however, there can be a cost that is introduced when speech is spectrally shifted. For example, men's vowels are more susceptible to downward shifts compared to vowels spoken by women and children, while there is a greater decline with upward shifts in children's vowels. These results imply that declines in perceptual performance are related to the absolute ranges of the formant frequencies or other features of the spectral envelope across age/sex classes (Assmann & Nearey, 2008). Relatable findings have been obtained for the perception of vowels (Daniloff *et al*., 1968; Assmann & Nearey, 2008) and connected speech (Assmann & Nearey, 2007).

The preserved intelligibility of frequency-shifted speech is consistent with a rapid adaptation to talker characteristics, while the breakdown in intelligibility with larger shifts suggests there are hard limits on this mechanism. However, some studies have shown that intelligibility improves with extended exposure (Rosen *et al.*, 1999; Nogaki *et al.*, 2007). Rosen and colleagues used a 4-channel vocoder and spectrally-shifted speech upward and reported that listener performance for identification of intervocalic consonants, medial vowels in monosyllables and words in sentences improved after increased exposure to spectrally-shifted

speech even when initial performance was as low as 1%. Nogaki and colleagues used an 8-channel vocoder and also spectrally-shifted speech upward, and they reported similar results to Rosen *et al*., namely significant improvements after training for vowel, consonant and sentence recognition. Notably, both studies not only performed upward spectral shifts with their speech stimuli, but they also spectrally reduced the speech stimuli. The combination of the two spectral manipulations were performed to simulate both spectrally-shifted and compressed speech, two aspects of spectral distortion often associated with cochlear implant devices. One interpretation of these findings is that the limits on adaptation to talker variability are not fixed, but can be extended somewhat as a function of experience. It remains to be seen, however, whether such adaptation occurs with foreign-accented speech.

Our previous research found that spectrally reduced foreign-accented speech results in impaired intelligibility for listeners in single- and multiple-talker conditions (Kapolowicz *et al*., submitted). Listeners were able to adapt with increased exposure when presented with spectrally reduced speech spoken by a single foreign-accented talker, but not to the level of performance for unprocessed speech, where listeners had full access to spectral resolution. The detriment was much greater when listeners were presented with spectrally reduced foreign-accented speech spoken by multiple interleaved talkers, and no improvement occurred with increased exposure. In conditions where foreign-accented speech was unprocessed, we found that intelligibility scores were also greater when listening to speech spoken by the same foreign-accented talker over time as compared to five different talkers. Perceiving speech spoken by different talkers in native speech has been shown to depend on spectral cues (Summerfield, 1981; Nearey, 1989; Mullennix *et al*., 1989; Johnson 1991, Wong & Diehl, 2003) although adaptation in native

speech occurs rapidly, perhaps because native talkers have more canonical and/or familiar speech patterns than foreign-accented talkers (Wade *et al*., 2007; Baese-Berk & Morrill, 2015).

These results provide evidence for the importance of spectral cues when perceiving foreign-accented speech, especially in multi-talker conditions. It is, however, unclear how spectral scaling interacts with intelligibility of foreign-accented speech for native listeners. The research presented here investigates the potential interaction of spectrally-shifted foreign-accented speech by testing the ability of listeners to adapt to spectrally-shifted foreign-accented speech over time in a simulated multiple-talker condition. Given our previous results showing that spectral cues are important when perceiving foreign-accented speech, it was hypothesized that *spectrally-shifted* speech from a single foreign-accented talker to simulate multiple talkers would cause a reduction in intelligibility scores compared to when listeners heard *unprocessed speech* from a single foreign-accented talker. This finding would provide further evidence that talker-specific spectral cues are important when perceiving foreign-accented speech, since only the spectral information was manipulated (and temporal cues remain unchanged). This result would, therefore, also provide additional evidence that talker normalization is a perceptual process that listeners undergo when perceiving foreign-accented speech.

An additional control experiment was performed to investigate the role that spectrally-shifted foreign-accented speech has on the ability of listeners to discriminate talker identity. Another control experiment was performed to test whether spectral scaling would affect perceived “naturalness” of foreign-accented speech. Perceived “naturalness” is a subjective measure entailing that a speech stimulus sounds normal or natural to the listener (Parrish, 1951). Assmann *et al*. (2006) showed that listeners judge frequency-shifted sentences as more natural

when F0 and mean formant frequencies were shifted following the covariation in natural voices: Listeners assigned ratings of “masculine” to voices with low F0 and low formant frequencies and “feminine” to voices with high F0 and high formant frequencies. Assmann *et al*. also found, however, that frequency-shifted sentences from male talkers received higher ratings of “masculine” than sentences derived from female talkers, even when shifted to the female range. Similarly, sentences from female talkers received higher ratings of “feminine” than sentences derived from male talkers despite scale factor assignment being appropriate for the opposite gender. This result indicates that factors other than F0 and average formant frequencies contribute to perceived gender. Cleary *et al*. (2005) studied how acoustically different (regarding the average F0 and formant frequencies) utterances needed to be for a child to classify the utterances as being spoken by two different talkers. They found that the average spectral characteristics (F0 and formant frequencies) needed to differ by at least 11% for children to perceive the voices as coming from two different talkers. In the present experiments, it was predicted that spectrally-shifting F0 and the spectral envelope in the same direction (up or down) from the same talker beyond a certain limit would convince listeners that they were perceiving a different talker. It was also predicted that spectrally-shifted speech would be perceived as being natural except in the case of extreme shifts that went beyond the natural range of human speech. It was also expected that foreign-accented speech would be perceived as being less natural when compared to native-accented speech even when the speech was unshifted.

5.2 Method and procedure

5.2.1 Speech processing

Frequency-scaled sentences that follow the covariation in natural voices (Assmann & Nearey, 2008) were constructed by processing Harvard sentences previously obtained from five foreign-accented talkers. (See Chapter 2 for a detailed description of speech materials and talker group assignment). Synthesized versions of each sentence were obtained for each talker using the STRAIGHT vocoder (Kawahara, 1997) to shift F0 and the spectral envelope (four versions of each sentence with frequency scaling in the same direction, up or down). The spectral envelope of each sentence was shifted up or down by a scale factor of 1 + 0.08 entailing an 8% shift for each scaled stimulus. F0 for each sentence was shifted up or down by a scale factor of 1 + 0.296 entailing about a 30% shift for each scaled stimulus. Speech from one of five native- or foreign-accented talkers was randomly selected for each listener, and the original plus four frequency-scaled versions of the sentences were presented to listeners to simulate five different talkers. Two additional scaled versions of each sentence were spectrally-shifted to simulate extremely unnatural upwards and downwards shifts beyond the normal human speech range for the naturalness control experiment (allowing for a comparison of the original plus six frequency-scaled versions of the sentences).

5.2.2 Listeners

Two hundred fifty-six monolingual (160 for control experiments; 64 for intelligibility experiment), native talkers of English (age range: 18-52 years, mean: 21.6 years) with normal hearing were recruited for participation. Participants had only ever resided in Texas and reported variable exposure to foreign-accented speech. All participants reported normal hearing and

passed a hearing screening at 20 dB hearing level at octave frequencies from 250 to 8000 Hz in both ears. Participants in the control experiments were students at The University of Texas at Dallas and were compensated with course credit. Participants in the intelligibility experiment were also university students and were paid $15.00 for their participation. Experiments were conducted in a sound-attenuating booth. Stimuli were presented at a comfortable level through Sennheiser HD-598 headphones using Tucker-Davis System 3 and RP2.1 hardware. Stimuli and conditions were randomized and presented using custom Matlab scripts. All procedures were reviewed and approved by The University of Texas at Dallas Institutional Review Board.

5.2.3 Design and procedure: talker identity and naturalness ratings

The talker identity experiment was a 2 x 2 x 5 x 5 mixed design: accentedness (between) by same/different talker (within) by baseline shift factor (between) by test baselines against shift factors (within). There were 10 listener groups: 5 groups (5 baseline shift factors) by 2 accents (foreign-accented and native-accented). Each listener heard 5 shifts by 2 sentences per talker for each shift by 5 talkers, which equals 50 trials (2 sentences per trial). The total number of sentences heard for each listener for the talker identity experiment was 100 sentences. Sentences were selected randomly without replacement. Listeners were asked after hearing the two sentences in each trial if the sentences were spoken by the same talker or two different talkers; the response choice was binary.

The naturalness experiment was a 2 x 5 x 7 repeated measures design: accent by talker by shift factor. Listeners heard 70 sentences total for this experiment, and they used a 6-point Likert scale (ranging from extremely natural to extremely unnatural) to rate the level of naturalness for each sentence. Listeners were informed that they would hear computer-processed speech that

could come from males and females of any age range. They were also informed that the speech could range from native-accented speech to heavily foreign-accented speech. Listeners were instructed *not* to rate the level of naturalness for each sentence based on how intelligible the words were nor on how heavily a foreign accent was perceived, but rather to focus on distinguishing between voices which sound like they could come from an actual human being (which should be rated as more natural) and voices that sound more fictitious, such as a cartoon character or a monster (which should be rated as less natural).

The naturalness control experiment was run immediately after the talker identity control experiment was completed. The same listeners participated in both control experiments. Although no sentences were repeated for the talker identity experiment, a subset of sentences were repeated to listeners for the naturalness experiment, but should not present any confounds since the naturalness experiment follows the talker identity experiment. Though it is beyond the scope of this dissertation research, a future experiment could examine whether sentence repetition is a factor when listeners are assigning naturalness ratings (*i.e.*, would a previously heard sentence bias a listener to consider the repeated sentence as sounding more natural due to familiarization?).

5.2.4 Design and procedure: intelligibility of spectrally-shifted foreign-accented speech

The experiment was a 4 x 4 mixed design: 4 conditions (between) by 4 blocks (within). For the single-talker condition, listeners heard a single foreign-accented talker across 40 sentences, and this talker was randomly selected from the five foreign-accented talkers presented in the multiple-talker condition. For the multiple-talker condition, listeners heard five different foreign-accented talkers across 40 sentences. For the simulated multiple-talker condition,

listeners heard unprocessed speech as well as spectrally-shifted speech from same talker across 40 sentences to simulate five different talkers. For the control condition, listeners heard sentences that were processed through the vocoder but remained unshifted from the same foreign-accented talker to account for artifacts introduced from the vocoder. The talker in the control condition was also randomly selected from the five foreign-accented talkers presented in the multiple-talker condition. To gain familiarity with the procedure, listeners in all conditions heard the same native-accented talker for a 10-sentence practice block (block 1). Listeners' responses for Block 1 were not included in statistical analyses. The remaining 40 sentences were divided into four 10-sentence blocks (blocks 2-5). The procedure was blocked using 10-sentence increments to consider if adaptation to stimuli occurred earlier than in the final exposure block. In each trial, the target sentence was randomly selected (without replacement) from previously recorded/processed sentences. Intelligibility scores were calculated as the ratio of correctly identified keywords to total number of presented keywords.

