NORTHWESTERN UNIVERSITY

# Perceptual Learning of Accented Speech by First and Second Language Listeners

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By
Angela Cooper

EVANSTON, ILLINOIS

June 2016

## ABSTRACT

Acoustic variability permeates our communicative environments, arising from differences in talker or accent, speaking style, and a range of environmental factors, which can make the mapping between the auditory input and stored linguistic representations a challenge. However, listeners have been shown to be remarkably flexible in their ability to adapt to these changing conditions. The aim of this dissertation is to better understand the processes that underlie perceptual adaptation to speech variation, with a particular interest in the role of predictive strength and uncertainty during adaptation. The current experiments examined how the perceptual system utilizes different types of linguistic knowledge and signal-based information during adaptation and how linguistic experience modulates this process. The present work hypothesized that the predictive strength of disambiguating information (that is, the degree to which it narrows the space of possible category options for the incoming input) would mediate adaptation, with more predictive information increasing listeners' certainty about how to categorize the input, thereby facilitating adaptation, relative to less predictive information.

Chapter 2 investigated this issue by presenting listeners with Mandarin-accented English sentences followed by native-accented feedback that either aligned with the target at 1) all linguistic levels, 2) syntactic and sub-lexical levels with real words or 3) syntactic and sub-lexical levels with non-words. Contrary to our initial prediction that adaptation performance would vary as function of predictive strength, listeners provided with any kind of native English-accented feedback (matched or mismatched sentences) significantly outperformed those not presented with this kind of feedback, suggesting that listeners drew

upon points of alignment between the native- and Mandarin-accented sentences beyond the lexical. These findings indicate that native English listeners can leverage any linguistically relevant information during adaptation, even if it is not highly predictive of the input.

Chapters 3 and 4 further explored the issue of predictive strength by comparing the contribution and interaction of different types of information, namely linguistic knowledge (e.g., lexical vs. semantic contextual feedback) and signal-based information (single vs. multiple talker training) during passive exposure to a novel English accent (NSAE). The impact of uncertainty during adaptation was also explored by comparing first (L1) and second (L2) language listeners, who vary in the amount of linguistic uncertainty they maintain (about the language overall but also regarding specific phonemic contrasts). The results revealed that both L1 and L2 listeners were capable of drawing upon varied types of information (regardless of their predictive strength) in order to enhance their certainty and make them willing to adjust the relevant categories, as evidenced by significantly improved recognition of NSAE-accented pronunciations following exposure. Moreover, linguistic experience was found to play a dynamic role in adaptation, demonstrating a combination of plasticity and stability within the system. When exposed to accent patterns involving contrasts that also exist in their native language, L2 listeners demonstrated significant learning; however, their perceptual system appeared to be resistant to adaptation when encountering items containing challenging L2 contrasts. These results point to perceptual adaptation involving a dynamic interaction of prior knowledge and uncertainty with current beliefs about the observed auditory input.

## ACKNOWLEDGMENTS

First and foremost, I would like to extend my heartfelt thanks to Ann R. Bradlow, my advisor and mentor, for her unwavering support and encouragement throughout my graduate career. I am grateful for her engaging and insightful discussions, her patient and detailed critiques, and for her confidence in my research that has enabled me to move forward with my work more securely. It has been an absolute pleasure learning from and collaborating with her—an honour I hope will continue in the future. Thanks also to my other committee members, Matt Goldrick and Nina Kraus, for their helpful comments and encouraging me to push myself to think about the research in different ways.

I must also thank the funding agencies and grants that have supported my graduate career, including the Social Sciences and Humanities Research Council of Canada for their doctoral fellowship (#752-2011-0082), The Graduate School for their Graduate Research Grant, Ann Bradlow's NIH-NIDCD grant (R01-DC005794), and the numerous travel grants awarded to me by the Linguistics department, the Cognitive Science program and The Graduate School.

My fellow grad students and friends, Angela Fink and Erin Gustafson, deserve my gratitude for their valuable advice on research and statistics, being fabulous travel buddies and for all of our ladies nights filled with tasty foods and good company. Thank you especially to David Potter, who, throughout everything, all of the trials and tribulations, has remained a constant source of support—my best friend who always inspires me with his passion and enthusiasm. His encouragement at every step of my graduate career has been invaluable.

Thanks also to Chun Liang Chan and my research assistant Alexandra Saldan for all of their research and technical support. I am also grateful to Mirjam Ernestus and her research group at the Max Planck Institute for Psycholinguistics in Nijmegen for being so welcoming and supportive, allowing me to collect the data from my second language speakers efficiently and effectively.

I have also been extremely fortunate to have a wonderfully supportive network of friends outside of the department, particularly through my involvement with Glenwood Dance Studio. I am so grateful for the mental reprieve and respite from the toils of graduate life that dance has provided me. The studio has been a safe haven for me, filled with dancing, laughter and silliness, and I am especially thankful to Annaleah Tubbin and Rose Mulvey for their enduring friendship. I owe a particular debt of gratitude to Erin DeWitt, who has been tremendously supportive throughout this challenging dissertation year, lending a patient ear and always willing to be there for me.

Finally, I must express my sincerest thanks to my family, who taught me from an early age the value of pursuing knowledge. They have been champions of my success from the beginning, always believing that I could accomplish anything.

*"The accent of one's birthplace remains in the mind and in the heart as in one's speech."*

*- Francois de La Rochefoucauld*

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1: INTRODUCTION

A critical step to understanding speech, spoken word recognition, is contingent on our ability to map the incoming auditory input onto stored linguistic representations. Discrete phonemes must be extracted from the continuous acoustic information, which must then be grouped into individual word forms. Multiple candidate word forms that either partially or completely overlap with the input become activated in the lexicon, at which point, the listener must select the most likely lexical candidate, given the sensory input as well as other linguistic and nonlinguistic contextual information (Dahan & Magnuson, 2006). Rampant acoustic variability in our communicative environments, stemming from talker or accent characteristics, speaking rate, and environmental noise, makes this mapping an even more complicated process. Despite this, native language listeners display remarkably flexible perceptual systems, able to efficiently accommodate this variability and accurately recognize varied pronunciations of spoken words (Cutler, 2012).

## 1. Perceptual learning

Successful speech perception arises in part as a result of the perceptual system's capacity to adapt how it responds to sensory input, integrating current information in the environment with prior experience and stored linguistic knowledge. This perceptual learning process is evident from the more accurate and efficient processing of speech produced by familiar relative to unfamiliar talkers (e.g., Newman & Evers, 2007; Nygaard & Pisoni, 1998) and familiar versus unfamiliar accents (e.g., Adank, Evans, Stuart-Smith, & Scott, 2009; Smith, Holmes-Elliott, Pettinato, & Knight, 2014; Witteman, Weber, & McQueen,

2013). Indeed, short- or long-term exposure to a given talker or accent facilitates spoken word recognition, as it enables the system to cue into the specific acoustic patterns that characterize their speech and internalize this knowledge for future use.

The nature of these adaptation processes has been a subject of considerable research over the past several decades, with investigations into the representational locus and specificity of these perceptual adjustments, the types of internal knowledge and external cues that facilitate adaptation and the conditions under which adaptation will actually take place (e.g., Bertelson, Vroomen, & De Gelder, 2003; Bradlow & Bent, 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2006; Maye, Aslin, & Tanenhaus, 2008; McQueen, Cutler, & Norris, 2006; Norris, McQueen, & Cutler, 2003; Sjerps & McQueen, 2010). In their seminal paper, Norris et al. (2003) investigated adaptation to within-category variation by presenting listeners with an ambiguous sound [#] (midway between [f] and [s]) in lexical contexts intended to bias them to perceive the ambiguous sound as either [f] or [s] (e.g., "gira[#]e" or "dino[#]aur", respectively). Participants' interpretation of the ambiguous sound during a lexical decision task later impacted a subsequent phonetic categorization task, where listeners were asked to identify items along an auditory [f]-[s] continuum. Listeners' category boundaries were found to have shifted after exposure to the ambiguous fricative, with those who had heard the [#] in [f]-biasing lexical contexts producing more [f] categorizations than those who heard the [#] in [s]-biasing contexts, and those in the [s]-biased group showing the reverse pattern. Such phonetic category recalibration was not found for listeners who were exposed to the ambiguous fricative in nonword contexts, indicating that lexical knowledge was necessary to disambiguate [#] and for perceptual learning to take place. Learning was

also remarkably rapid, occurring after exposure to only 20 items (spread out throughout 200 other words and nonwords).

This lexically-guided category retuning paradigm has been utilized in a number of subsequent studies that have provided a more detailed window into the mechanisms that drive perceptual learning. The perceptual adjustments induced by lexically-disambiguating exposure have been found to generalize to novel lexical items (McQueen et al., 2006), pointing to a pre-lexical locus of adjustment. That is, in order for listeners to interpret an atypical or ambiguous sound in a previously unheard word, the boundaries of the relevant phonetic categories would have had to have been adjusted (rather than encoding this information into specific lexical items). Furthermore, these category re-adjustments have been shown to endure for at least 25 minutes, even when exposed to prototypical pronunciations of the ambiguous sound in the interim (Kraljic & Samuel, 2005) and even up to 12 hours (Eisner & McQueen, 2006).

The specificity of perceptual learning has also been the subject of investigation (e.g., Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Kraljic & Samuel, 2006), with a number of factors implicated in whether or not listeners will generalize to novel talkers, including the type of phonetic contrast involved, the amount of relevant variation in the signal and the acoustic similarity of the talkers. Eisner and McQueen (2005) reported talker-specific perceptual learning, finding that talker-general phonetic category recalibration only occurred if the vowel [ɛ] used in a continuum of [ɛs]-[ɛf] was produced by a novel talker but the fricatives were produced by the original talker. If the entire continuum was produced by the novel talker, no evidence of learning was obtained. It is important to note that prior

work finding talker-specific learning (Eisner & McQueen, 2005; Kraljic & Samuel, 2005) used fricatives as the critical ambiguous sounds. Kraljic and Samuel (2006), using the same lexically-guided retuning paradigm, found significant learning for both the trained and a novel talker when exposed to an ambiguous stop consonant (midway between /d/ and /t/). Furthermore, single-talker learning not only generalized to a new talker, but it also generalized to a novel set of featurally-related consonants (/b/ and /p/), suggesting that perceptual adjustments are made at a sub-phonemic level. The disparity in the specificity of learning between fricatives and stop consonants has been attributed to the fact that stop consonants provide relatively less indexical information that would cue a speaker's identity (being a temporal contrast) relative to fricatives (being a spectrally-cued contrast).

The specificity of perceptual adjustments can also depend on the variation provided during exposure. When presented with multiple talkers who share the same foreign accent, perceptual learning was more likely to generalize to a novel talker that shares that accent than if exposed to a single talker (Bradlow & Bent, 2008). And, when presented with multiple talkers with different foreign accents, perceptual learning was more likely to generalize to a novel foreign accent than if exposed to a single foreign accent (Baese-Berk et al., 2013). High variability training is posited to promote generalizable learning, as it allows for the extraction of systematic patterns of deviations from native-accented norms shared by the talkers, resulting in more robust and generalizable adjustments. There have been cases where single talker training on a regional accent characteristic (e.g., a segment ambiguous between two categories) has been found to transfer to a novel talker, which has been attributed to the acoustic similarity between the talkers (Reinisch & Holt, 2014). Cross-talker generalization

of lexically-guided phonetic category retuning for [s]-[f] was found to be dependent on whether the fricatives produced by the novel talker was sampled from the same or a distinct perceptual space as the trained talker.