5.3 Results

5.3.1 Talker identity

A mixed effects logistic regression model on judgments of talker identity indicated a significant effect of shift difference [$\chi^2 = 543.39$, $p < 0.001$], talker difference [$\chi^2 = 1934.49$, $p < 0.001$], and a significant interaction between shift difference and talker difference [$\chi^2 = 1013.46$, $p < 0.001$], a significant interaction between talker difference and foreign accent [$\chi^2 = 14.23$, $p < 0.001$] and a 3-way interaction between shift difference, talker difference and foreign accent [$\chi^2 = 24.48$, $p < 0.001$] on listeners' abilities to correctly ascertain whether the stimuli in each trial

were spoken by the same talker or two different talkers. Notably, the main effect of foreign accent and the interaction between shift difference and foreign accent were not significant.

Fig. 5.1 displays listeners' judgments for the conditions in which the talkers in each trial were the same. Fig. 5.2 shows listeners' judgments for the conditions in which the talkers in each trial were different. Overall, the results indicate that shifting the spectral envelope and F0 information of speech stimuli resulted in listeners' judgments of the same talkers as being convincingly *different* talkers. Additionally, the results indicate that listeners' correct answers did not depend on the talkers' accentedness; however, when the first talker was different from the second talker in a given trial, correct judgments depended on the talkers' accentedness.

5.3.2 Naturalness ratings

A nested factor repeated measures analysis of variance revealed a significant main effect of accent [$F(1,9922) = 355.60$, $p < 0.001$], a significant main effect of spectral shifting [$F(6,9022) = 3094.72$, $p < 0.001$], a significant interaction between talker and accent [$F(8,9022) = 27.01$, $p < 0.001$], and a significant interaction between accent and spectral shifting [$F(6,9022) = 14.05$, $p < 0.001$]. Figs. 5.3 and 5.4 show the mean naturalness ratings across shift factors for each native- and foreign-accented talker, respectively. Fig. 5.5 shows the mean ratings for each accent condition. The figures show that, for extreme downward spectral shifting (-3), male voices were perceived as extremely unnatural regardless of the accent of the talker. This was expected, since this extreme shift (which is outside of the normal human speech range) was meant to serve as an anchor point for "extremely unnatural" voices. Extreme upward shifts (3) were still perceived as less natural for male and female speech, though not to the extent that extreme downward shifts had on male voices. Only male voices were perceived as somewhat

unnatural when spectrally-shifted down two steps, and this pattern was greater for the male foreign-accented talkers. These results replicate previous findings by Assmann and Nearey (2008). They found that downward shifts had a greater impact on men's voices compared to women's and children's voices. Specifically, they reported a greater impairment in vowel identification when listeners' heard male voices that were spectrally-shifted down compared to when listening to women's or children's voices shifted down. Upward scaling by a factor of two for male and female talkers did not severely impact naturalness ratings in either accent condition, an effect also reported by Assmann and Nearey (2008), who found that upward scaling of children's voices had a greater negative impact on vowel identification for listeners compared upward scaling of adult male and female voices. Fig. 5.5 shows that listeners rated foreign-accented talkers as less natural than native-accented talkers in all conditions including when speech was unshifted.

5.3.3 Intelligibility

A 4 x 4 mixed design analysis of variance revealed a significant main effect of processing condition [$F(3,60) = 6.08$, $p < 0.01$] and a significant main effect of block [$F(3,180) = 4.64$, $p < 0.01$]. The interaction between block and condition was not significant [$F(9,180) = 1.25$, $p = 0.27$]. The following pairwise comparisons using Bonferroni corrections were significantly different across blocks 2 through 5: unprocessed single-talker and spectrally-shifted single-talker, $p < 0.001$, unprocessed single-talker and unprocessed multiple-talker, $p < 0.001$, vocoder control and unprocessed single-talker, $p < 0.01$. The following pairwise comparisons were significantly different in block 5: unprocessed single-talker versus spectrally-shifted single-talker, $p < 0.05$, unprocessed single-talker versus unprocessed multiple-talker, $p < 0.001$,

vocoder control versus unprocessed multiple, $p < 0.05$. A post-hoc comparison using Dunnett's tests revealed a significant difference between scores in block 2 and block 5 for the vocoder control condition, $p < 0.05$, and a trend toward a significant difference between the scores in block 2 and block 5 for the unprocessed single-talker condition, $p = 0.09$. Results are shown in Fig. 5.6.

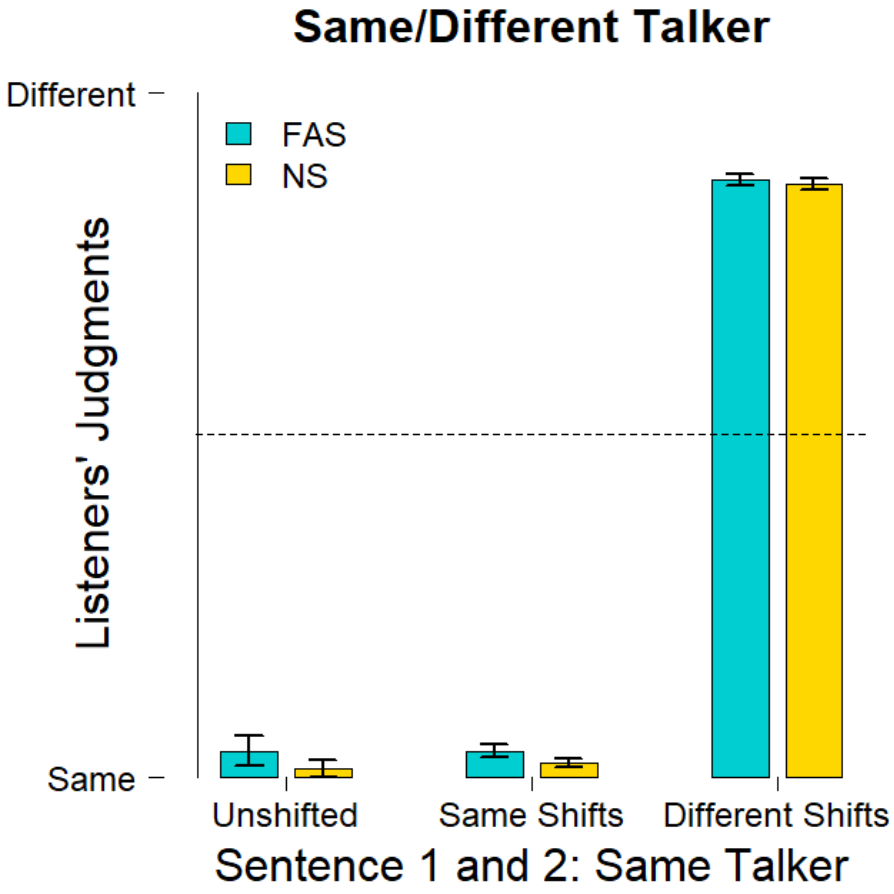

Fig. 5.1. Perception of talker identity for spectrally-shifted speech: same talker (n=80 per accent condition: NS = native speech, FAS = foreign-accented speech; n=16 per shift factor condition). Judgments were binary (same or different talkers). The Unshifted condition represents stimuli for which the talkers were the same in a given trial and their speech was unprocessed. The Same Shifts condition comprises trials for which the stimuli were processed with the same shift factor and the talkers for each stimulus were the same. The Different Shifts condition represents when the stimuli in each trial were processed with different shift factors, and the talkers for each

stimulus were the same. Chance performance is given by the horizontal dotted line. Standard error of the means (*SEM*) are also shown by the error bars.

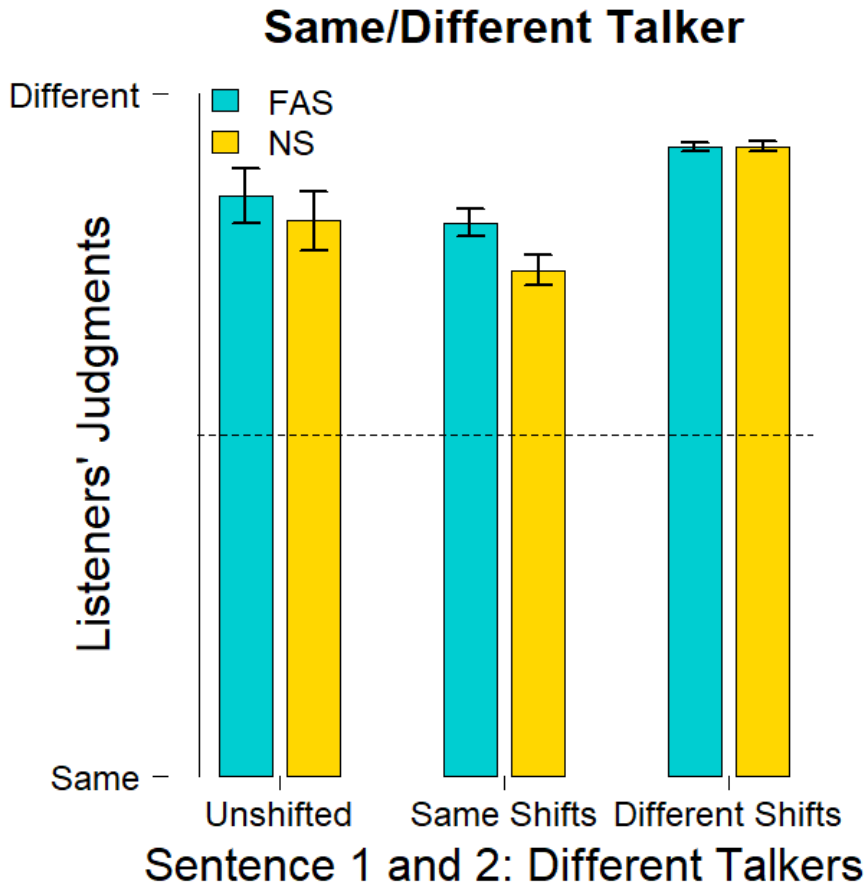

Fig. 5.2. Perception of talker identity for spectrally-shifted speech: different talkers (n=80 per accent condition; n=16 per shift factor condition). Judgments were binary (same or different talkers). The Unshifted condition represents when the talkers in each stimulus were different for a given trial, and their speech was unprocessed. The Same Shifts Condition is comprises trials for which the stimuli were processed with the same shift factor, and the talkers for each stimulus were different. The Different Shifts condition represents when the stimuli in each trial were processed with different shift factors, and the talkers for each stimulus were different. Chance performance is given by the horizontal dotted line. *SEM* bars are also given.

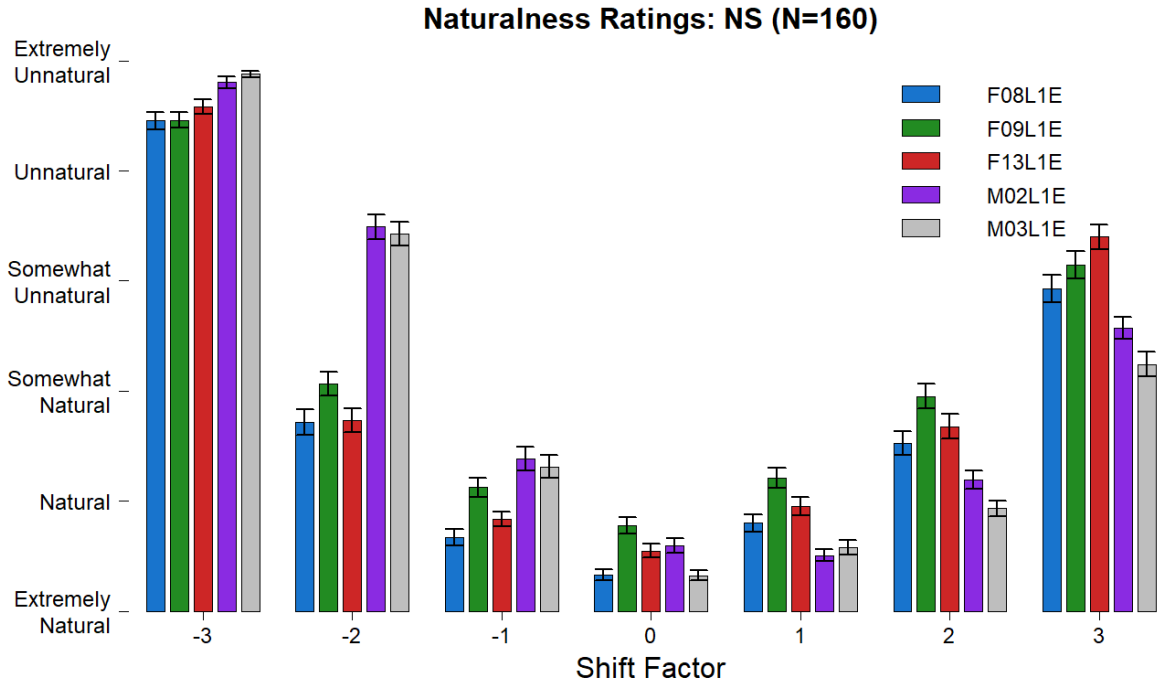

Fig. 5.3. Mean naturalness ratings for each talker in the native-accented speech (NS) condition across shift factors. Talker gender is denoted by the first letter in the talker code (F = female talker; M = male talker). A shift of 0 entails that no shifts were made to the speech stimuli. Positive shift factors indicate upward scaling of the spectral envelope and fundamental frequency for each sentence, and negative shift factors indicate downward scaling of the spectral envelope and fundamental frequency for each sentence. The spectral envelope of each sentence was shifted up or down by a scale factor of 1 + 0.08 entailing an 8% shift. The fundamental frequency for each sentence was shifted up or down by a scale factor of 1 + 0.296 entailing about a 30% shift. *SEM* bars are also shown.

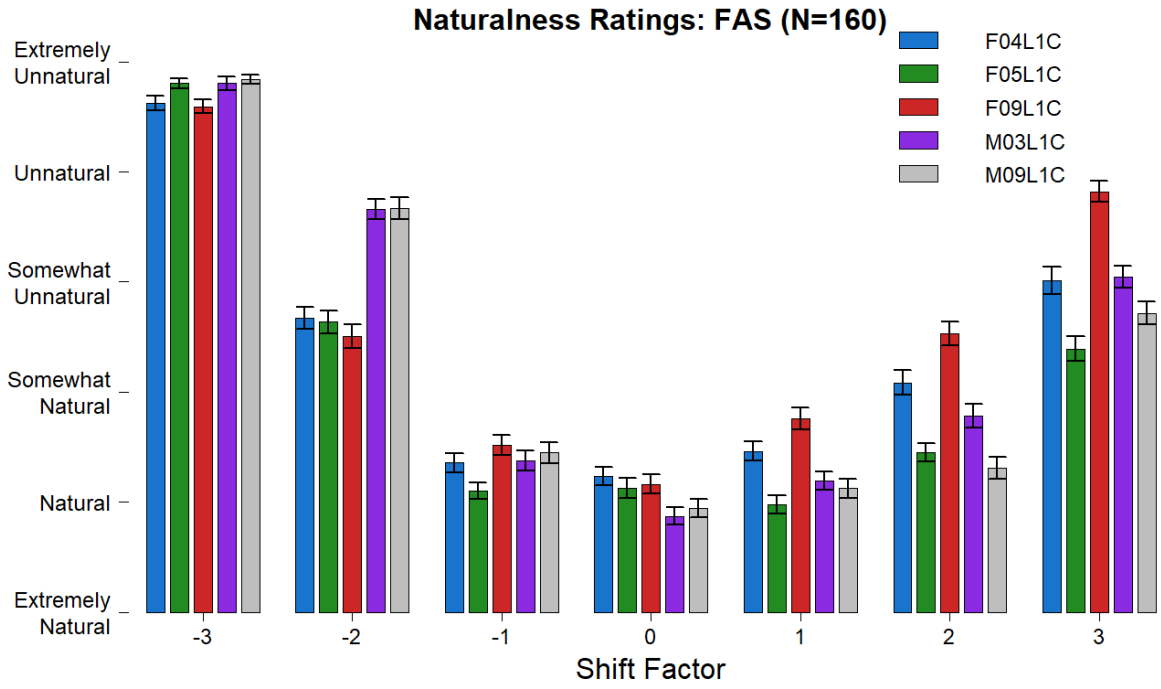

Fig. 5.4. Mean naturalness ratings for each talker in the foreign-accented speech (FAS) condition across shift factors. Talker gender is denoted by the first letter in the talker code. A shift of 0 entails that no shifts were made to the speech stimuli. Positive shift factors indicate upward scaling of the spectral envelope and fundamental frequency for each sentence, and negative shift factors indicate downward scaling of the spectral envelope and fundamental frequency for each sentence. The spectral envelope of each sentence was shifted up or down by a scale factor of 1 + 0.08 entailing about an 8% shift. The fundamental frequency for each sentence was shifted up or down by a scale factor of 1 + 0.296 entailing about a 30% shift. *SEM* bars are also shown.

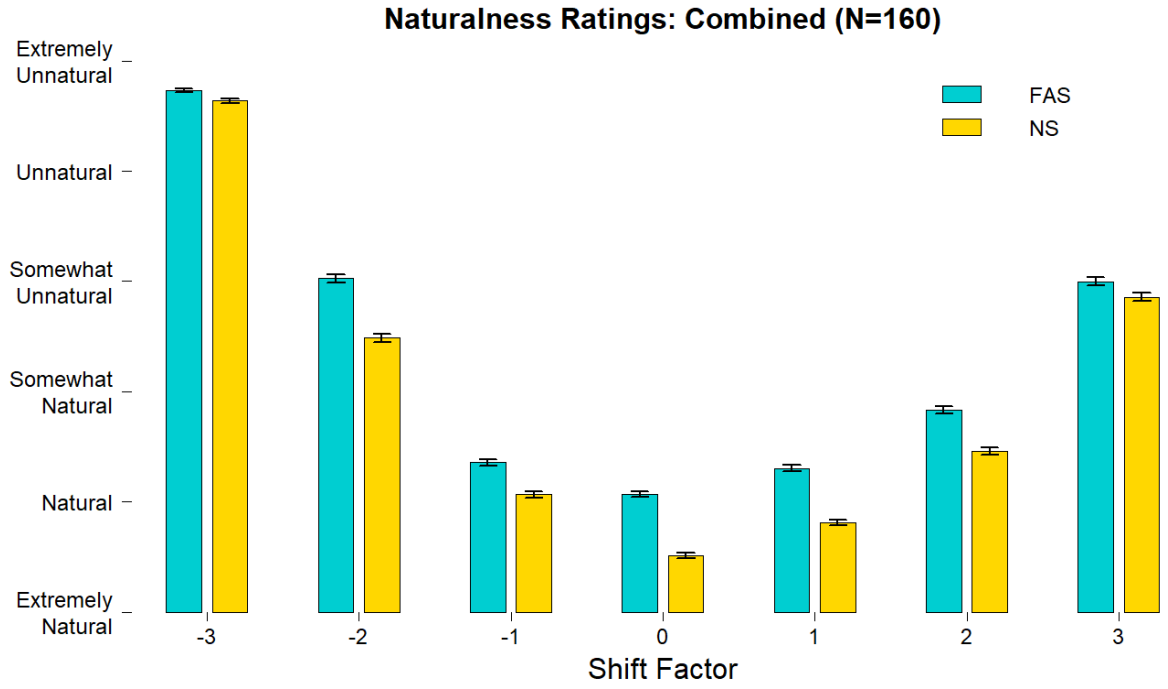

Fig. 5.5 Mean naturalness ratings for each accent condition across shift factors. A shift of 0 entails that no shifts were made to the speech stimuli. Positive shift factors indicate upward scaling of the spectral envelope and fundamental frequency for each sentence, and negative shift factors indicate downward scaling of the spectral envelope and fundamental frequency for each sentence. The spectral envelope of each sentence was shifted up or down by a scale factor of 1 + 0.08 entailing about an 8% shift. F0 for each sentence was shifted up or down by a scale factor of 1 + 0.296 entailing about a 30% shift. *SEM* bars are also shown.