Recent work has also shown that category recalibration can occur not only for a single pair of categories but also for a set of categories (Maye et al., 2008; Weatherholtz, 2015). Understanding how listeners adapt to multiple patterns of deviation is particularly important, in large part because most foreign and regional accents involve more than one deviation pattern, and the factors that contribute to tracking multiple patterns relative to a single pattern may differ. Moreover, a language is structured such that it utilizes a relatively small set of features to distinguish many different contrasts or systems of contrasts. Foreign-accented speech often systematically deviates from native-accented speech as a product of talkers' inability to appropriately utilize certain features. Multiple accent patterns within a foreign accent may therefore share the use of common underlying features, which could contribute to listeners' patterns of adjustments and generalization during adaptation. Maye et al. (2008) exposed listeners to a novel, vowel chain-shifted accent of English where front vowels were lowered (e.g., "keep" /kip/ ➔ "kip" /kɪp/; "witch" /wɪtʃ/ ➔ "wetch" /wɛtʃ/, "rest" /ɹɛst/ ➔ "rast" /ɹæst/). Listeners first heard a story passage (*The Wizard of Oz*) produced by a synthesized voice in a standard American English accent and then completed a lexical decision task. In a second session, listeners heard the same story passage produced by the same synthesized voice, which now contained altered pronunciations of the critical front vowel items, before completing the identical lexical decision task. Their findings revealed evidence for perceptual learning, as shown by a significant increase between the first and

second sessions in lexical endorsement rates for front vowel lowered items. That is, words that were originally perceived to be nonwords (e.g., “wetch”) were more likely to be perceived as words following exposure to the novel accent. However, listeners apparently did not completely re-map their vowel space, as listeners still considered standard pronunciations of the items to be words (both “witch” and “wetch” were accepted as forms of “witch”). Learning was also found to be direction-specific, with no increase in endorsement rates to items with front raised vowels. The authors interpreted these findings as indicating that perceptual learning does not involve a general relaxing of criteria for what is an acceptable exemplar of a vowel category but rather constitutes targeted category shifts. Additionally, in their first experiment, the authors reported evidence that learning generalized to an untrained region of the vowel space (back vowel lowered items), though they were unable to replicate this finding in their second experiment.

Weatherholtz (2015) further pursued this issue, noting that Maye et al. (2008)’s use of synthesized speech and a within-subject design (where the same items were repeated in both lexical decision tasks) might not have resulted in an accurate reflection of how cross-category variation (e.g., *wooden* pronounced as *w*[o]*den*, where one category is substituted for a different category) is adapted to in naturally-produced speech. Employing a between-subjects exposure-test design, listeners were exposed either to a story containing a novel vowel chain shift (back vowel lowering) or standard American English pronunciations. At test, novel accent-exposed listeners demonstrated significant learning, with higher endorsement rates and word recognition of novel, back vowel lowered items. Despite being exposed to only a single talker, learning was also found to generalize to new talkers with the

same accent, regardless of their acoustic similarity to the trained talker. Moreover, when exposed to a back vowel lowered chain shift, listeners generalized learning to back vowel raised and front vowel lowered items. Though, listeners exposed to back vowel raised variants did not display generalization to untrained back vowel lowered items. The author concluded that perceptual adaptation involves dynamic adjustments to talker-independent representations, which can include either a broadening or targeted shifts of phonetic categories. The apparent discrepancy between Maye et al. (2008) and Weatherholtz (2015) with respect to the amount of generalization demonstrated might be in part dependent on the directionality and location of the vowel shift, as even Weatherholtz (2015) found varying degrees of generalization depending on whether the vowel shift was raised or lowered.

While prior research has predominantly focused on lexically-guided learning, the perceptual system has been found to leverage different non-lexical dimensions during adaptation (Bertelson et al., 2003; Cutler, McQueen, Butterfield, Norris, & Planck, 2008; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008; Mitterer, Chen, & Zhou, 2011). Cutler et al. (2008), in a similar paradigm as Norris et al. (2003), found that listeners utilized native language phonotactics to disambiguate ambiguous sounds. Exposure consisted of nonword contexts containing an ambiguous sound between [f]-[s], where only one of the sounds was phonotactically legal (*frul* vs. **srul* or **fnud* vs. *snud* ). A similar boundary adjustment was yielded as in prior studies with lexically-guided learning, indicating that lexical information is not required for learning to take place. Other types of linguistic information, including phonological (Hervais-Adelman et al., 2008) as well as semantic contextual (Mitterer et al., 2011), have similarly been found to drive perceptual adjustments.

Non-auditory cues, such as visual lip movements, can also be used in the recalibration of phonetic categories. Bertelson et al. (2003) presented a sound ambiguous between [b]-[d] in conjunction with a video of a face articulating either [b] or [d]. In a subsequent phonetic categorization task (with no video presentation), listeners in the visually [b]-biasing group produced a higher proportion of [b] responses relative to the visually [d]-biasing group. Taken together, these findings indicate that perceptual learning processes will leverage contextual information, lexical or otherwise, that serves to sufficiently constrain how to categorize the incoming input.

While these findings demonstrate the perceptual system's remarkable plasticity in the face of abundant variation, the system must restrict the re-adjustment of category boundaries to a certain extent in order to provide a degree of stability to sustain existing categories and representations. Samuel and Kraljic (2009) posited that the system maintains a balance between stability and plasticity by only recalibrating categories when it has sufficient evidence to do so ('Conservative Adjustment & Restructuring Principle'), such as if there is an indication that the pronunciation is a stable property of the speaker (or group of speakers). Kraljic, Samuel, and Brennan (2008) assert that listeners are subject to a primacy bias, such that they learn and adapt to their initial experiences with the speaker, which inform what they consider to be characteristic of that speaker. Indeed, listeners who were first exposed to a speaker's nonstandard pronunciations adapted their categories accordingly; however, when the same nonstandard pronunciations were presented after exposure to the speaker producing standard pronunciations, no such learning occurred. Furthermore, if the nonstandard pronunciations can be attributed to some external factor, such as a pen in the speaker's

mouth, the system remains stable, as such pronunciations are deemed to be incidental and not a lasting property of the speaker's productions.

Taken together, the perceptual system fine-tunes its performance in acoustically-varied contexts by being "aggressively opportunistic" (p. 1217, Samuel & Kraljic, 2009), drawing upon any relevant information, both internally-generated (e.g., phonotactic or lexical knowledge) and externally-provided (e.g., visual cues), to recalibrate category boundaries to enable more efficient and accurate perception in the future. The degree to which the system is willing to generalize its adaptation (or adapt at all) has been found to be dependent on whether it can determine if the pronunciation variants are a stable property of the talker or group of talkers, which can be modulated by the acoustic similarity between exposure and novel talkers, variability provided during exposure, and the perceived reliability of the variant (that is, whether its atypicality can be attributed to talker-external factors).

**2. Frameworks of perceptual learning**

Given the wealth of behavioural evidence of the flexibility inherent in successful speech perception, numerous models have been proposed (or adapted from existing speech perception models) to account for perceptual adaptation to speech variation (e.g., Guediche, Blumstein, Fiez, & Holt, 2014; Kleinschmidt & Jaeger, 2015; Mirman, McClelland, & Holt, 2006; Norris et al., 2003; Sohoglu & Davis, 2016). Such models include the feedforward MERGE model (Norris et al., 2003) and the interactive Hebb-TRACE model, which employs a Hebbian learning algorithm to adjust connection weights between the auditory input and pre-lexical representations (Mirman et al., 2006). Guediche et al. (2014; see Panichello, Cheung, & Bar, 2013, for a similar perspective in the visual domain) proposed that predictive

coding might be a valuable framework to unify the different strands of research on adaptation. The authors noted that different types of disambiguating information that have been examined, including internal (e.g., linguistic knowledge) and external (e.g., feedback) sources, could be used by the perceptual system to generate predictions. During speech perception, the incoming auditory input is compared against these predictions, and any discrepancies between them will yield an internally-generated error signal. This error signal will then lead to adjusted predictions in an effort to improve the match between these predictions and the sensory input. For example, when encountering an ambiguous sound [#] between [f]-[s] in the context of [ɹi#], listeners' lexical knowledge generates a prediction that the item should be "reef", as "reese" is not a real word in English. However, in comparing the auditory information to the stored phonemic representations associated with that item, the discrepancy between [#] and [f] generates a prediction error signal. As a result, adaptive adjustments are made such that future predictions would include [#] as a possible exemplar of [f], improving the alignment between predictions and subsequent input. When they later encounter [#] in items such as [li#], they would then be more likely to interpret it as "leaf" rather than "lease", as a product of this category boundary adjustment.

Recent work has taken this a step further by providing a formally explicit framework to model perceptual adaptation (the ideal adapter framework, Kleinschmidt & Jaeger, 2015), whereby speech perception involves a combination of "prediction and inference under uncertainty" (p. 76). The authors posit that for speech perception to occur, listeners build generative linguistic models, which can be defined as knowledge about the distribution of acoustic cues associated with each linguistic unit (e.g., a phonetic category). Similar to

Guediche et al. (2014), listeners utilize knowledge of higher-level linguistic units, comparing them to determine how well each one predicts the incoming signal. Because of the variability inherent in speech perception (e.g., talker or accent-related differences), accurate perception also relies on utilizing the contextually-appropriate generative model. However, because listeners cannot ever truly know the exact nature of the generative model for any given talker or situation, they maintain uncertain beliefs about it. Thus, adaptation is a process where listeners update their beliefs about the cue distributions of a talker- or situation-specific generative model. If there is insufficient information available about the relevant generative model (e.g., a novel male English talker), listeners will leverage a combination of prior beliefs (e.g., their experience with other male English speakers), and whatever they can extract from current observations.

If we consider speech perception as a problem of inference under uncertainty, then it is necessary to note that uncertainty can stem from multiple sources, from uncertainty about the identity of the talker or the accent to uncertainty about what prior experience will be relevant to determining the nature of the generative model (e.g., whether it is appropriate to utilize beliefs associated with a generative model of Canadian-accented or Boston-accented or Mandarin-accented English). Listeners are posited to infer what is relevant based on a combination of top-down cues, such as visual information about the speaker or being explicitly told the speaker's origins, and bottom-up cues from the speech signal.