Together, these results show that spectrally-shifting speech from a single foreign-accented talker to simulate multiple talkers leads to perceptual patterns that were also observed when listeners heard multiple foreign-accented talkers. Specifically, intelligibility scores in the simulated multiple-talker condition and in the multiple-talker condition are much lower than intelligibility scores for perception in the single-foreign-accented talker condition. These results also indicate that there is an initial detriment to perception of unshifted vocoded foreign-accented speech, but listeners were able to adapt by the 5th block. No adaptation occurred in other

conditions, although there was a trend toward adaptation for the unprocessed single-talker condition, and intelligibility was much higher, overall, in this condition, even early on (in block 2) as compared to the other conditions.

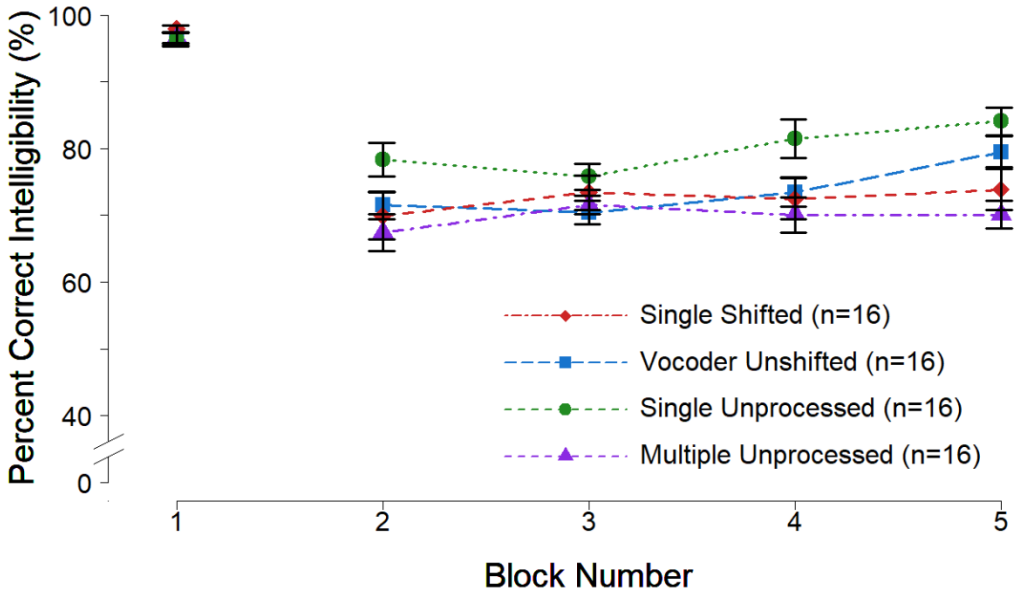

Fig. 5.6. Mean intelligibility scores expressed as percent correct across blocks. *SEM* bars are also shown. Block 1 (practice session) was not included in statistical analyses.

5.4 Discussion

The research presented in previous chapters provides evidence that talker variability is more detrimental when listening to foreign-accented talkers compared to when listening to native-accented talkers. We also showed that limiting listeners' access to spectral information further impairs perception of foreign-accented speech. Here, to further examine the role of spectral cues in perception of foreign-accented speech, we shifted the spectral envelope and F0 of single talkers to simulate multiple different talkers. We hypothesized that if listeners were focusing on these specific spectral cues, then they would be sensitive to changes in this domain

despite all other elements of the speech being left unchanged. Specifically, we expected that the pattern of mean intelligibility scores from listeners who were exposed to speech from a single foreign-accented talker that was simulated (via manipulations solely in the spectral domain) to sound like speech from five different foreign-accented talkers would match the pattern observed for listeners who were exposed to speech from multiple talkers rather than if they were exposed to unprocessed speech from the same foreign-accented talker. Given our previous results, we also expected that mean intelligibility scores would not improve with increased exposure in the simulated multiple-talker condition (adaptation would *not* occur). These results were confirmed, providing further evidence that listeners rely heavily on spectral cues when listening to foreign-accented speech.

A control experiment was performed to determine whether the spectral shifts used in the intelligibility experiment convinced listeners that that they were hearing different talkers when, in fact, they were hearing the same talker whose speech was processed to sound like different talkers. The results from this experiment showed that listeners were convinced that they were hearing different talkers when listening to spectrally-shifted speech from the same talker. Interestingly, when listeners heard spectrally-shifted speech from *different* talkers for each trial, they were better able to discriminate that the two talkers were different if the talkers had a foreign accent compared to when the talkers were native-accented. This implies that listeners are more sensitive to changes in talker identity for foreign-accented speech. This finding was beyond the scope of this dissertation, but it warrants further analyses for future research.

Another control experiment was conducted to explore whether the spectral scaling would have an effect on perceived naturalness of the speech stimuli. Results were in-line with previous

results reported in the literature, namely that downward spectral scaling of speech from male talkers is perceived as less natural, and extreme downward shifts were perceived as extremely unnatural for speech from male and female talkers. Also, there was an overall tendency for listeners to report that foreign-accented speech sounded less natural even when speech was unprocessed. This finding could be due to either a listener-biased effect or from variations from native-accented speech in the temporal and/or spectral domain (likely a combination of listener biases *and* deviations from target speech norms (Yi *et al*., 2014)).

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSION

Phonologically equivalent utterances can exhibit a high degree of acoustic variability across different talkers, yet listeners are able to resolve this problem relatively easily. Talker normalization is a theoretical framework that explains how listeners are able to resolve the problem of talker variability by relying on talker-specific spectral patterns recovered from the speech signal. Given the greater degree of variability that can often be found in foreign-accented speech, it was an overarching goal of this dissertation to consider how such spectral patterns provide a consistent source of information that listeners rely on when perceiving foreign-accented speech.

Frequency information has been shown to provide important cues for listening to native-accented speech in various adverse conditions, such as in background noise (Loizou *et al*., 2003; Shannon *et al*., 2004). Although foreign-accented speech can also be considered an adverse listening condition, the distortions reside within the speech signal, itself, rather than extrinsic to it, as is the case for listening to speech in background noise. Perceptual distortions arising from foreign-accented speech are partially observed in the spectral domain. It has been reported that frequency information can provide specific cues that listeners rely on when listening to foreign-accented speech. Arslan and Hansen showed that the midfrequency range (1500-2500 Hz) is the most sensitive band regarding foreign-accented talker variations (1997a). In a separate study, Arslan and Hansen (1997b) also showed that listeners are more sensitive to specific frequency ranges depending on the perception task: for intelligibility/speech recognition, the first formant (F1) was shown to be the most important; for accent discrimination ability, F2 and F3 were

shown to be more important. These results provide evidence that listeners are particularly sensitive to the spectral information when perceiving foreign-accented speech. The work that comprises this dissertation further explored the role that spectral information provides when perceiving foreign-accented speech. Specifically, these experiments examined listeners' dependence on talker-specific spectral cues by testing perception of foreign-accented speech in conditions where spectral information was manipulated in single- and multiple-talker conditions:

1. In Chapter 3, we tested how varying the spectral resolution affects intelligibility and accent detection and showed that the number of spectral channels needed for perception of foreign-accented speech was much greater than for perceiving native-accented speech in terms of intelligibility and accent detection (Fig. 3.1, Fig. 3.2, and Fig. 3.3). Although we did not increase the spectral resolution beyond 9 channels, it is hypothesized that further increasing the number of channels would continue to increase intelligibility scores for perception of foreign-accented speech, though only to a certain extent. With vocoded native speech, a plateau was reached at only 9 channels, and this plateau could have occurred with fewer channels (we did not test performance with 6, 7 or 8 channels). Nevertheless, we would predict that there would be some additional benefit gained by increasing the spectral resolution beyond 9 channels for perceiving foreign-accented speech, since it has been shown that more channels are needed for other adverse listening conditions, such as listening to native speech with background noise (Shannon *et al.*, 2004). A future experiment could determine whether intelligibility performance would continue to increase to the level of performance when perceiving unprocessed foreign-accented speech by adding additional channels. Determining such a limit could motivate research to improve upon pre-existing technology that currently offers limited and/or distorted spectral

resolution, such as CI devices, since interactions between native and non-native talkers are becoming commonplace in speech communication.

2. In Chapter 4, we compared how limiting spectral resolution affects listeners' abilities to adapt to foreign-accented speech with increased exposure in single- and multiple-(interleaved) talker conditions. Intelligibility scores were much lower, initially, but listeners were able to adapt with increased exposure, though only if they were exposed to the same talker over time (Fig. 4.1 and Fig. 4.2). Although it is beyond the scope of this dissertation, it is speculated that listeners were able to rely (at least partly) on temporal cues within the speech signal to aid with perception when spectral resolution was limited, since listeners adapted to hearing a single foreign-accented talker with limited spectral resolution (Fig. 4.1). In this experiment, temporal cues were not manipulated. It has been shown that there is less within-talker variation for duration patterns in foreign-accented speech compared to native speech (Baker *et al*., 2011). Given the consistent temporal patterns that they reported for an individual foreign-accented talker, in our experiment, listeners could have utilized these consistent temporal cues in a talker-dependent manner when adapting to foreign-accented speech from a single talker with reduced spectral resolution. Baker and colleagues (2011) also showed that there is more between-talker variation for foreign-accented speech than for native speech. Since temporal patterns across different foreign-accented talkers are more variable, that could be why we saw no improvement with increased exposure in our condition where listeners were exposed to multiple foreign-accented talkers with increased exposure (Fig. 4.2). Other sources of information, such as lexical context, could have also assisted listeners with perception; however, this cannot explain the

difference in performance observed when listeners heard the same foreign-accented sentences spoken by either a single or several different talkers over time.