**3. Current research**

The principal aim of this dissertation was to provide a better understanding of the processes that drive perceptual adaptation to variation in speech. In particular, the present

research sought to understand how the perceptual system leverages different types of information during adaptation, varying in the degree to which they predict the relevant phonemic categories, and how linguistic experience mediates this process. These issues can be conceptualized from the perspective of prediction and uncertainty (Guediche et al., 2014; Kleinschmidt & Jaeger, 2015). Listeners are posited to maintain beliefs about the relevant cue distributions of linguistic units over the course of speech perception (Kleinschmidt & Jaeger, 2015), making predictions based on disambiguating information (internally-generated or externally-provided) about how well their beliefs align with the observed sensory input. Given that, all levels of linguistic knowledge (e.g., phonemic, lexical, syntactic, pragmatic) should be available for the listener to generate these predictions. Prior work has predominantly focused on the utility of lexical information during perceptual learning (e.g., Norris et al., 2003; Maye et al., 2008); however, less is known about the relative contribution of different types of lexical and other higher-levels of linguistic information (e.g., semantic context) and lower-level information (e.g., phonetic information) in learning. The present work hypothesizes that the predictive strength of the disambiguating information, that is the degree to which the information narrows the space of possible category options, modulates the speed and generalizability of learning. Predictive strength may also contribute to listeners' level of uncertainty. For example, being provided a subtitle that exactly matches the speech being presented is 100% predictive. As such, listeners can have high certainty as to the identity of the appropriate categories that need to be adjusted. A pragmatic context sentence, however, would be relatively less predictive and should thus increase the listeners' uncertainty about which specific adjustments need to be made, which might subsequently

slow adaptation. Moreover, in natural communicative contexts, listeners likely draw upon multiple sources of information during adaptation (e.g., visual cues, linguistic information, talker variability). How then do these sources of information interact with one another? It could be the case that certain sources may be negatively impacted by other types of information. For example, Clopper (2012) found an attenuated semantic predictability effect in speech-in-noise perception as a product of dialect variability. It had been posited that as energetic masking increases in speech-in-noise contexts, listeners weight higher-level cues such as semantic context lower and attend more to lower-level acoustic information (Mattys, Brooks, & Cooke, 2009). Clopper (2012) asserted that talker and dialect variability functions as a source of noise, in much the same way that speech-shaped noise yields varying degrees of energetic masking, and that increasing variability leads to a concomitant increase in "noise", thus reducing the semantic predictability benefit.

In addition to the uncertainty that may arise from the nature of the disambiguating information, the present work argues that uncertainty can also stem from listeners' linguistic experience. Second language (L2) speakers, by virtue of having less exposure to and experience with the language, will have a more impoverished L2 linguistic knowledge base, which can have a range of different effects on second language processing (Cutler, 2012). Native listeners have a much broader base of experience with speakers of English, yielding prior beliefs about English that are more generalized (or less constraining) than L2 listeners, who have more constrained prior beliefs due to their relatively smaller set of encounters. This could result in a generally heightened uncertainty level for L2 listeners—uncertainty about the appropriate cue distributions for L2 linguistic units as well as uncertainty about the

applicability of their beliefs derived from their more limited experience with English speakers.

While prediction and uncertainty are proposed to be at work in adaptation to multiple forms of variation yielding ambiguity (e.g. environmental noise, vocoded speech), the present work focuses on their role in adaptation to foreign-accented speech specifically. Accented speech produces variation along multiple linguistic dimensions, thereby providing an excellent testing ground for the influence of different levels of linguistic information as well as listeners' prior linguistic knowledge (being a first or second language user) on perceptual learning processes. Moreover, with English being spoken as a second language by over 500 million people worldwide (Lewis, Simons & Fennig, 2014), the growing need for both first and second language speakers to accommodate accented speech is uncontroversial. Increasing our understanding of the factors that mediate more or less successful adaptation to accented speech has practical ramifications that could contribute to the development of training paradigms for listeners with extensive contact with accented speakers.

This dissertation sought to provide insight into how sources of prediction and uncertainty modulate perceptual learning processes with three experiments. Chapter 2 describes an experiment examining the relative contributions of different levels of linguistic information, varying in their predictive strength, in the adaptation to Mandarin-accented English sentences. The use of naturally-produced, sentence-length materials more closely resembles the kind of speech samples listeners might encounter in real-world communicative contexts with accented speakers. Participants first transcribed Mandarin-accented English sentences in speech-shaped noise. Feedback sentences by a native talker were then presented

that either aligned with the target at 1) all linguistic levels, 2) syntactic and sub-lexical levels with real words or 3) syntactic and sub-lexical levels with non-words and were compared against conditions providing non-English "feedback" or accent only exposure.

Chapter 3 examined how different types of information (e.g., linguistic knowledge vs. signal-based cues such as talker variability) interacted with one another by exposing listeners to a novel accent of English, specifically constructed so as to control for the number and type of accent deviation patterns. In this experiment, the predictive strength was manipulated along various dimensions: the type of feedback provided to disambiguate the accented items (lexical, semantic context) and the number of talkers during exposure (single, multiple). Following exposure, listeners were tested with lexical decision and word identification tasks and compared with listeners who had not received accented exposure.

Finally, Chapter 4 investigated the issue of linguistic experience and its impact on perceptual adaptation processes. Dutch-English bilinguals were tested with the same paradigm introduced in Chapter 3 to test whether their general uncertainty about the language as well as more specific uncertainty about the distributions of certain phonemic contrasts would slow learning relative to native English listeners.

This dissertation document is comprised of 5 chapters. This introductory chapter provides a general overview of the prior literature and of the research questions investigated in the dissertation. Chapters 2 – 4 are written as separate research papers and, as such, each has their own introduction, methods, results and discussion sections. Chapter 5 is a concluding chapter that summarizes the overall findings, discusses the potential implications and future directions of this research.

# CHAPTER 2: LINGUISTICALLY-GUIDED ADAPTATION TO FOREIGN-ACCENTED SPEECH

## 1. Introduction

With English being spoken as a second language by over 500 million people worldwide (Lewis, Simons & Fennig, 2014), the growing need to communicate across a language barrier (i.e., between native and non-native speakers) is uncontroversial. One could address the potential problem of communication across a language barrier in one of two ways: 1) training non-native speakers to modify their foreign accents, or 2) training listeners to be more flexible in their perceptual accommodation to foreign accents. Training non-native speakers to achieve native-like production targets is a well-documented challenge, sometimes requiring weeks of training to yield improvements on only a single segmental contrast (e.g., Hirata, 2004; Thomson & Derwing, 2015). On the other hand, in most communicative contexts, listeners must contend with extensive talker-related variability in their speech input, ranging from differences in vocal tract size to the presence of a speech impairment or foreign accent. As such, listeners are highly experienced at adapting to systematic deviations from long-term linguistic regularities such as those present in foreign-accented speech—adaptation which can occur in as short a time frame as a few sentences (e.g., Clarke & Garrett, 2004) and which can have lasting effects on listeners' perceptual systems (Zhang & Samuel, 2014). The present study thus focuses on elucidating this perceptual learning process as a linguistically-guided process involving multiple levels of linguistic information.

Perceptual learning for speech has been found to leverage information outside of the speech signal itself, including visual lip movements (e.g., Bertelson, Vroomen, & De Gelder, 2003) and lexical knowledge (e.g., Maye, Aslin, & Tanenhaus, 2008), to facilitate the disambiguation of distorted or ambiguous speech. Such visually- or lexically-guided disambiguations are posited to yield adaptive adjustments to phonetic categories, resulting in improved classification of subsequent exposures to the ambiguous sound in novel words. A considerable body of research has investigated the specific role of lexical information in adapting to variability (e.g., Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Kraljic & Samuel, 2007; Maye et al., 2008; Mitterer & McQueen, 2009; Norris, McQueen, & Cutler, 2003). For instance, Norris et al. (2003) presented listeners with an ambiguous sound between [f] and [s] in lexical contexts intended to bias them to perceive the ambiguous sound as either [f] or [s] (e.g., "giraffe" or "dinosaur", respectively). When later asked to identify sounds along an auditory [f]-[s] continuum, listeners' phonetic category boundaries were found to have shifted after exposure, suggesting that listeners' experience with the ambiguous sound in lexically-biasing contexts enabled them to recalibrate the appropriate phonetic categories.

Similarly, rapid-adaptation to noise-vocoded speech (i.e., speech that has been filtered in a manner that mimics the course-grained filtering of a cochlear implant) has been found when listeners are provided feedback as to the lexical content of the speech (Davis et al., 2005). Listeners were first asked to transcribe a vocoded sentence, followed by two repetitions of the target sentence. One group received the same sentence in clear (non-vocoded) speech followed by vocoded speech (distorted-clear-distorted condition, DCD), and

the other group received the same sentence as vocoded speech followed by clear speech (distorted-distorted-clear condition, DDC). The DCD condition allowed listeners to hear a repetition of the vocoded sentence after the lexical content of the sentence had been revealed (in the clear presentation), enhancing the intelligibility of that vocoded sentence, a phenomenon that has been termed a "pop-out" effect. This enabled listeners in the DCD condition, over the course of 30 trials, to adapt to the noise-vocoded speech significantly faster than the DDC group.

One way to conceptualize the efficacy of such lexically-guided learning is to consider that it provides information to the perceptual learning system that is used to internally-generate predictions about future speech input (Guediche, Blumstein, Fiez, & Holt, 2014; see Panichello, Cheung, & Bar, 2013, for a similar perspective in the visual domain). Any discrepancy between these predictions and the incoming speech input would yield an error signal, which would subsequently initiate adaptive adjustments by the perceptual system in an effort to improve the match between future predictions and the speech input. Under this perspective, different levels of linguistic information (e.g., phonological, lexical, semantic or pragmatic context) in the feedback interval should all be available to help generate predictions for upcoming speech input that should in turn facilitate perceptual adjustments. Foreign-accented speech yields variation along multiple linguistic dimensions, ranging from the level of individual speech sounds to word-level stress patterns to phrase-level intonational contours, thereby providing an excellent testing ground for the influence of different levels of linguistic information on perceptual learning processes.

While there has been a considerable focus in prior research on lexical information as a source of disambiguating information, relatively little work has investigated the efficacy of other possible connections between the incoming speech input and different levels of linguistic representation (Cutler et al., 2008). In Davis et al. (2005), clear (non-vocoded) feedback provided information not only about the lexical content of the target sentences but also about the phonemic, prosodic and syntactic content (as feedback and target sentences were identical, aside from the vocoding). It is conceivable then that sub- and supra-lexical information offered by feedback trials could have provided additional support for connections to stored linguistic knowledge, yielding predictions about upcoming speech input and facilitating adaptation. The present work investigated the relative contribution of different levels of linguistic information by examining the extent to which feedback trials needed to align with the Mandarin-accented sentence, manipulating the degree of match on different linguistic dimensions. Given that naturally-occurring conversations rarely include exact repetitions of "distorted" utterances in their "undistorted" form, it is important to establish that linguistically-guided adaptation does not depend entirely on exactly matching "feedback." It is possible that other levels of linguistic structure, particularly those that can potentially be the focus of interlocutor entrainment over the course of a conversation (e.g., prosody), could constrain and guide perceptual adaptation. This then could provide a critical link between constructed perceptual adaptation training regimens in the laboratory to naturally-occurring, real-world, experience-dependent adaptation.

Using a similar paradigm as Davis et al. (2005), listeners in the present study first transcribed Mandarin-accented English sentences in noise. After each transcription, feedback

sentences produced by a native English talker were presented that either aligned with the target on 1) all linguistic levels (Lexical Match), 2) sub-lexical, prosodic and syntactic levels with real words (Lexical Mismatch) or 3) sub-lexical, prosodic and syntactic levels with non-words (Jabberwocky). These three conditions were compared to two control conditions, 4) non-English (Korean) speech feedback (Language Mismatch), and 5) Mandarin-accented only exposure (Accent Control). These controls allowed us to establish a baseline amount of improvement as a result of accent exposure (Accent Control) as well as the degree to which hearing speech (Language Mismatch) in the feedback interval facilitated learning. It is important to note that a basic assumption of this paradigm is that the Lexical Match condition would (trivially) result in (near) perfect recognition of the target sentence following feedback (i.e., feedback-guided within-trial correction via "revelation" of the intended sentence in clear, native-accented speech), whereas within-trial improvement would be significantly less than perfect in all other conditions. Thus, the critical measure of feedback-guided perceptual adaptation in all conditions was generalization to novel sentences (i.e., recognition accuracy improvement across trials).