3. In Chapter 4, we also compared intelligibility scores for unprocessed foreign-accented speech, where spectral resolution was not limited, in single- and multiple-talker conditions. We found that intelligibility scores were higher, overall, when listeners were exposed to unprocessed speech from the same foreign-accented talker over time (talker-dependent perception) compared to when they heard unprocessed speech from five different foreign-accented talkers (accent-dependent perception) (Fig. 4.1 and Fig. 4.2). It has been shown that listeners adapt to variability in native-accented speech in a talker-dependent manner, where indexical properties stemming from a single vocal source can facilitate perceptual learning (*e.g*., Nygaard *et al*., 1994; Nygaard & Pisoni, 1998). However, contrary results have been reported in the literature regarding adaptation to a single or multiple foreign-accented talkers. Bent and Holt (2013) also found that perception was better when listening to speech spoken by a single foreign-accented talker compared to multiple foreign-accented talkers who share the same native language. Bradlow and Bent (2008) reported no difference in listeners' performance when trained with speech from the same Chinese-accented talker who was presented during testing or multiple Chinese-accented talkers. These reported differences could be due to variations in stimuli as well as training paradigms across studies. Perhaps a more discernable reason for the different conclusions reported is that differences in baseline intelligibility scores for each talker are present. Bent and Holt stated that baseline intelligibility scores across conditions were closely matched. Bradlow and Bent noted baseline intelligibility scores for each talker in their multi-talker condition ranging from 79 to 88 percent, but only 74 percent in their talker-specific (single talker)

condition. Baseline intelligibility scores for the subset of Chinese-accented talkers presented in our experiments ranged from 58 to 74 percent. Also, the talker was randomly chosen from one of the talkers presented in the multiple talker condition for listeners assigned to our single-talker condition. Despite our baseline intelligibility scores being lower than the talkers presented in Bradlow and Bent's study, our listeners still achieved high intelligibility scores in the single talker condition after exposure to just a few sentences. This result provides evidence against the possibility that our stimuli were too difficult or that our baseline intelligibility scores were too low for adaptation to occur. It, therefore, seems more plausible that listeners find it less difficult to rapidly adapt to foreign-accented speech in a talker-dependent manner rather than in an accent-dependent (talker-independent) manner. Another important point to consider is listener familiarity with foreign-accented speech. Listener familiarity with a particular foreign accent has been shown to affect perceptual adaptation to foreign-accented speech, where increased familiarity with a particular accent aids with perceptual learning (Witteman *et al*., 2013). Ongoing work is being conducted to investigate whether our results are influenced by listeners' familiarity with Chinese-accented English. Listeners who participated in these experiments reported variable exposure to Chinese-accented English. Further meta-analyses could show if listeners with more familiarity to Chinese-accented English had even higher intelligibility scores (more adaptation) in the single foreign-accented talker condition. Also, listeners who reported having more exposure to Chinese-accented English may have adapted in the multiple-foreign-accented talker condition.

4. Chapter 5 examines how spectrally-shifted speech (F0 and the spectral envelope) affects listeners' abilities to discriminate between the same or different talkers. Previous research

has shown that listeners can utilize spectral properties in speech to obtain indexical information about a particular talker. For example, listeners can identify a talker's gender with exposure to only formant frequencies (Fellowes *et al*., 1997). Source characteristics, such as F0 and vocal tract length also convey information about individual talkers (Bachorowski & Owren, 1999). Accurate talker identification has been shown to be important for speech perception. When talker presentation varies rapidly, listeners are slower and less accurate in identifying speech sounds (Mullennix & Pisoni, 1990), and listeners are able to recall fewer words when perceiving speech from multiple talkers (Martin *et al*., 1989). Talker normalization is a relatively easy, even unconscious, perceptual process for native speech, but, given our results from Chapter 4, this process becomes even more critical when perceiving foreign-accented speech. Here, we manipulated the spectral envelope and F0 of a single talker up or down by 8% and 30%, respectively, and observed that listeners consistently judged spectrally-shifted speech from the same talker as being a different talker, as long as the shift factors were different across the two sentences in each trial (Fig. 5.1 and Fig. 5.2). Interestingly, when F0 is removed from the speech signal, listeners can still accurately identify a particular talker (Fellowes *et al*., 1997). In this experiment, however, we observed that listeners were particularly sensitive to F0 because shifting F0, together with the spectral envelope, convinced listeners that they were hearing different talkers. Point 6, below, presents a summary of our findings regarding how perceiving the same talker as a different talker can be detrimental to intelligibility of foreign-accented speech.

5. In Chapter 5, we also investigated how spectrally-shifted speech can affect perceived naturalness. Our results (displayed in Fig. 5.3, Fig. 5.4, and Fig. 5.5) were consistent with

previous studies reporting that downward scaling of speech from male talkers results in judgements perceived as being less natural, and spectrally scaling speech from male and female talkers downward beyond the normal human range results in speech being perceived as extremely unnatural (Assmann & Nearey, 2008). Also, foreign-accented speech was judged as being less natural than native-accented speech even when the spectral information was unshifted. In this experiment, listeners were explicitly instructed *not* to rate the level of naturalness for each talker based on perceived intelligibility nor on perceived level of foreign-accentedness. Given our results presented for this experiment, it could entail that our observations for adaptation to spectrally-shifted foreign-accented speech are, at least partially, correlated with naturalness. Future investigations should examine this possibility, as the vocoder used to manipulate the spectral envelope and F0 could have resulted in detrimental effects to foreign-accented speech stimuli compared to native speech stimuli.

6. Chapter 5 also covers our analysis of the intelligibility scores of listeners who were exposed over time to spectrally-scaled foreign-accented speech that simulated a multiple-talker condition, compared to listeners exposed to either unprocessed speech from the same foreign-accented talker or to unprocessed speech from five different foreign-accented talkers who shared the same foreign-accent. We found that performance patterns for the simulated multiple-talker condition matched performance for the unprocessed multiple-talker condition, despite only manipulating the spectral envelope and F0 and maintaining the original temporal patterns (Fig. 5.6). Contrary to the previous conjecture that listeners may have adapted to a single foreign-accented talker with limited resolution by relying on temporal patterns, this does not seem evident in our results. Temporal patterns remained unchanged in the spectrally-shifted single-

talker (simulated multi-talker) condition, yet intelligibility was as low as in the condition where listeners heard unprocessed speech from different foreign-accented talkers who share the same native language. Conflicting evidence has been reported regarding the role of temporal cues when perceiving foreign-accented speech. Although it is beyond the scope of this dissertation, it is worth mentioning some of these results. Tajima and coworkers (1997) found that intelligibility of Chinese-accented English improved when temporal modifications were made to align the duration of the accented speech stimuli with the duration of native English speech while maintaining the spectral and source characteristics of the Chinese-accented talker. They also reported that intelligibility performance of speech from a native English talker decreased when temporal patterns were aligned with the duration patterns of Chinese-accented English. Other studies have reported that, although foreign-accented talkers are rated as being more fluent if their speaking rate is faster (Lennon, 1990), comprehension actually drops when listening to foreign-accented speech with a faster speaking rate (*e.g.*, Anderson-Hsieh & Koehler, 1988). In short, although our results indicate the importance of talker-dependent spectral patterns for perception of foreign-accented speech, listeners are also sensitive to temporal information.

Results from this work have both theoretical and clinical implications. The findings presented in this dissertation expand upon previously existing research attempting to outline the perceptual process involved in foreign-accented speech perception in single- and multiple-talker conditions for normal-hearing listeners in quiet conditions. Another implication is that the perception of foreign-accented speech would be compromised in situations where fine spectral detail is limited, such as when talking on the phone, using VoIP systems or when using speech recognition devices.

Some of our experiments tested the ability of listeners to process foreign-accented speech with limited spectral resolution. This signal processing was done using a tone vocoder to limit the number of frequency bands in the speech signal, which, therefore, limits the spectral cues available to listeners. For each band, the envelope was extracted using full-wave rectification followed by low-pass filtering, and the envelope was used to modulate the tone. This process simulates listening to speech with a multi-channel CI device. Given this, our results could also offer valuable insight into how CI users perceive foreign-accented speech, and how this process might be more difficult for them when they are exposed to several different foreign-accented talkers. Decreased spectral resolution is only one issue that CI users experience when perceiving speech. CI users also commonly incur frequency-place mismatch due to limitations on the insertion depth of the device's electrode arrays. Earlier studies simulated this effect in normal-hearing listeners and found that the resulting frequency shift caused a decrease in perceptual performance for sentences, vowels and consonants (Dorman *et al*., 1997; Fu & Shannon, 1999; Zhou *et al*., 2010). Here, we extended these findings by testing perceptual performance of spectrally-shifted sentences produced by foreign-accented talkers. Although we did not directly test CI users in our experiments, it can be inferred from our results that the added difficulties reported by CI users when they are perceiving foreign-accented speech are partly due to their limited access to talker-specific spectral cues within the speech signal. It is difficult to test the effect of frequency-place mismatch, alone, in CI users due to other perceptual confounds *(e.g*., limited spectral resolution, cross-channel interaction).

Another interesting application of this work could explore perceptual adaptation to various other types of adverse listening conditions. As aforementioned, it has already been

shown that perception of native speech in various types of adverse listening conditions requires greater spectral resolution (Shannon *et al.*, 2004). Given our results presented in Chapter 3, it is also the case that listening to foreign-accented speech requires more spectral detail than when listening to native speech. Other adverse listening conditions where the distortion resides within the signal, such as perceiving dysarthric speech, were not examined here; however, it seems plausible to hypothesize that our results would generalize to perception of dysarthric speech. Specifically, perception of dysarthric speech would require greater spectral resolution. It could also be speculated that listeners would adapt to a single talker with dysarthria more easily than to multiple talkers with dysarthria. Some relevant data were reported by Borrie *et al*., (2017), who found that listeners benefited from familiarization with a talker with dysarthria, and adaptation was greater when training was with a talker who shared similar perceptual features as the talker presented in the testing session.

In conclusion, these results provide evidence for the hypothesis that spectral information is an important source within the speech signal that aids listeners when perceiving foreign-accented speech by providing listeners with talker-specific invariant representations of highly variable inputs. Perceptual performance decreased when manipulating these cues either by reducing the spectral resolution, or by presenting listeners with foreign-accented speech in a multiple-talker condition or by introducing spectral shifts. These results, indicating the importance of spectral information when perceiving foreign-accented speech, suggest that a talker normalization framework may account for this specific perceptual process.