If listeners leverage multiple sources of information during perceptual adaptation, one could hypothesize that the greater the number of connections between the disambiguating information and the input, the more efficiently and confidently the system can refine its predictions about future Mandarin-accented input. This would predict that conditions where the Mandarin-accented target and native-accented feedback trials overlap to a greater extent on a larger number of linguistic dimensions (Lexical Match) should demonstrate greater adaptation relative to conditions with less overlap (Lexical Mismatch and Jabberwocky).

Alternatively, for the sake of generalization to novel Mandarin-accented input, the highly generalizable sub- and supra-lexical information present in the Lexical Mismatch and Jabberwocky conditions may be sufficient for the perceptual system to promote adaptation to novel items. Under this scenario, all English feedback conditions should yield comparable learning, showing larger gains relative to the Language Mismatch and Accent Control conditions.

## 2. Methods

### *2.1 Participants*

One hundred English monolingual listeners, self-reporting no speech or hearing deficits at the time of testing, were included in the present study and received course credit or were paid for their participation (F=75; *M age*=19.8 years). Monolingual American English listeners were defined as having no experience with a language other than English prior to the age of 11 for more than 5 hours per week. Listeners were randomly assigned to one of 5 conditions (n=20 per condition). Data collection was stopped once 20 participants were reached in each condition, as prior work has found reliable between-group differences with samples of that size (e.g., Adank, Hagoort, & Bekkering, 2010; Mitterer & McQueen, 2009; Witteman, Bardhan, Weber, & McQueen, 2014).

### *2.2 Stimuli*

The target materials were a set of 26 sentences taken from the revised Bamford Kowal-Bench (BKB) Standard Sentence Test (Bamford & Wilson, 1979). These items are declarative, monoclausal sentences, each containing 3 to 4 keywords (e.g., "The boy fell from the window"). They were produced by a male, Mandarin-accented talker of medium-

intelligibility, as determined in Bradlow and Bent (2008), as well as by a male, native American English talker. For the Lexical Mismatch condition, 26 Hearing-in-Noise Test sentences that did not overlap with the target BKB sentences were taken from the ALLSSTAR database (Bradlow et al., 2011), produced by a male native American English talker. For the Jabberwocky condition, the 26 HINT sentences from the Lexical Mismatch condition were adapted, replacing the content words with English pseudowords. The phonemes from the content words in each feedback sentence in the Lexical Mismatch condition were used to create the novel pseudowords in the Jabberwocky condition. Thus, the feedback sentences in both the Lexical Mismatch and Jabberwocky conditions contained the same phonemes and syntactic structure and were produced with highly similar phrase-level prosody (declarative intonation).

**Table 2.1 Training experiment setup for each condition**

<table>
<tr><td>Condition</td><td>Trial 1</td><td rowspan="6">Transcription</td><td>Trial 2</td><td>Trial 3</td></tr>
<tr><td>Lexical Match</td><td rowspan="5">Mandarin-accented in noise<br>“The children dropped the bag”</td><td>Native-accented in clear<br>“The children dropped the bag”</td><td rowspan="5">Mandarin-accented in noise<br>“The children dropped the bag”</td></tr>
<tr><td>Lexical Mismatch</td><td>Native-accented in clear<br>“The wife helped her husband”</td></tr>
<tr><td>Jabberwocky</td><td>Native-accented in clear<br>/The bɛft faɪzd her wʌldɚn/</td></tr>
<tr><td>Language Mismatch</td><td>Korean in clear<br>“ごくろうさまな話だ”</td></tr>
<tr><td>Accent Control</td><td>Mandarin-accented in noise<br>“The children dropped the bag”</td><td></td></tr>
</table>

Each target sentence was paired with the same feedback sentence across participants within the Lexical Match, Mismatch and Jabberwocky conditions. Target sentences contained an average of 15 phonemes (range 11-19), with an average 35% phonemic overlap (range 7-59%) from their associated feedback sentences (Lexical Mismatch or Jabberwocky conditions). With regards to the amount of lexical overlap, no content words were shared, and there were only 10 instances of function words overlapping between target and feedback sentences, which typically involved both sentences containing the function word "the". Therefore, this phonemic overlap largely stemmed from shared individual phonemes or clusters (e.g., "found" in the target sentence, "ground" in the feedback sentence), rather than entire content words. In terms of syntactic structure, all target and feedback sentences contained an initial noun phrase subject followed by a verb phrase. The verb phrase typically contained an XP: a direct object noun phrase (e.g., "The children helped their teacher"), prepositional phrase (e.g., "They washed in cold water") or adjectival phrase (e.g., "Sugar is very sweet").

*2.3 Procedure*

Participants underwent 2 blocks of foreign-accented speech training, adapted from Davis et al. (2005). The trial structure contained either 2 or 3 trials in a set (Table 2.1). In all conditions, listeners' task after the first trial was to transcribe a foreign-accented sentence, presented in speech-shaped noise at +5 dB SNR, with no limit on response time. After the transcription, participants received feedback in one of several different formats (see Table 2.1 for example sentences). Feedback consisted of either a native-accented production of the target sentence (Lexical Match), a mismatching sentence (Lexical Mismatch), a jabberwocky

sentence (Jabberwocky) or a Korean sentence (Language Mismatch). This was then followed by a repetition of the Mandarin-accented target sentence in noise. In these conditions, the clear production and the Mandarin-accented target repetition were separated by 500 ms. In the Accent Control condition, listeners heard one repetition of the target Mandarin-accented sentence in noise with no intervening feedback, to establish a baseline with respect to how much learning could occur from exposure to just the Mandarin-accented speech.

Each training block contained 2 blocked repetitions of 13 unique target sentences, whereby listeners heard and transcribed 13 target sentences (Block 1A) before receiving the same set of 13 sentences again (Block 1B). Block 2 then introduced a new set of 13 sentences (in a similarly constructed Block 2A and Block 2B). Which target sentences listeners received in the first and second training blocks were counterbalanced across participants. Training therefore consisted of 52 sentence sets (26 sentences sets x 2 repetitions), where each trial set could contain either 2 (Accent Control) or 3 productions (all other conditions). All participants heard the same number of Mandarin-accented trials over the course of the experiment (104 sentences), which included target transcription presentations and subsequent target repetitions. This task was administered over headphones at a comfortable listening volume in a sound-attenuated booth.

**3. Results**

Strict keyword accuracy was tabulated for Block 1A and Block 2A (i.e., a comparison of transcription accuracy for the first repetition of two different sentence sets), whereby each keyword was considered either ‘correct’ (coded as 1) or ‘incorrect’ (coded as 0). Homophones and obvious spelling errors were not considered incorrect; however, words with

inaccurate morpheme substitutions or omissions were scored as incorrect. Figure 2.1 depicts the mean improvement in accuracy (difference between Block 1A and Block 2A), and Figure 2.2 plots the mean keyword accuracy scores by participant for Block 1A against Block 2A. Logistic mixed effects regression models were implemented to analyze the data (Baayen, Davidson, & Bates, 2008), with keyword transcription accuracy as the dependent variable. A model was constructed with Helmert contrast-coded fixed effects of Condition (A: Accent Control vs. all other conditions; B: Language Mismatch vs. Jabberwocky, Lexical Mismatch and Lexical Match; C: Jabberwocky vs. Lexical Mismatch and Lexical Match; D: Lexical Mismatch vs. Lexical Match) and Block (1, 2) along with their interactions. The maximal random effects structure that would converge was implemented, which included random intercepts for participant and keyword, as well as random slopes for Condition by keyword and Block by participant.

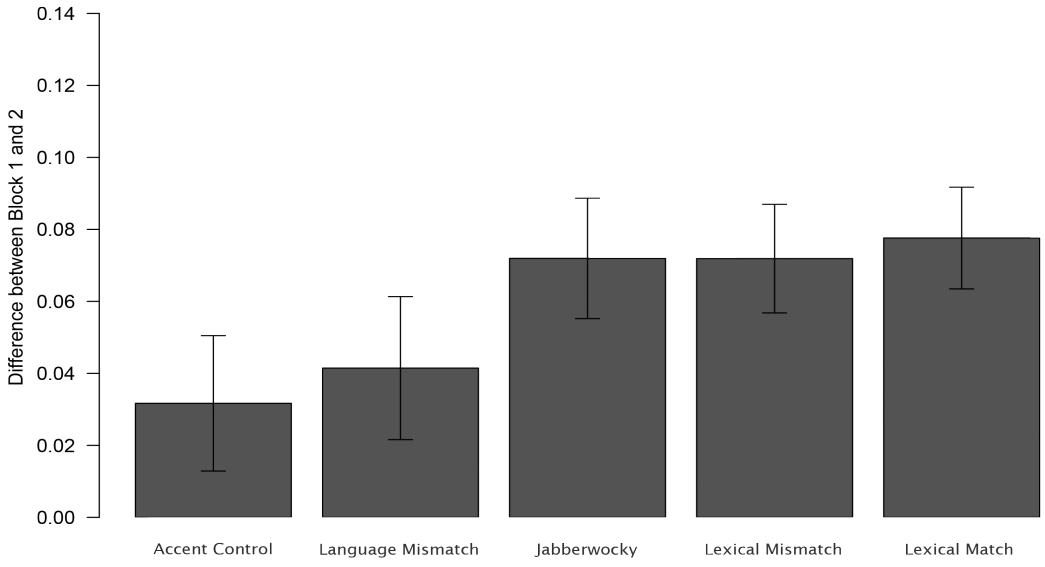


**Figure 2.1 Mean difference in proportion keyword correct between Block 1 and Block 2. Error bars indicate +/- 1 standard error.**

Results revealed a significant main effect of Condition B (Language Mismatch vs. Jabberwocky, Lexical Mismatch and Lexical Match; β=-0.51, SE β=0.25, $\chi^2(1)=4.1507$, p=0.04), with listeners in the Language Mismatch condition (*M*=86%) performing better across blocks relative to the English feedback conditions (*M*=83%). A main effect of Condition D (Lexical Mismatch vs. Lexical Match) was also obtained, with higher accuracy rates overall in the Lexical Match condition (*M*=85%) as compared to the Mismatch condition (*M*=82%). No other Condition effects reached significance ($\chi^2<2.63$, $p>0.10$). A highly significant main effect of Block (β=0.77, SE β=0.09, $\chi^2(1)=59.373$, $p<0.001$) indicates that, across conditions, keyword identification accuracy significantly improved from Block 1 to Block 2.