APPENDIX A

HARVARD SENTENCES

'S_07_01.wav'	'We talked of the sideshow in the circus.'
'S_07_02.wav'	'Use a pencil to write the first draft.'
'S_07_03.wav'	'He ran halfway to the hardware store.'
'S_07_04.wav'	'The clock struck to mark the third period.'
'S_07_05.wav'	'A small creek cut across the field.'
'S_07_06.wav'	'Cars and buses stalled in snow drifts.'
'S_07_07.wav'	'The set of china hit the floor with a crash.'
'S_07_08.wav'	'This is a grand season for hikes on the road.'
'S_07_09.wav'	'The dune rose from the edge of the water.'
'S_07_10.wav'	'Those words were the cue for the actor to leave.'
'S_11_01.wav'	'Oak is strong and also gives shade.'
'S_11_02.wav'	'Cats and dogs each hate the other.'
'S_11_03.wav'	'The pipe began to rust while new.'
'S_11_04.wav'	'Open the crate but don't break the glass.'
'S_11_05.wav'	'Add the sum to the product of these three.'
'S_11_06.wav'	'Thieves who rob friends deserve jail.'
'S_11_07.wav'	'The ripe taste of cheese improves with age.'
'S_11_08.wav'	'Act on these orders with great speed.'
'S_11_09.wav'	'The hog crawled under the high fence.'
'S_11_10.wav'	'Move the vat over the hot fire.'

'S_16_01.wav'	'The empty flask stood on the tin tray.'
'S_16_02.wav'	'A speedy man can beat this track mark.'
'S_16_03.wav'	'He broke a new shoelace that day.'
'S_16_04.wav'	'The coffee stand is too high for the couch.'
'S_16_05.wav'	'The urge to write short stories is rare.'
'S_16_06.wav'	'The pencils have all been used.'
'S_16_07.wav'	'The pirates seized the crew of the lost ship.'
'S_16_08.wav'	'We tried to replace the coin but failed.'
'S_16_09.wav'	'She sewed the torn coat quite neatly.'
'S_16_10.wav'	'The sofa cushion is red and of light weight.'
'S_19_01.wav'	'Acid burns holes in wool cloth.'
'S_19_02.wav'	'Fairy tales should be fun to write.'
'S_19_03.wav'	'Eight miles of woodland burned to waste.'
'S_19_04.wav'	'The third act was dull and tired the players.'
'S_19_05.wav'	'A young child should not suffer fright.'
'S_19_06.wav'	'Add the column and put the sum here.'
'S_19_07.wav'	'We admire and love a good cook.'
'S_19_08.wav'	'There the flood mark is ten inches.'
'S_19_09.wav'	'He carved a head from the round block of marble.'
'S_19_10.wav'	'She has a smart way of wearing clothes.'
'S_22_01.wav'	'The cement had dried when he moved it.'
'S_22_02.wav'	'The loss of the second ship was hard to take.'

'S_22_03.wav'	'The fly made its way along the wall.'
'S_22_04.wav'	'Do that with a wooden stick.'
'S_22_05.wav'	'Live wires should be kept covered.'
'S_22_06.wav'	'The large house had hot water taps.'
'S_22_07.wav'	'It is hard to erase blue or red ink.'
'S_22_08.wav'	'Write at once or you may forget it.'
'S_22_09.wav'	'The doorknob was made of bright clean brass.'
'S_22_10.wav'	'The wreck occurred by the bank on Main Street.'
'S_26_01.wav'	'Yell and clap as the curtain slides back.'
'S_26_02.wav'	'They are men who walk the middle of the road.'
'S_26_03.wav'	'Both brothers wear the same size.'
'S_26_04.wav'	'In some form or other we need fun.'
'S_26_05.wav'	'The prince ordered his head chopped off.'
'S_26_06.wav'	'The houses are built of red clay bricks.'
'S_26_07.wav'	'Ducks fly north but lack a compass.'
'S_26_08.wav'	'Fruit flavors are used in fizz drinks.'
'S_26_09.wav'	'These pills do less good than others.'
'S_26_10.wav'	'Canned pears lack full flavor.'
'S_53_01.wav'	'Press the pedal with your left foot.'
'S_53_02.wav'	'Neat plans fail without luck.'
'S_53_03.wav'	'The black trunk fell from the landing.'
'S_53_04.wav'	'The bank pressed for payment of the debt.'

'S_53_05.wav' 'The theft of the pearl pin was kept secret.'

'S_53_06.wav' 'Shake hands with this friendly child.'

'S_53_07.wav' 'The vast space stretched into the far distance.'

'S_53_08.wav' 'A rich farm is rare in this sandy waste.'

'S_53_09.wav' 'His wide grin earned many friends.'

'S_53_10.wav' 'Flax makes a fine brand of paper.'

'S_56_01.wav' 'The small red neon lamp went out.'

'S_56_02.wav' 'Clams are small, round, soft, and tasty.'

'S_56_03.wav' 'The fan whirled its round blades softly.'

'S_56_04.wav' 'The line where the edges join was clean.'

'S_56_05.wav' 'Breathe deep and smell the piney air.'

'S_56_06.wav' 'It matters not if he reads these words or those.'

'S_56_07.wav' 'A brown leather bag hung from its strap.'

'S_56_08.wav' 'A toad and a frog are hard to tell apart.'

'S_56_09.wav' 'A white silk jacket goes with any shoes.'

'S_56_10.wav' 'A break in the dam almost caused a flood.'

'S_58_01.wav' 'It is a band of steel three inches wide.'

'S_58_02.wav' 'The pipe ran almost the length of the ditch.'

'S_58_03.wav' 'It was hidden from sight by a mass of leaves and shrubs.'

'S_58_04.wav' 'The weight of the package was seen on the high scale.'

'S_58_05.wav' 'Wake and rise, and step into the green outdoors.'

'S_58_06.wav' 'The green light in the brown box flickered.'

'S_58_07.wav'	'The brass tube circled the high wall.'
'S_58_08.wav'	'The lobes of her ears were pierced to hold rings.'
'S_58_09.wav'	'Hold the hammer near the end to drive the nail.'
'S_58_10.wav'	'Next Sunday is the twelfth of the month over.'
'S_72_01.wav'	'A gold ring will please most any girl.'
'S_72_02.wav'	'The long journey home took a year.'
'S_72_03.wav'	'She saw a cat in the neighbor’s house.'
'S_72_04.wav'	'A pink shell was found on the sandy beach.'
'S_72_05.wav'	'Small children came to see him.'
'S_72_06.wav'	'The grass and bushes were wet with dew.'
'S_72_07.wav'	'The blind man counted his old coins.'
'S_72_08.wav'	'A severe storm tore down the barn.'
'S_72_09.wav'	'She called his name many times.'
'S_72_10.wav'	'When you hear the bell, come quickly.'

APPENDIX B

PRODUCTION QUESTIONNAIRE

Age:

Gender:

Do you have any hearing or speech impairments:

If yes, please describe:

Country of Birth:

Country, city and state/province where you have lived most of your life:

What is your primary language:

List any languages that you speak and/or understand fluently (including different dialects of Chinese):

How old were you when you started to learn English:

Was it American English or another dialect of English, such as British English:

How long have you been residing in the United States of America:

Have you lived in any other countries besides the USA and your country of birth?

If yes, which country/countries and for how long in each:

Are you an undergraduate student at UTD:

If you are a graduate student/other, where (country, city, state/province) did you obtain your degree(s):

How often do you interact with native speakers of American English outside of school (choose only one):

a. **Often, I have several close friends who are American and/or my roommate is American**
b. **Sometimes, during activities outside of school such as at church or clubs**
c. **Rarely, only when necessary**

Thank you for filling out this questionnaire!

For your privacy, this data will not be directly linked to you in any way.

<u>THIS PORTION IS TO BE FILLED OUT BY THE RESEARCHER</u>

Participant Code:
Name of Researcher Obtaining Data:
Date:
Additional Comments:

APPENDIX C

PERCEPTION QUESTIONNAIRE

Age:

Gender:

Major/Degree(s):

Do you have any hearing or speech impairments:

If yes, please describe:

Country, city and state/province of birth:

Country, city and state/province where you have lived most of your life:

What is your primary language:

List any languages that you speak and/or understand fluently:

How old were you when you started to learn English:

Was it American English or another dialect of English, such as British English:

How long have you been residing in the United States of America:

Have you lived in any other countries besides the USA and your country of birth?

If yes, which country/countries and for how long in each:

How often do you interact with non-native speakers of American English (choose only one):

d. **Often, I have several close friends/family who are from foreign countries and/or my roommate is from a foreign country**
e. **Sometimes, during activities outside of school such as at work, church or clubs**
f. **Rarely, only when necessary, such as in a classroom setting**

How familiar are you with American English spoken with a Chinese accent (choose only one):

a. **Very familiar, I have several close friends/family who are Chinese and/or my roommate is from a country where s/he speaks Chinese primarily**
b. **Somewhat familiar, I have socialized with native speakers of Chinese during activities outside of school such as at work, church or clubs**
c. **Not very familiar, I have heard Chinese people speak at school on occasion**
d. **Not at all familiar, I cannot tell the difference between Chinese accents or other Asian accents**

Thank you for filling out this questionnaire!

For your privacy, this data will not be directly linked to you in any way.

THIS PORTION IS TO BE FILLED OUT BY THE RESEARCHER

Participant Code:
Name of Researcher Obtaining Data:
Date:
Additional Comments:

REFERENCES

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, 38, 561-613.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning,* 42, 529-555.

Arslan, L.M., & Hansen, J.H.L. (1997a). Frequency characteristics of foreign accented speech. *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing,* 5, 1123-1126.

Arslan, L.M., & Hansen, J.H.L. (1997b). A study of temporal features and frequency characteristics in American English foreign accent. *Journal of the Acoustical Society of America*, 102, 28-40.

Assmann, P.F., Nearey, T.M., & Hogan, J.T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America,* 71, 975-989.

Assmann, P.F., Dembling, S., & Nearey, T.M. (2006). Effects of frequency shifts on perceived naturalness and gender information in speech. *Proceedings of the Ninth International Conference on Spoken Language Processing*, 889-892.

Assmann, P.F., & Nearey, T.M. (2007). Effects of frequency shifts on the identification of vowels and words in sentences. *Journal of the Acoustical Society of America,* 122, 3064-3065.

Assmann, P.F., & Nearey, T.M. (2008). Identification of frequency-shifted vowels. *Journal of the Acoustical Society of America*, 124, 3203-3212.

Assmann, P.F., Kapolowicz, M.R., Massey, D.A., Barreda, S., & Nearey, T.M. (2014). Perception of speaker sex in re-synthesized children's voices. *Journal of the Acoustical Society of America,* 135, 2424.

Assmann, P.F., Kapolowicz, M.R., Massey, D.A., Barreda, S., & Nearey, T.M. (2015). Links between the perception of speaker age and sex in children's voices. *Journal of the Acoustical Society of America,* 138, 1181

Bachorowsky, J.A., & Owren, M.J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America,* 106, 1054-1063.

Baese-Berk, M.M., Bradlow, A.R., & Wright, B.A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America,* 133, EL174-EL180.

Baese-Berk, M.M., & Morrill, T.H. (2015). Speaking rate consistency in native and non-native speakers of English. *Journal of the Acoustical Society of America Express Letters*, 138, EL223-EL228.

Baker, R.E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K.J., and Bradlow, A.R. (2011). Word durations in non-native English. Journal of Phonetics, 39, 1.

Bent, T., Loebach, J.L., Phillips, L., & Pisoni, D.B. (2011). Perceptual adaptation to sinewave-vocoded speech across languages. *Journal of Experimental Psychology: Human Perception and Performance,* 37, 1607-1616.

Bent, T. & Holt, R.F. (2013). The influence of talker and foreign-accent variability on spoken word identification. *Journal of the Acoustical Society of America,* 133, 1677-1686.

Borrie, S.A., Lansford, K.L., & Barrett, T.S. (2017). *Journal of Speech, Language, and Hearing Research,* 60, 3110-3117.

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

Burk, M.H., & Humes, L.E. (2007). Effects of training on speech recognition performance in noise using lexically challenging words. *Journal of Speech, Language, and Hearing Research*, 50, 25-40.

Chang, Y. & Fu, Q.J. (2006). Effects of talker variability on vowel recognition in cochlear implants. *Journal of Speech, Language, and Hearing Research*, 49, 1331-1341.

Chang, C.Y., & Fox, R.A. (2010). Production and perception of lexical tones in Beijing and Taiwan Mandarin. *Journal of the Acoustical Society of America,* 127, 2023.

Clarke, C.M., & Garrett, M.F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America,* 116, 3647-3658.

Cleary, M., & Pisoni, D.B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. *The Annals of Otology, Rhinology, and Laryngology,* 111, 113-118.

Cleary, M., Pisoni, D.B., & Kirk, K.I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 48, 204-223.

Clopper, C.G., & Pisoni, D.B. (2004). Perceptual dialect categorization by an adult cochlear implant user: a case study. *International Congress Series,* 1273, 235-238.

Cooper, A. & Bradlow, A.R. (2016). Linguistically guided adaptation to foreign-accented speech. *Journal of the Acoustical Society of America: Express Letters*, 140, EL378-EL384.

Cousins, K.A.Q., Dar, H., Wingfield, A., & Miller, P. (2014). Acoustic masking disrupts time-dependent mechanisms of memory encoding in word-list recall. *Memory and Cognition.* 42, 622-638.

Daniloff, R.G., Shiner, T.H., & Zemlin, W.R. (1968). Intelligibility of vowels altered in duration and frequency. *Journal of the Acoustical Society of America*, 44, 700-707.

Dorman, M.F., Loizou, P.C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America*, 102, 2403-2411.

Dorman, M.F., Loizou, P.C., & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *Journal of the Acoustical Society of America*, 102, 2993-2996.

Eckert, M.A., Walczak, A., Denslow, S., Horwitz, A., & Dubno, J.R. (2008). Age-related effects on word recognition: reliance on cognitive control systems with structural declines in speech-responsive cortex. *Journal of the Association for Research in Otolaryngology*, 9, 252-259.

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, 4, 148.

Faulkner, K.F., & Pisoni, D.B. (2013). Some observations about cochlear implants: challenges and future directions. *Neuroscience Discovery*, DOI: 10.7243/2052-6946-1-9.

Fellowes, J.M., Remez, R.E., & Rubin, P.E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception and Psychophysics*, 59, 839-849.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1276.

Floccia, C., Bunler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research,* 38, 379-412.

Fu, Q.J., & Shannon, R.V. (1999). Effects of electrode configuration and frequency allocation on vowel recognition with the nucleus-22 cochlear implant. *Ear and Hearing*, 20, 332-344.

Graddol, D. (2006). *English Next* (The British Council, London).

Green, T., Katiri, S., Faulkner, A., & Rosen, S. (2007). Talker intelligibility differences in cochlear implant listeners. *Journal of the Acoustical Society of America,* 121, EL223-229.

Guest, D., Kapolowicz, M.R., Hossain, S., Montazeri, V., & Assmann, P.F. (2016). Perception of voice gender in cochlear implant simulations of children's speech. *Journal of the Acoustical Society of America,* 139, 2124

Guion, S., Flege, J.E., Liu, H.M., & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21, 205-228.

Hanulikova, A., & Weber, A. (2012). Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, 74, 613-629.

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

Heinrich, A. & Schneider, B.A. (2011). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *Quarterly Journal of Experimental Psychology,* 64, 186-205.

Hillenbrand, J.M., & Clark, M.J. (2009). The role of F0 and formant frequencies in distinguishing the voices of men and women. *Perception and Psychophysics*, 71, 1150-1166.

Holt, L.L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of the Acoustical Society of America,* 120, 2801-2817.

Huang, J., & Holt, L.L. (2012). Listening for the norm: Adaptive coding in speech categorization. *Frontiers in Psychology,* 3, 10.

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

Jenkins, J. (2000). *The phonology of English as an International Language*. Oxford, UK: Oxford University Press.

Ji, C., Galvin, J.J., Chang, Y., Xu, A., & Fu, Q.J. (2014). Perception of speech produced by native and nonnative talkers by listeners with normal hearing and listeners with cochlear implants. *Journal of Speech, Language, and Hearing Research* 57, 532-542.

Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative perception. *Language and Speech*, 34, 265-279.

Kapolowicz, M.R., Montazeri, V., & Assmann, P.F. (2016). The role of spectral resolution in foreign-accented speech perception. *Proceedings of Annual Conference of the International Speech Communication Association,* INTERSPEECH 2016, 3289-3293.

Kawahara, H. (1997). Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. *Proceedings of the ICASSP,* 1303-1306.

Kirk, K.I., Hay-McCutcheon, M., Sehgal, S.T., & Miyamoto, R.T. (2000). Speech perception in children with cochlear implants: effects of lexical difficulty, talker variability, and word length. *Annals of Otology, Rhinology & Laryngology*, 109, 79-81.

Lane, H. (1963). Foreign accent and speech distortion, *Journal of the Acoustical Society of America.* 35, 451-453.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.

Liang, E.J.C., Liu, R., Lotto, A.J., & Holt, L.L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology*, 3, *203*.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.

Loizou, P. C., Mani, A., & Dorman, M.F. (2003). Dichotic speech recognition in noise using reduced spectral cues. *Journal of the Acoustical Society of America,* 114, 204-223.

Luce, P.A., Feustel, T.C., & Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, 25, 17-32.

Martin, C.S., Mullennix, J.W., Pisoni, D.B., & Sommers, W.V. (1989) Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 676-684.

Mattys, S.L., Davis, M.H., Bradlow, A.R., & Scott, S.K. (2012). Speech recognition in adverse conditions: a review. *Language and Cognitive Processes,* 27, 953-978.

Morrill, T., Baese-Berk, M., & Bradlow, A.R. (2016). Speaking rate consistency and variability in spontaneous speech by native and non-native speakers of English. *Proceedings of the International Conference on Speech Prosody*, 2016, 1119-1123.

Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Mullennix, J.W., & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379-390.

Munro, M.J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition,* 20, 139-154.

Munro, M.J., & Derwing, T.M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*, 289-306.

Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, 85, 2088-2113.

Nearey, T.M. & Assmann, P.F. (2007). Probabilistic 'sliding-template' models for indirect vowel normalization, in Experimental Approaches to Phonology, eds. M. J. Solé, P. S., Beddor, and M. Ohala. Oxford University Press.

Nogaki, G., Fu, Q.J., & Galvin, J.J. III (2007). Effect of training rate on recognition of spectrally shifted speech. *Ear and Hearing,* 28, 132-140.

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

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

Padilla, M., & Shannon, R.V. (2002). Could a lack of experience with a second language be modeled as a hearing loss? *Journal of the Acoustical Society of America,* 112, 2385.

Parrish, W.M. (1951). The concept of naturalness. *Quarterly Journal of Speech*, 37, 448-450.

Peterson, G., & Barney, H. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America,* 24, 175-184.

Pisoni, D.B., & Koen, E. (1981). "Some comparisons of intelligibility of synthetic and natural speech at different speech-to-noise ratios," In *Research on Speech Perception Progress Report No. 7*, Bloomington: Indiana University, Speech Research Laboratory, 243-254.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reinisch, E., & Holt, L.L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 539-555.

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

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Daielsson, H., Lyxell, B., *et al*., (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience,* 7, 31.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *Journal of the Acoustical Society of America,* 106, 3629-3636.

Shannon, R.V. (1989). Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. *Journal of the Acoustical Society of America*, 85, 2587-2592.

Shannon, R.V. (1992). Temporal modulation transfer functions in patients with cochlear implants. *Journal of the Acoustical Society of America*, 91, 2156-2164.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science,* 270, 303-304.

Shannon, R.V., Fu, Q.J., & Galvin III, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-laryngologica Supplementum,* 552, 50-54.

Smith, D.R.R., Patterson, R.D., Turner, R, Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America*, 117, 305-318.

Strange, W., Jenkins, J., & Johnson, T. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, 74, 695-705.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1074-1095.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*. 25, 1-24.

Tamati, T.N., & Pisoni, D.B. (2015). The perception of foreign-accented speech by cochlear implant users, *18th International Congress of Phonetic Sciences*, Glasgow, UK.

Trude, A.M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations of adaptation to foreign accents. *Journal of Memory and Language*, 69, 349-367.

U.S. Census Bureau. (2013). *Language spoken at home for the population 5 years and over*. Retrieved from http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_12_3YR_C16001&prodType=table

Van Engen, K.J., & Peelle, J.E. (2014). "Listening effort and accented speech," *Frontiers in Human Neuroscience*, 8, 577.

Van Wijngaarden, S. J. (2001). The intelligibility of Non-native Dutch speech. *Speech Communication.* 35, 103–113.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, 64, 122-144.