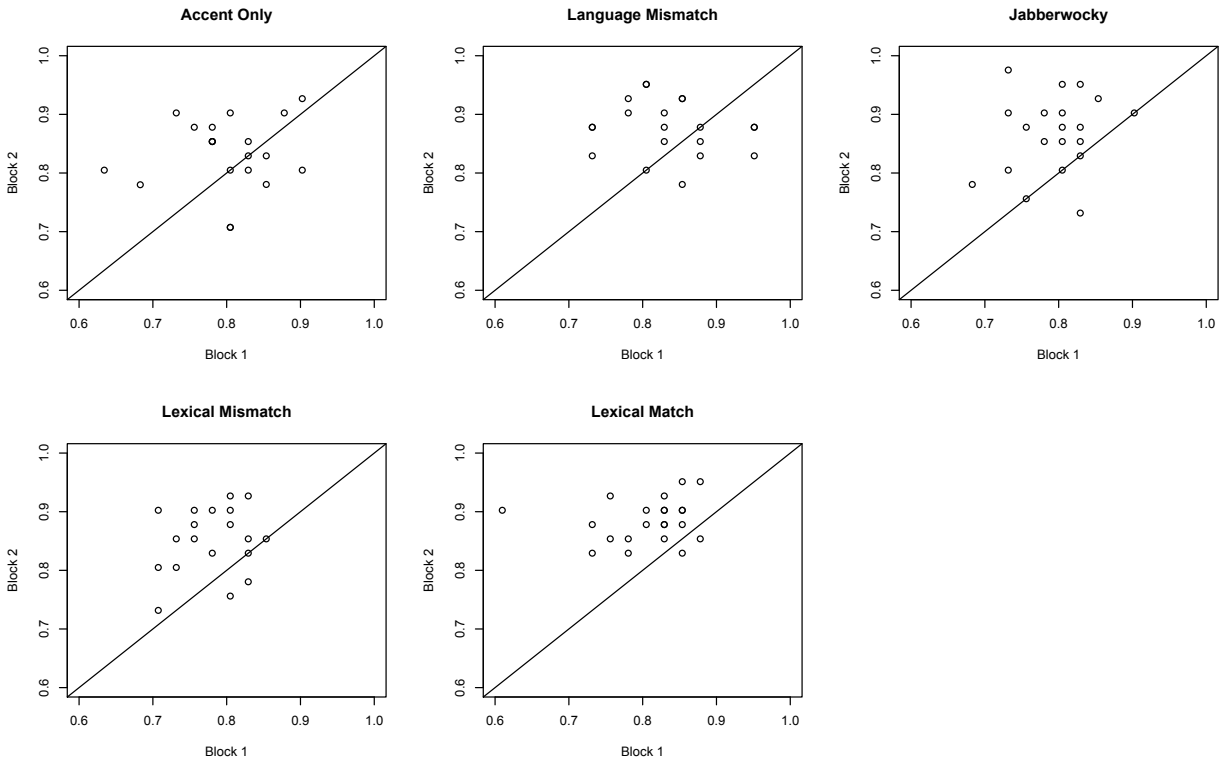


**Figure 2.2 Mean proportion keyword correct for Block 1 (x-axis) and Block 2 (y-axis) for each participant by condition. Points above line indicate adaptation (accuracy gain from Block 1 to 2).**

Critically, a significant Condition A (Accent Control vs. all other conditions) x Block interaction was found (β=0.77, SE β=0.33, $\chi^2(1)=5.0077$, p=0.025), indicating that the presence of speech feedback significantly improved performance relative to exposure to only Mandarin-accented trials. The Condition B (Language Mismatch vs. Jabberwocky, Lexical Mismatch and Lexical Match) x Block interaction was marginally significant (β=0.58, SE β=0.34, $\chi2(1)=2.8279$, p=0.09). As evidenced by Figures 2.1 and 2.2, there was a distinct numerical trend for greater adaptation to occur in conditions with English feedback relative to Korean. The remaining Condition x Block interactions were not significant ($\chi^2<0.3844$, $p>0.54$), indicating that the English feedback conditions did not differ with respect to the amount of adaptation that occurred from Block 1 to 2.

## 4. Discussion

The present work found enhanced adaptation in the English feedback conditions (Lexical Match, Lexical Mismatch and Jabberwocky) relative to Language Mismatch and Accent Control conditions. These findings suggest that listeners' perceptual systems leveraged linguistic information present in the externally-provided feedback, in the form of native-accented speech, resulting in improved sentence recognition. Consistent with research reporting the facilitative effect of externally-provided, matching lexical feedback on adaptation, significantly larger gains were made in the Lexical Match relative to the Accent Control condition (Davis et al., 2005; Hervais-Adelman et al., 2008; Mitterer & McQueen, 2009). However, prior work using this particular paradigm (e.g., Davis et al., 2005) always involved a within-trial match of feedback and target sentences (i.e., the target and its feedback sentence were always identical to each other). The present work revealed that the feedback did not need to match the target in order for enhanced adaptation to occur, suggesting that connections with non-lexical linguistic dimensions guided the adaptation process.

One might predict that providing listeners with the identity of the target sentence would yield the largest gains over the course of training, as the within-trial feedback, being the same sentence as the Mandarin-accented input, aligned with the target on all linguistic dimensions. This might be expected to enable listeners to better access stored linguistic knowledge, as a result of the concomitant increase in intelligibility of the target repetition, to then notice any discrepancies between the input and their predictions, and adjust the appropriate categories accordingly in a manner that could be generalized to the novel input of

the subsequent trials. The relative predictive strength of the different types of information in the feedback trials were initially predicted to be tied to the certainty with which a listener would make a perceptual adjustment, whereby highly predictive information (Lexical Match) might increase listeners' certainty about how to categorize the input, thereby facilitating adaptation, relative to less predictive information (Lexical Mismatch or Jabberwocky). However, the present findings suggest a divergence between predictive strength and certainty, such that listeners drew upon varied types of information (regardless of its predictive strength) in order to enhance their certainty and make them willing to adjust the relevant categories. No additional advantage was found for the Lexical Match condition over the Lexical Mismatch and Jabberwocky conditions. All three conditions were found to result in significantly greater adaptation relative to accent only exposure. While not sharing any lexical content words, there were nonetheless points of connection between the feedback and Mandarin-accented input within trial sets in these conditions. This degree of overlap in the phonemic content, prosodic and syntactic structure of these sentences may have facilitated the connection to stored linguistic knowledge and improved predictions about future, novel samples of Mandarin-accented input. These findings provide insight into how the perceptual system leverages different sources of linguistic information during adaptation, namely that some degree of linguistically-relevant alignment (sharing of English language features) of Mandarin-accented target and feedback trials was sufficient to promote generalized adaptation under these listening conditions. As the perceptual system did not need a lexical match in order to see enhanced comprehension of Mandarin-accented speech, this suggests that the system leverages multiple sources of linguistic information (e.g., lower-level

phonemic content, higher-level lexical or pragmatic information) present in the surrounding communicative context.

Future work should investigate just how many points of connection one can strip away before the information no longer becomes useful for the perceptual system. For example, would misaligning the syntactic or prosodic structure between target and feedback trials inhibit adaptation? Would the system be able to utilize just phonotactically legal strings of native-accented phonemes outside of any syntactic or prosodic frame? It also remains for future research to determine how the degree of linguistically-relevant alignment interacts with the intelligibility of the speech. It is conceivable that a benefit for greater alignment between target and feedback trials would emerge in lower intelligibility conditions (e.g., higher levels of noise or a more heavily foreign-accented speaker), as it may require listeners to leverage as many sources of information as possible. More difficult access to the linguistic content of the speech input may make it challenging for the perceptual system to only have, for example, native-accented exemplars of lower-level phonemic information to use as a source for prediction-generation and adjustments. As a result, converging sources of information, such as higher-level lexical or pragmatic information, may play a larger role during adaptation in such adverse contexts.

# CHAPTER 3: KNOWLEDGE- AND SIGNAL-DRIVEN PERCEPTUAL LEARNING OF NOVEL ACCENTED SPEECH

## 1. Introduction

In real-world listening situations, listeners must contend with an extensive amount of variability in their speech input, from differences in vocal tract size to the presence of a speech impairment or foreign accent. However, listeners are remarkably efficient at adapting to systematic deviations from long-term linguistic regularities by exploiting a range of signal-based and knowledge-based sources of information to facilitate the mapping between stored linguistic representations and speech input that deviates from them. These adjustments have been found to occur very rapidly, in as short a time frame as a few sentences (Clarke & Garrett, 2004), and can have lasting effects on listeners' perceptual systems (Zhang & Samuel, 2014). This process is known as perceptual learning: lasting changes in our perceptual system that result from it constantly attempting to refine how it responds to its environment (Goldstone, 1998).

The kinds of information that listeners utilize to make these changes in the perceptual learning of speech can be classified in one of two ways: 1) knowledge-based, which includes prior linguistic knowledge (e.g., phonotactics, lexical items) and 2) signal-based, which includes information extant in the signal that one would not necessarily need prior long-term linguistic knowledge to extract (e.g., visual cues, variability within the signal, etc.). Both of these types of information have been found to be effective in facilitating perceptual speech learning (e.g., Bertelson, Vroomen, & De Gelder, 2003; Bradlow & Bent, 2008; Cutler,

McQueen, Butterfield, Norris, & Planck, 2008; Norris, McQueen, & Cutler, 2003). However, the relationship between the different types of information that listeners utilize in perceptual speech learning has yet to be fully clarified. The present study will seek to examine how these types of information interact with one another.

*1.1 Knowledge-driven perceptual learning*

Considerable attention has been paid to the role of lexical information in the perceptual learning of acoustic-phonetic properties in L1 speech (e.g., Eisner & McQueen, 2006; Kraljic & Samuel, 2005, 2006; Maye, Aslin, & Tanenhaus, 2008; McQueen, Cutler, & Norris, 2006; Norris et al., 2003). This literature has largely focused on lexically-guided phonetic retuning, whereby listeners are presented with ambiguous sounds (e.g., a sound midway between /f/ and /s/) in lexical contexts intended to bias them to perceive the ambiguous sound as, for example, either [f] or [s] (e.g., "giraffe" or "dinosaur", respectively). Because there are no such words as "girasse" or "dinofaur", listeners learn to associate the ambiguous sound with either [f] or [s]. In an identification test of an auditory [f]-[s] continuum, listeners' category boundaries were found to have subsequently shifted after exposure, suggesting that listeners were able to utilize lexical knowledge to make adjustments to pre-lexical phonetic categories. Similar findings implicating the role of lexical information in perceptual learning have also been found for adaptation to noise-vocoded speech (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005), in a paradigm where listeners are provided access to the lexical content of the speech while adapting to the distorted speech.

While prior work on this issue has focused predominantly on adaptation guided by lexical information, some research has examined other linguistic sources of guidance during adaptation. Cutler et al. (2008), using the same paradigm described above, found that presenting listeners with an ambiguous sound between /f/-/s/ in nonword contexts where only one of the segments was phonotactically legal also resulted in the retuning of phonetic category boundaries, in much the same ways as lexically-biasing contexts. Moreover, Hervais-Adelman, Davis, Johnsrude, and Carlyon (2008) reported that access to the phonological content of the speech signal (in the form of nonwords) was sufficient to induce comparable adaptation to noise-vocoded speech as access to lexical content.

*1.2 Signal-driven perceptual learning*

Plasticity in the speech perception system is not confined to knowledge-driven perceptual adjustments. Indeed, the signal itself can provide cues that can influence the results of this re-tuning. One of the seminal studies in perceptual learning examined the recalibration of native phonetic categories guided by audio-visual cues (Bertelson et al., 2003). During an initial exposure phase, listeners were presented with 3 different audio stimuli ([aba], [ada] or [a#a]), where [#] was a sound ambiguous between /b/ and /d/) that were dubbed onto video of a speaker producing /aba/ or /ada/. Post-exposure auditory-only identification tests revealed that listeners were strongly influenced by the visual context of the exposure phase (relative to pre-exposure baseline identification tests), such that listeners considered the ambiguous auditory-only token to be whatever sound the face was producing during the audio-visual exposure phase. These findings provide evidence that, not only does visual information influence sound perception during audio-visual exposure (similar to the

well-known McGurk effect; McGurk & MacDonald, 1976), but visual information can also recalibrate the auditory speech perception system in a way that influences auditory classification of subsequent stimulus exposures even when a visual signal is not available.