Weatherholtz, K., & Jaeger, T.F. (2016). Speech perception and generalization across talkers and accents. Oxford Research Encyclopedias. DOI: 10.10193/acrefore/9780199384655.013.95.

Winn, M.B., & Litovsky, R.Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *Journal of the Acoustical Society of America,* 137, 1430-1442.

Witteman, M.J., Weber, A., & McQueen, J.M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 75, 537-556.

Witteman, M.J., Weber, A., & McQueen, J.M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin and Review*, 21, 512-519.

Wong, P.C., & Diehl, R.L. (2003). Perceptual normalization for inter- and intratalker variation on Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, 46, 413-421.

Yi, H.G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, 8, 768.

Zhou, N., Xu, L., & Lee, C.Y. (2010). The effects of frequency-place shift on consonant confusion in cochlear implant simulations. *Journal of the Acoustical Society of America*, 128, 401-409.

BIOGRAPHICAL SKETCH

Michelle Rae Kapolowicz moved to Texas from her birth state of Connecticut at a very young age. She studied Philosophy (Major) and Linguistics (Minor) at The University of Texas at Arlington, graduating *Magna Cum Laude* in 2008. She enrolled in the Applied Cognition and Neuroscience Program at The University of Texas at Dallas and received her Master of Science degree in 2010. During this time, she was working in Dr. L. Tres Thompson's Aging and Memory Research Laboratory in a collaborative effort to reformulate chronic micro-drive electrodes for *in vivo* single-unit electrophysiology of hippocampal place cells. A year later, she entered the Cognition and Neuroscience Ph.D. Program at The University of Texas at Dallas, where she continued to work with Dr. Thompson characterizing early plasticity in limbic regions after acute noise exposure in an animal model of tinnitus while cross-training in Dr. Peter Assmann's Speech Perception Laboratory to begin research for her dissertation. Her dissertation work investigates how talker-specific spectral cues aid listeners when perceiving foreign-accented speech.

CURRICULUM VITAE

Michelle R. Kapolowicz

School of Behavioral and Brain Sciences, The University of Texas at Dallas
800 West Campbell Rd., Richardson, TX 75080-3021
Work Email: michelle.kapolowicz@utdallas.edu

Education

2017 **Ph.D.**, *Cognition and Neuroscience*, The University of Texas at Dallas
Advisor: Peter F. Assmann, Ph.D.

2010 **M.S.**, *Applied Cognition and Neuroscience*, The University of Texas at Dallas
Advisor: L. T. "Tres" Thompson, Ph.D.

2008 **B.A.**, *Philosophy* (Major) *Linguistics* (Minor), The University of Texas at Arlington
Magna Cum Laude

Continuing Education & Professional Development

2012 **Plexon Neurophysiology Workshop**
Plexon Inc., Dallas, TX

2010 **Rodent Surgery/Biomethodology Workshops**
The University of Texas at Dallas, Richardson, TX

Peer-Reviewed Publications

MR Kapolowicz & LT Thompson, (*2016*). Acute high-intensity noise induces rapid Arc protein expression but fails to rapidly change GAD expression in amygdala and hippocampus of rats: Effects of treatment with D-cycloserine. *Hearing Research,* 342, 69-79.

MR Kapolowicz, V Montazeri, & PF Assmann, (*2016*). The role of spectral resolution in foreign-accented speech perception. *Proceedings of Annual Conference of the International Speech Communication Association,* INTERSPEECH 2016, 3289-3293.

MR Kapolowicz, V Montazeri, & PF Assmann, (*submitted*). Perceiving foreign-accented speech with decreased spectral resolution in single- and multiple-talker conditions. *Journal of the Acoustical Society of America: Express Letters.*

Conference Presentations

MR Kapolowicz, DR Guest, V Montazeri, & PF Assmann, *Effect of frequency shifts on talker recognition in native- and foreign-accented speech*. 174th Meeting of the Acoustical Society of American, December, 2017.

DR Guest, **MR Kapolowicz,** V Montazeri, & PF Assmann, *Perception of voice gender in children's voices by cochlear implant users*. Boston, MA: 173rd Meeting of the Acoustical Society of America and the 8th Forum Acousticum, June, 2017.

MR Kapolowicz, V Montazeri, JF Kuang, & PF Assmann, *Adaptation to foreign-accented speech with decreased spectral resolution*. Honolulu, HI: 5th Joint Meeting of The Acoustical Society of America and the Acoustical Society of Japan, December, 2016.

DR Guest, **MR Kapolowicz**, S Hossain, V Montazeri, & PF Assmann*, Perception of voice gender in cochlear implant simulations of children's speech.* Salt Lake City, UT: 171st Meeting of the Acoustical Society of America, May, 2016.

PF Assmann, **MR Kapolowicz**, DA Massey, S Barreda, & TM Nearey, *Links between the perception of speaker age and sex in children's voices*. Jacksonville, FL: 170th Meeting of the Acoustical Society of America, November, 2015.

MR Kapolowicz, PF Assmann, & LT Thompson, *Tinnitus-inducing noise trauma and D-cycloserine alter Arc protein expression in amygdalo-hippocampal circuitry.* Baltimore, MD: Annual Midwinter Meeting of the Association for Research in Otolaryngology, February, 2015.

MR Kapolowicz, JI Sedillo, MM Makkieh, & LT Thompson, *Tinnitus-inducing noise trauma and D-cycloserine alter amygdalo-hippocampal excitatory biomarkers*. Washington, DC: Annual Meeting of the Society for Neuroscience, November 2014.

JI Sedillo, **MR Kapolowicz**, MM Makkieh, AR Møller, & LT Thompson, *Evidence for multisystem plasticity in non-classical auditory regions in early stages of tinnitus*. Washington, DC: Annual Meeting of the Society for Neuroscience, November, 2014.

PF Assmann, **MR Kapolowicz**, DA Massey, S Barreda, & TM Nearey, *Perception of speaker sex in re-synthesized children's voices*. Providence, RI: 167th Meeting of the Acoustical Society of America, May, 2014.

MR Kapolowicz, EH Kardosi, FK Alahmady, JI Sedillo, & LT Thompson, *The effect of noise trauma on Arc and GAD expression in a rat model of tinnitus.* San Diego, CA: Annual Meeting of the Society for Neuroscience, November, 2013.

CE Barnes, **MR Kapolowicz**, & LT Thompson, *Using a gap detection startle paradigm to assess gender differences in the development of tinnitus*. Dallas, TX: UT Dallas Undergraduate Research Scholar Award Presentation, March 2013.

MR Kapolowicz & LT Thompson, *Amygdalo-hippocampal plasticity in a rat model of tinnitus: implications for aging and Alzheimer's disease.* Las Colinas, TX: Dallas Aging and Cognition Conference, January 2013.

MR Kapolowicz, SB Templet, S Yan, & LT Thompson, *Hippocampal place cells and interneurons in tinnitus: Time course of behavioral and single-unit plasticity*. New Orleans, LA: Annual Meeting of the Society for Neuroscience, October 2012.

MR Kapolowicz, SJ Lek, S Templet, R Rennaker, & LT Thompson, *Hippocampal place-cell plasticity and basolateral amygdala responses to auditory stimuli in a rat model of tinnitus*. Washington, DC: Annual Meeting of the Society for Neuroscience, November 2011.

Invited Speaker

Children's Voices, Foreign Accent, and the Perception of Speech in Adverse Conditions. Callier Center for Communication Disorders, Dallas, TX. (September 29, 2017).

Teaching

Guest Lecturer

Language, Ethology, Piaget and Development. The University of Texas at Dallas, Richardson, TX: Invited Guest Lecturer for PSY 3360/CGS 3325: Historical Perspectives on Psychology: Minds and Machines since 1600 (August 4, 2016).

Acute high-intensity noise induces rapid Arc/Arg 3.1 expression but fails to change GAD expression: region-specific effects of treatment with D-cycloserine. The University of Texas at Dallas, Richardson, TX: Invited Guest Lecturer for NSC 4357: Neurobiology of Learning and Memory (March 21, 2016).

Hearing and Language. The University of Texas at Dallas, Richardson, TX: Invited Guest Lecturer for NSC 3361: Behavioral Neuroscience (March 3, 2016)
Sleep. The University of Texas at Dallas, Richardson, TX: Invited Guest Lecturer for NSC 3361: Behavioral Neuroscience (November 13, 2014).

The effects of d-cycloserine on ARC and GAD expression in a rat model of tinnitus. Brookhaven College, Farmers Branch, TX: Invited Guest Lecturer for PSYC 2301: General Psychology (September 7, 2012).

Teaching Assistant

Introduction to Neuroscience (Fall 2016, Spring 2017, Fall 2017)

Redesigning Behavioral and Brain Sciences Freshman Experience Course (Summer 2015)

Historical Perspectives (Summer 2014, Summer 2017)

Behavioral Neuroscience (Summer 2013, Fall 2014, Spring 2015, Fall 2015, Spring 2016)

Brain and Memory (Spring 2013)

University Service Activities

Graduate Student mentor for UT Dallas Undergraduate Honors Thesis for Tina Dam: Comparing effects of non-traumatic and traumatic noise exposure on Arc protein expression in limbic regions (2015)

Graduate Student mentor for UT Dallas Undergraduate Honors Thesis for Claire Barnes: Using a gap-startle paradigm to assess gender differences in the development of tinnitus (2013)

Assistant to Cognition and Neuroscience Program Head (Fall 2013, Spring 2014)

Community Service Activities

City of Plano: Animal Services: Foster Volunteer
Plano, TX (May 31, 2016 – June, 9, 2017)

Special Awards Judge for the Dallas Regional Science and Engineering Fair
Area: Acoustics, Representing the Acoustical Society of America
Dallas, TX (February 27, 2016, February 25, 2017)

Funding

University of Texas at Dallas $1000.00 Small Research Grant (May, 2017)

School of Behavioral and Brain Sciences' $2000.00 Research Stipend for Women in Bio-Behavioral Sciences (Fall 2012)

University of Texas at Dallas $900.00 Graduate Student Grant to attend Plexon's Neurophysiology Workshop (April 2012)

University of Texas at Dallas $1000.00 Graduate Student Travel Award (November 2011, October 2012, November 2013, November 2014, September 2016, December 2017)

Professional Affiliations

International Speech Communication Association*, member* (June 2016 – present)
Acoustical Society of America, *member* (September 2013 – present)
Society for Neuroscience, *member* (May 2011 – present)