Sjerps and McQueen (2010) also reported findings consistent with signal-driven perceptual learning. In a similar paradigm as Norris et al. (2003), listeners were exposed to a signal-correlated noise[1] version of /θ/ ([#]), which replaced either /f/ or /s/ in items in a lexical decision task (simple word-nonword identification for each item). They were tested in a cross-modal identity priming task, where they heard aurally-presented prime words that were either unrelated or contained the noise segment in place of an /f/ or /s/ (e.g., [dɛ#] "deaf" or [moʊ#]) followed by visually-presented letter strings (e.g., *deaf* or *mouse*). Even for listeners who received a lexical bias indicating that the sound should be interpreted as /s/ during exposure (e.g. "mouse"), listeners perceived [#] as an instance of [f] (i.e., auditory "mou#" did not prime visual *mouse*, but auditory "dea#" did prime visual *deaf*). The authors argue that [#] and [f] were more spectrally similar than [#] and [s], indicating that signal cues can override lexically-biased perceptual adaptation (i.e., the influence of higher-level information) in certain cases. Listeners learned to associate [#] to [f] in [f]-biasing words, but did not learn to associate [#] to [s] in [s]-biasing words, providing evidence for signal-driven perceptual learning.

Another factor implicated in perceptual learning is input variability. While trial-to-trial variability has been found to lead to decrements in word recognition and recall performance as a result of the increased processing demands required to compensate for the

[1] Signal-correlated noise has a flat spectrum within the amplitude, duration and spectral range of speech.

variability across a test or exposure session (e.g., Goldinger, Pisoni, & Logan, 1991; McLennan & Luce, 2005; Mullennix, Pisoni, & Martin, 1989), its impact on speech processing is not completely negative. In particular, input variability has been found to be an important source of information in successful perceptual speech learning. Infant phonological processing in early word learning has been found to be enhanced when training was completed with multiple talkers relative to a single talker (Rost & McMurray, 2009). Similarly, research on adult listeners' adaptation to L1 speech has reported augmented perceptual learning of foreign-accented speech (e.g. Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009) and dialects (Clopper & Pisoni, 2004) with multi-talker training relative to training with a single talker. Variability has been posited to facilitate perceptual learning by providing listeners with a range of varied speech input from which they can discover systematicities. From these systematicities, listeners can then make more generalized adjustments to long-term representations, which will subsequently yield more generalizable learning. Indeed, training with 5 different foreign accents of English (Mandarin-, Romanian-, Thai-, Hindi-, and Korean-accented English) led to more generalizable, accent-independent adaptation to foreign-accented English (i.e., to both Mandarin-accented English, a trained accent, and to Slovakian-accented English, a novel accent) than training on a single foreign accent of English (Baese-Berk et al., 2013).

*1.3 Generalization in perceptual learning*

One of the critical features of linguistic competence is the ability to handle novelty, whether it be new lexical items or unfamiliar talkers; thus, generalization should be considered an essential test for perceptual learning for speech. While lexically-guided

phonetic adjustments have been shown to generalize to novel items, the results on whether these adjustments generalize across speakers has been mixed, with speaker-specific adjustments for certain speech contrasts (e.g., fricatives; Eisner & McQueen, 2005; Kraljic & Samuel, 2005) and more generalized learning for other contrasts (e.g., stop voicing; Kraljic & Samuel, 2007). The authors of these studies have claimed that the presence of speaker-specific cues in fricatives (spectrally-based cues) and the lack of such cues in stop contrasts (temporally-based cues) modulated whether or not learning generalized to novel speakers (Kraljic & Samuel, 2007). Variability has also been implicated in generalized perceptual learning, whereby talker-general learning of naturally-produced foreign-accented speech was only found following training with multiple talkers but not with single-talker training (Bradlow & Bent, 2008). However, other recent work has found robust generalization of back vowel lowering accent patterns to novel talkers, both acoustically similar and dissimilar to the trained talker (Weatherholtz, 2015), despite only training listeners with a single talker. Weatherholtz (2015) exposed listeners to a back vowel lowered chain shift (e.g., /ʊ/ "wooden" ➔ /oʊ/ "woden") before testing them on items in lexical decision and word identification tasks to see if the pre-test exposure would cause listeners to loosen their criteria for word (versus non-word) identification (e.g. "woden" accepted as a word due to the learned vowel lowered chain shift). The author noted that one possible explanation for his finding robust generalization while prior work did not relates to the nature of the foreign-accented or ambiguous sounds being employed, namely vowels as opposed to consonants. Vowels are inherently more perceptually variable relative to consonants, which may have resulted in listeners being more willing to adapt to atypical vowel pronunciations.

In addition to generalization to novel talkers, prior work has investigated the extent to which adaptation generalizes from trained to untrained phonemes. Kraljic and Samuel (2006) exposed listeners to an ambiguous /d/-/t/ contrast in lexically-biasing contexts, which impacted subsequent categorization of both /d/-/t/ and /b/-/p/ continua. However, in a series of experiments using the visually-guided recalibration paradigm, a study by Reinisch, Wozny, Mitterer, and Holt (2014) was unable to obtain generalized learning from /aba/-/ada/ to other phonetic contexts (e.g., /ubu/-/udu/) or other contrasts (phonetically-cued in the same way, e.g., /ama/-/ana/), with adaptation restricted to the phonetic context presented during exposure. Similarly, Maye et al. (2008) found successful learning of front vowel lowering accent patterns to which listeners were exposed but which did not generalize to back vowel lowered or front vowel raised items. The authors concluded that adaptation was vowel- and direction-specific. In contrast, Weatherholtz (2015) found listeners who were exposed to a system of back vowel lowering were able to generalize their learning to a system of back vowel raising and front vowel lowering, which may have arisen from a general broadening of listeners' vowel categories.

*1.4 Current research*

The present work considers both knowledge-driven and signal-driven perceptual learning as sharing a common mechanism—namely that they both provide information that can be used to generate predictions about future speech input. Guediche, Blumstein, Fiez and Holt (2014) posited that a variety of forms of information (e.g. long-term linguistic knowledge, visual cues) that can disambiguate ambiguous or distorted speech input is used to internally-generate predictions about future input. Any discrepancy between these predictions

and the incoming input will yield a prediction error signal, which will subsequently trigger adaptive adjustments in an effort to improve the match between future predictions and incoming input. For example, when listening to a foreign-accented talker produce the word *giraffe,* he might sound like he is saying something more like "girass", where the [f]-sound is closer to an [s]-sound. However, listeners draw upon knowledge-based information (such as lexical knowledge) and consider that the produced word is likely *giraffe* because "girass" is not a real word in English. This mismatch between what listeners predicted was said based on their knowledge of English words (in this case, *giraffe*) and what was actually produced (something closer to "girass") leads them to make adaptive adjustments to their [f]-sound category, expanding it to include this ambiguous sound (e.g., Norris et al., 2003). As a result, the next time they hear the word "giraffe" or a word containing an [f]-sound produced by this talker, their predictions will be better tuned to what will actually occur in the speech input, yielding improved comprehension. Moreover, signal-based information, such as talker variability, could also provide information to generate predictions. Exposure to multiple talkers enables listeners to extract systematicities from the speech and make talker-general predictions. For instance, if multiple talkers produce an ambiguous /f/-/s/ sound (instead of an unambiguous /f/), listeners, when encountering a novel talker, might predict that such an ambiguous sound would also belong to the /f/ category for that talker. Conversely, listeners exposed to only a single talker would adjust their predictions about that specific talker's category boundaries; however, they might not have strong evidence to make adjusted predictions for other talkers.

Kleinschmidt and Jaeger (2015) further pursued the notion of predictive processes in adaptation in a formally explicit Bayesian model of perceptual adaptation, the "ideal adapter framework", which considers speech perception as involving prediction and inference under uncertainty. The model posits that the perceptual system holds beliefs about talker-specific (or situation-specific) generative linguistic models—that is, beliefs about the particular distribution of cues for each sound category based on prior experience observing these cue distributions. Listeners utilize knowledge of higher-level linguistic units, comparing them to determine how well each one predicts the incoming signal. Adaptation occurs in part as a product of updating their beliefs about the statistics of the relevant categories for a given speaker (or a particular situation/context). This can entail shifting the mean of a particularly category in the direction of the observed values or increasing the variance of that category (both routes are predicted to be possible by this model).

If listeners are drawing upon different sources of information to help generate predictions and update their beliefs about cue distributions within upcoming speech input, then we would posit that the predictive strength of the information, by which we mean the degree to which it narrows the space of possible options or the extent to which it reduces uncertainty about how a stimulus should be categorized, would modulate the speed and generalizability of learning. As such, different levels of linguistic information (phonological, lexical, semantic or pragmatic context) may be more or less predictive about upcoming speech input. While access to lexical information does appear to play a substantive role in perceptual speech learning (e.g. Norris et al., 2003; Maye et al., 2008), less is known about the relative contributions of lexical and other higher-levels of linguistic information (e.g.,

semantic context) in generating such predictions. Moreover, listeners likely draw upon multiple concurrent sources of information to inform perceptual adjustments; however, relatively little is known about how different types of information interact with each other.

The present work sought to examine how different types of prediction-generating information (knowledge- and signal-based) contribute to and interact with one another in perceptual adaptation. Listeners were exposed to a novel accent of English, manipulating predictive strength along various dimensions: the type of feedback regarding the Non-Standard American English (NSAE)-accented items (lexical, semantic context), the number of talkers during exposure (single, multiple), and phoneme type (vowels, consonants). A novel accent pattern was constructed such that it contained both vowel deviation patterns (a word such as “bleak” pronounced as “blick) as well as consonant patterns (“threw” pronounced as “trew”). Exposure consisted of either a written presentation of a word or a moderately-predictive semantic context followed by an auditory presentation of the target item produced by either a single talker or multiple talkers throughout training. Following exposure, listeners were tested with lexical decision and word identification tasks and compared with listeners who had not received NSAE-accented exposure. For the lexical decision task, adaptation would manifest as a higher proportion of items (considered nonwords in a Standard American English accent) endorsed as lexical items by trained listeners relative to control listeners. Additionally, for the word identification task, a higher proportion of items transcribed based on the novel accent patterns (e.g., “blick” identified as “bleak”) would be indicative of adaptation.

We hypothesize that the predictive strength of the disambiguating information modulates learning, whereby information provided in the lexical conditions should provide more robust predictions (by virtue of being an exact lexical match to the target word) relative to the semantic context conditions, which provides a strong possible candidate for the identity of the target. This would predict a significant main effect of feedback type, with listeners in the lexical conditions demonstrating higher lexical endorsement rates for NSAE-accented items and word identification accuracy relative to those in the semantic conditions. Alternatively, if the perceptual system can leverage any type of linguistically-relevant disambiguating information (lexical or semantic context) to generate predictions about upcoming speech input and update its beliefs about the relevant cue distributions, then this would predict no significant effect of Feedback, whereby both feedback conditions would not significantly differ from each other but both differ significantly from control listeners (who only completed the test tasks and did not receive training).

Additionally, multi-talker exposure should have greater predictive strength than single-talker exposure, as listeners have more talkers over which to abstract and extract information with which to make more robust talker-general predictions. This would predict a Variability x Talker interaction, such that listeners in the multi-talker conditions would outperform single-talker conditions on items produced by the untrained test talker but perform similarly on the trained talker.

It is not yet clear how information derived from variability in the signal interacts with knowledge-based information in yielding predictions. Are certain sources of linguistic information negatively impacted by signal variability? Indeed, Clopper (2012) found an

attenuated semantic predictability benefit in speech-in-noise perception as a product of dialect variability. It is conceivable that high talker variability will attenuate the effectiveness of semantic predictability in facilitating adaptation. This would predict a significant Variability x Feedback x Talker interaction, such that listeners in the multi-talker condition with semantic context feedback would show less adaptation relative to those in the single-talker condition with semantic context feedback for the trained talker, whereas this difference may not arise in the lexical conditions. Alternatively, while variability may attenuate the semantic predictability benefit for listeners, there may be sufficient disambiguating information to overcome this difficulty and facilitate learning, predicting no significant difference as a function of talker variability for the semantic context conditions.

Finally, prior work has typically only used either vowels (e.g., Maye et al., 2008; Weatherholtz, 2015) or consonants (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2007), with vowel deviation patterns typically yielding robust talker-generalization but more limited generalization with consonant deviation patterns. Weatherholtz (205) posited that this discrepancy may arise because consonants contribute relatively more to lexical identity and word recognition as compared to vowels. This is likely due to a number of factors including the higher number of consonants relative to vowels in phonemic systems and the fact that consonants appear to more tightly constrain lexical selection than vowels (Cutler, Sebastián-Gallés, Soler-Vilageliu, & Van Ooijen, 2000; Nespor, Peña, & Mehler, 2003), which may stem from listeners accruing experience with the fact that vowels perceptually vary more in context than consonants. As a result of vowels' perceptually mutable nature, listeners may be more willing to generalize knowledge about vowel variation to a novel talker as compared to

consonantal variation. This might predict that single-talker exposure for consonants results in relatively more uncertainty about how they should be classified relative to vowels when encountering a novel talker, which might result in a Phoneme Type x Talker interaction. Lexical endorsement rates and word identification accuracy for items containing consonant deviation patterns would be higher for trained relative to a novel talker, whereas this difference might not exist for vowel deviation patterns.

## 2. Methods

### *2.1 Participants*

One hundred and sixteen American English listeners, which included 35 participants (F=23; *M age*=19.3 years) tested in the lab and 80 participants (F=50; *M age*=36.6 years) tested on Amazon Mechanical Turk (an online service that provides an on-demand human workforce for a variety of different tasks), participated in the study and received course credit or were paid for their participation. American English listeners were defined as having English be their primary language prior to school and to be the language of instruction during school. Listeners were randomly assigned to one of 5 conditions (Control=21; MT Lexical=23, MT Semantic context=27; ST Lexical=20; ST Semantic context=22).

### *2.2 Procedure*

The experimental setup is outlined in Table 3.1. Participants in training conditions completed two blocks of training, each preceded by a probe lexical decision task. Following training, they completed two test tasks: 1) lexical decision, and 2) word identification. Trained listeners took approximately 45-60 minutes to complete these tasks. The Control condition only completed the lexical decision and word identification tasks.

**Table 3.1 Experimental setup. Each letter represents a unique talker.**

| | Probe 1 | Training 1 | Probe 2 | Training 2 | LexDec | Word ID |
|---|---|---|---|---|---|---|
| Single talker | A | A-A-A-A | A | A-A-A-A | A-E | A-E |
| Multi-talker | A | A-B-C-D | A | A-B-C-D | A-E | A-E |

Training consisted of 2 phases of passive exposure to a novel accent (Non-Standard American English, NSAE) accompanied by one of two types of feedback. Participants would first view a written presentation of either the word (Lexical conditions) or the context sentence (Semantic Context conditions). Once they had finished reading the display, they would click "NEXT" to then hear an auditory presentation of the training item produced in NSAE. In single talker conditions, all items were produced by the same talker and presented in blocks of 15 trials each. In multi-talker conditions, training items were divided between 4 talkers and were presented blocked by talker (as illustrated in Table 3.1). The item order was identical between single- and multi-talker conditions.

Prior to each training phase, participants completed a probe lexical decision task, each containing 36 trials. Each item was presented individually, and participants were asked to respond as quickly as possible as to whether they thought the item was a word or a nonword. Participants used a mouse to click one of two radial response options on the screen. They were instructed that the speaker in these probe blocks was a talker who spoke with a NSAE accent. Item presentation was randomized within each block, and the order of the probe blocks was counterbalanced across participants.

The lexical decision task after training followed the same procedure as the probe task, whereby participants responded "word" or "nonword" to each individually presented item. A total of 189 trials were presented, blocked by talker, with the trained talker always preceding the generalization talker. Participants were instructed that the talker was either someone they had heard during training or a novel talker. They were also informed that the novel talker produced the same accent as they had been exposed to during training.

Finally, participants completed a word identification task, which consisted of a total of 132 trials. Similar to the previous task, trials were blocked by talker (66 trials each), and the trained talker block was presented before the generalization talker block. Each item was presented individually, and participants were asked to type the word they heard. If they did not believe the word to be a real word, they could type 'X'. There was no limit on response time.

*2.3 Stimuli*

Following Maye et al. (2008), a novel accent (NSAE) was created by implementing a set of cross-category pronunciation deviations from Standard American English (SAE) speech, outlined in Table 3.2. These particular deviations were selected as they are known to be common pronunciation problems for non-native speakers of English (Avery & Ehrlich, 1992) and thus could be considered plausible deviations that listeners might encounter in foreign-accented speech. Unlike prior work (e.g., Eisner & McQueen, 2005; Maye et al., 2008), the NSAE accent contained both vowel and consonant deviation patterns. The training stimuli were naturally-produced by 4 phonetically-trained, male native speakers of Standard

American English. Stimuli used in test tasks were produced by one of the trained speakers as well as fifth male native speaker of English.

**Table 3.2 Trained NSAE accent deviation patterns**

| **NSAE-accented segments** | | |
|---|---|---|
| /i/ ➔ /ɪ/ | 'cream' [krim] ➔ 'crim' [krɪm] | Vowel lowering |
| /eɪ/ ➔ /ɛ/ | 'cake' [keɪk] ➔ 'kek' [kɛk] | Vowel lowering |
| /ɛ/ ➔ /æ/ | 'west' [wɛst] ➔ 'waest' [wæst] | Vowel lowering |
| /z/ ➔ /s/ | 'rose' [roʊz] ➔ 'rous'[roʊs] | Fricative devoicing |
| /θ/ ➔ /t/ | 'thirst' [θɝst] ➔ 'turst' [tɝst] | Interdental fricative ➔ Alveolar stop |
| /d/ ➔ /t/ (word-finally) | 'word' [wɝd] ➔ 'wert' [wɝt] | Final stop devoicing |

Training materials were a set of 120 intended real words. Here, "intended" real words indicate that these items would be considered real words when pronounced with a SAE accent; however, in NSAE, certain items appeared to be non-words (e.g. "snake" /sneɪk/ ➔ "snek" /snɛk/). Each item only contained one deviation pattern. The 120 training items were divided between the six accent deviation patterns (16 items per consonant deviation pattern, 24 items per vowel deviation pattern). The output of the NSAE accent had one of two possible outcomes: 1) Minimal pair change (real word "bait" /beɪt/ ➔ real word "bet" [bɛt]) or 2) Lexicality change (real word "pose" /poʊz/ ➔ non-word [poʊs]). In all conditions, 60 items underwent a Minimal pair change, and 60 items underwent a Lexicality change.

For the Semantic Context condition, a moderately predictive sentence was created for each training item (120 total). Moderate predictability was determined based on a norming experiment on Amazon Mechanical Turk, where native English-speaking participants (n=45) were presented with a sentence and had to fill in the blank (e.g., "I stirred ____ into my coffee"). The probability that the training item was provided as a response was on average 38% (range 7%-80%). Critically, no sentences were 0% or 100% predictive of the training item.

Each probe task contained 36 items, which included 12 trained pattern NSAE items (e.g., "brief" ➔ [brɪf]), 36 real words, and 12 non-words (e.g., [flɑɹ]). All of the trained pattern items involved a lexicality change (real word ➔ non-word). None of these items were seen during the training blocks, and they were produced by the speaker from the single-talker training conditions.

The lexical decision task presented a total of 189 items: 54 trained pattern NSAE items, half of which were presented during training and half novel, 25 untrained pattern items, 80 real words, and 30 non-words. Similar to the probe task, trained pattern items involved lexicality changes, such that they would be considered non-words to untrained listeners. Untrained pattern items underwent a novel set of pronunciation deviation patterns (Table 3.3) and were divided evenly between the 5 deviation patterns. Half of these items were produced by the single-talker training talker, and half were produced by a novel, generalization talker. Which items were produced by which talker was counterbalanced across participants. In both the probe and lexical decision tasks, real and non-word items did not contain any NSAE-accented segments (Tables 3.2 and 3.3), only using segments that

were unaffected by the NSAE accent. Nonword items, similar to trained and untrained items, were constructed to differ minimally from real words (differing by one phoneme).

**Table 3.3 Untrained NSAE accent deviation patterns**

| **NSAE-accented segments** | | |
|---|---|---|
| /ɪ/ ➔ /eɪ/ | 'witch' [wɪtʃ] ➔ 'weich' [weɪtʃ] | Vowel lowering |
| /u$^{w}$/ ➔/ʊ/ | 'fruit' [fɹut] ➔ 'frut' [fɹʊt] | Vowel lowering |
| /v/ ➔ /f/ | 'vote' [voʊt] ➔ 'fote' [foʊt] | Fricative devoicing |
| /ð/ ➔ /d/ | 'brother' [bɹʌðɚ] ➔ 'brudder' [bɹʌdɚ] | Interdental fricative ➔ Alveolar stop |
| /b/ ➔ /p/ (word-finally) | 'scrub' [skrʌb] ➔ 'skrup' [skrʌp] | Final stop devoicing |

The word identification task contained a total of 132 items, including 72 trained pattern NSAE items and 60 untrained pattern items. Half of the trials involved minimal pair changes, and half were lexicality changes. Additionally, half of the items were produced by the talker from the single-talker training condition, and half by the novel, generalization talker. All of these items were not seen in the training blocks.

## 3. Results

### *3.1 Lexical Decision Training Probe task*

Lexical decisions were tabulated in each of the probe blocks, with 'word' responses coded as 1 and 'nonword' responses coded as 0. Response times were not analyzed, as the interpretation of these findings would be difficult given the design of the study, whereby there is considerable ambiguity as to the correctness of 'word' responses to target trials

(given that the lexical status of these items was designed change as a function of exposure). Lexical decision data were submitted to logistic linear mixed effects regression (LMER) models (Baayen et al., 2008). The first model examined trained pattern items and contained contrast-coded fixed effects for Probe Block (1, 2), Talker Variability (Single, Multiple) and Feedback (Lexical, Semantic Context) along with their interactions. The maximal random effects structure that would converge was implemented, which included random intercepts for participants and items, a by-participant random slope for Probe Block and by-item random slopes for Talker Variability and Feedback. Model comparisons were conducted to determine which factors made a significant contribution to the model. A significant main effect of Probe Block was found ($\beta$=1.84, SE $\beta$=0.15, $\chi^2(1)$=96.39, $p<0.001$), indicating that listeners significantly increased their lexical endorsement rate of trained pattern items from Block 1 to Block 2. No other effects or interactions were significant ($\chi^2<2.35$, $p>0.13$). A similar model was constructed examining responses to the nonword items. A significant effect of Probe Block was found ($\beta$=1.4, SE $\beta$=0.22, $\chi^2(1)$=39.49, $p<0.001$), indicating a tendency for word responses to increase, even to nonword items. All other main effects and interactions did not reach significance ($\chi^2<2.2$, $p>0.13$).

To examine whether listeners' endorsement patterns differed as a function of item type (Figure 3.1), an additional model was constructed with fixed effects of Probe Block (1, 2) and Item Type (Trained Pattern vs. Nonword). Significant main effects of Block ($\beta$=1.59, SE $\beta$=0.14, $\chi^2(1)$=86.96, $p<0.001$) and Item Type ($\beta$=1.6, SE $\beta$=0.34, $\chi^2(1)$=18.22, $p<0.001$) were found, along with a significant Block x Item Type interaction ($\beta$=0.57, SE $\beta$=0.20, $\chi^2(1)$=7.69, $p=0.006$). Follow-up models investigating this 2-way interaction, examining item

type for each Block, revealed significantly higher endorsement rates for trained pattern items in both Block 1 (β=1.32, SE β=0.39, $\chi^2(1)=10.37$, p=0.001) and Block 2 (β=2.02, SE β=0.38, $\chi^2(1)=21.07$, p<0.001). Thus, the interaction indicates that while there was a tendency for a higher lexical endorsement rate for the trained pattern items relative to nonword items prior to training, this difference was significantly larger by Block 2.

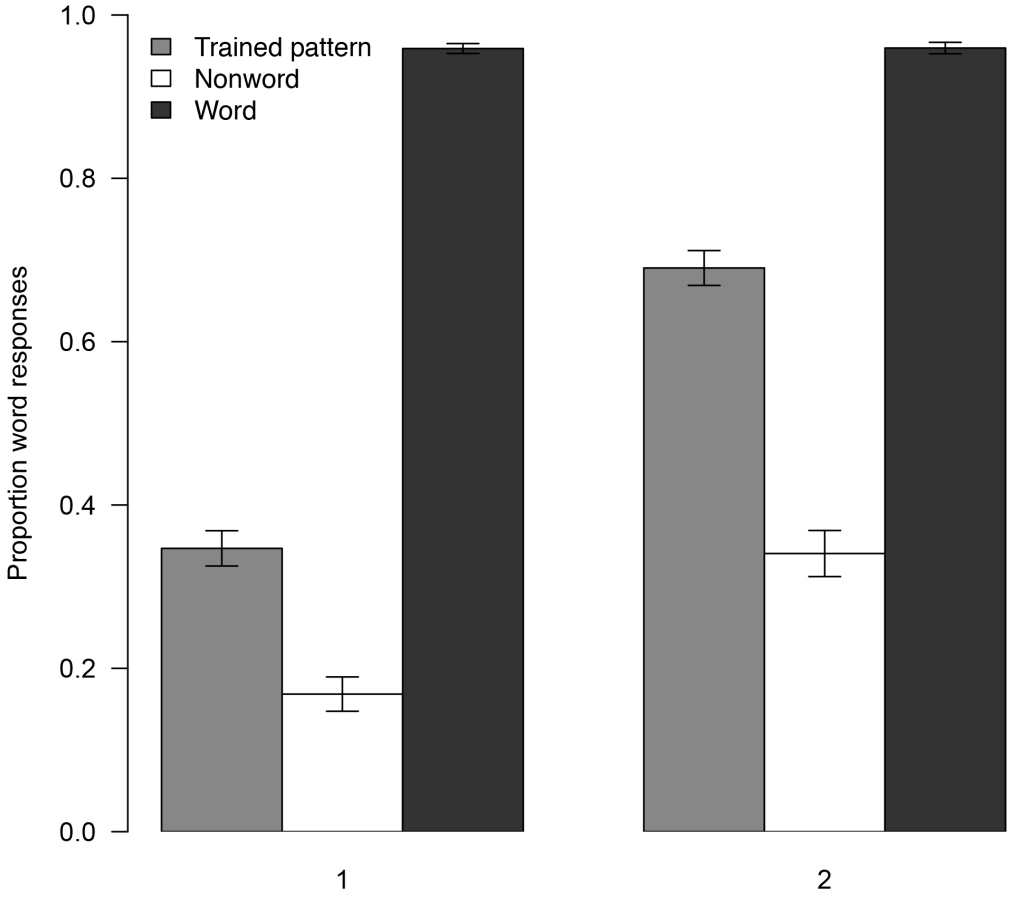


**Figure 3.1 Mean proportion of word responses to trained pattern, nonword and word items in Probe Block 1 and Probe Block 2. Error bars denote +/- 1 standard error.**

*3.2 Lexical Decision Test task*

*3.2.1 Trained and Untrained Accent Pattern*

The proportion of word responses was calculated for word, nonword, and NSAE-accented (trained and untrained patterns) items (Figure 3.2). Endorsement rates for nonword and NSAE-accented items produced by the trained talker were submitted to an LMER model containing contrast-coded fixed effects for Training (Control vs. Trained groups), Talker Variability (Single, Multiple), and Feedback (Lexical, Semantic Context), and Helmert contrast-coded fixed effects for Item Type (A: Nonword vs. Trained pattern + Untrained pattern; B: Trained pattern vs. Untrained pattern) along with their interactions. To make additional comparisons within Item Type, an additional model was run where Item Type was coded as Nonword + Untrained pattern vs. Trained pattern and Nonword vs. Untrained pattern. As such, the critical *p* value was set to 0.025. Random intercepts were specified for participants and items. Random slopes for Item Type by participant and Training, Talker Variability and Feedback by item were also included. A significant main effect of Training was found ($\beta$=2.54, SE $\beta$=0.49, $\chi^2(1)$=25.239, $p<0.001$), indicating that trained participants responded "Word" more frequently than control participants across item types. Significant effects of Item Type, both for Nonword vs. Trained pattern + Untrained pattern ($\beta$=1.53, SE $\beta$=0.30, $\chi^2(1)$=24.21, $p<0.001$) and Trained pattern vs. Untrained pattern ($\beta$=1.21, SE $\beta$=0.12, $\chi^2(1)$=5.5748, p=0.018), were also obtained, with more lexical endorsements to NSAE-accented items relative to nonwords and more endorsements for trained pattern items relative to untrained pattern items. Critically, Training x Item Type interactions were also significant, as participants who underwent training had higher lexical endorsement rates for NSAE-accented items (trained pattern + untrained pattern) over nonword items ($\beta$=1.45, SE $\beta$=0.62,

$\chi^2(1)=4.99$, p=0.025) and trained pattern items over untrained pattern items (β=1.59, SE β=0.52, $\chi^2(1)=8.54$, p=0.003) relative to the control group.

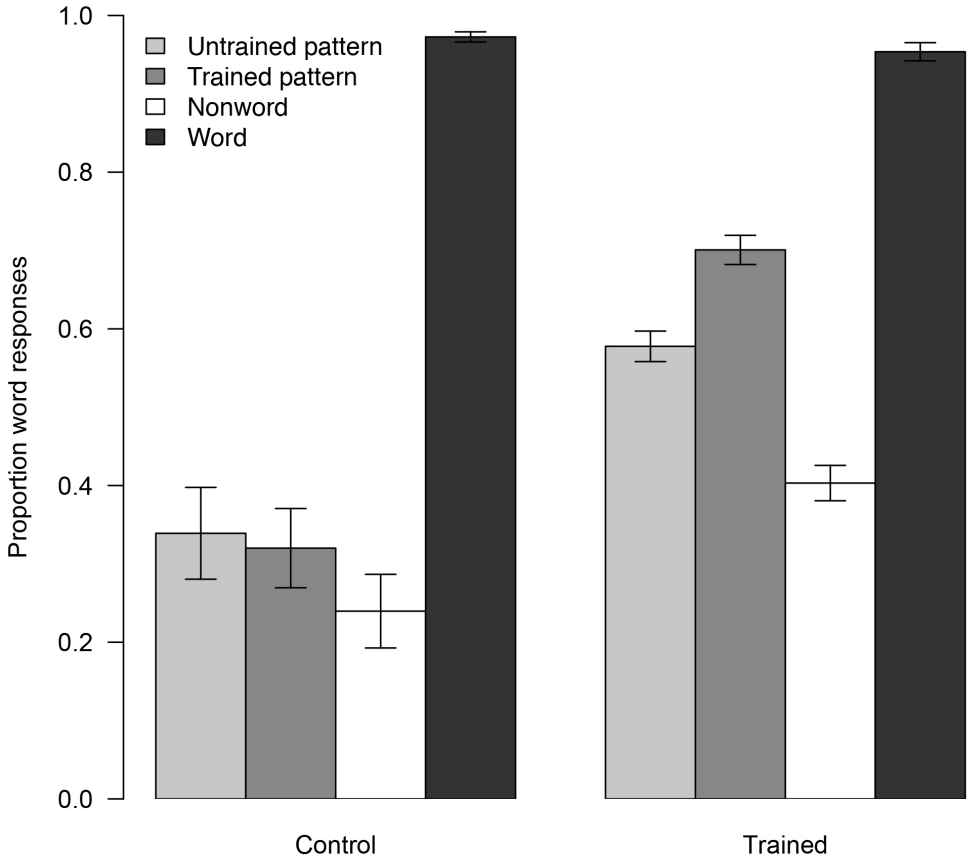


**Figure 3.2 Mean proportion of word responses by Item Type for control and trained listeners**

A significant Training x Item Type (Nonword + Untrained Pattern vs. Trained Pattern) interaction was also found (β=2.33, SE β=0.61, $\chi^2(1)=13.64$, p=0002); however, the Training x Item Type (Nonword vs. Untrained Pattern) was not significant ($\chi^2=2.75$, p=0.6). This indicates that while trained listeners saw significantly higher lexical endorsements for trained pattern items over nonwords relative to control listeners, endorsement rates for

nonwords and untrained pattern items did not significantly differ between trained and control groups. There was, however, a numerical trend for trained listeners to have higher endorsement rates for untrained pattern items than nonwords. Main effects of Talker Variability and Feedback and their interactions with Item Type did not reach significance ($\chi^2<1.60$, $p>0.21$).

*3.2.2 Trained and Untrained Talker*

To compare the influence of training type on participants' ability to generalize their knowledge of the NSAE accent to an untrained talker (Figure 3.3), an additional LMER model was constructed with lexical decisions to nonword and trained pattern items for both trained and untrained talkers as the dependent variable. Contrast-coded fixed effects of Training (Control vs. Trained groups), Talker Variability (Single, Multiple), Feedback (Lexical, Semantic Context) were included along with Helmert contrast-coded effects of Item Type (A: Nonword vs. Trained Pattern; Trained Pattern-Trained talker vs. Trained Pattern-Untrained talker). The same random effects structure as the previous model was implemented. Main effects of Training, ($\beta=2.83$, SE $\beta=0.48$, $\chi^2(1)=31.13$, $p<0.001$), Nonword vs. Trained pattern items ($\beta=2.91$, SE $\beta=0.45$, $\chi^2(1)=34.86$, $p<0.001$) and Trained talker vs. Untrained talker ($\beta=0.20$, SE $\beta=0.07$, $\chi^2(1)=7.78$, $p=0.005$) were obtained. A significant Training x Nonword vs. Trained Pattern interaction was found ($\beta=3.41$, SE $\beta=0.92$, $\chi^2(1)=12.57$, $p=0.0004$); however, the Training x Trained talker vs. Untrained talker interaction was not significant ($\chi^2=0.0099$, $p=0.9206$). All remaining main effects and interactions did not reach significance ($\chi^2<1.9$, $p>0.17$).

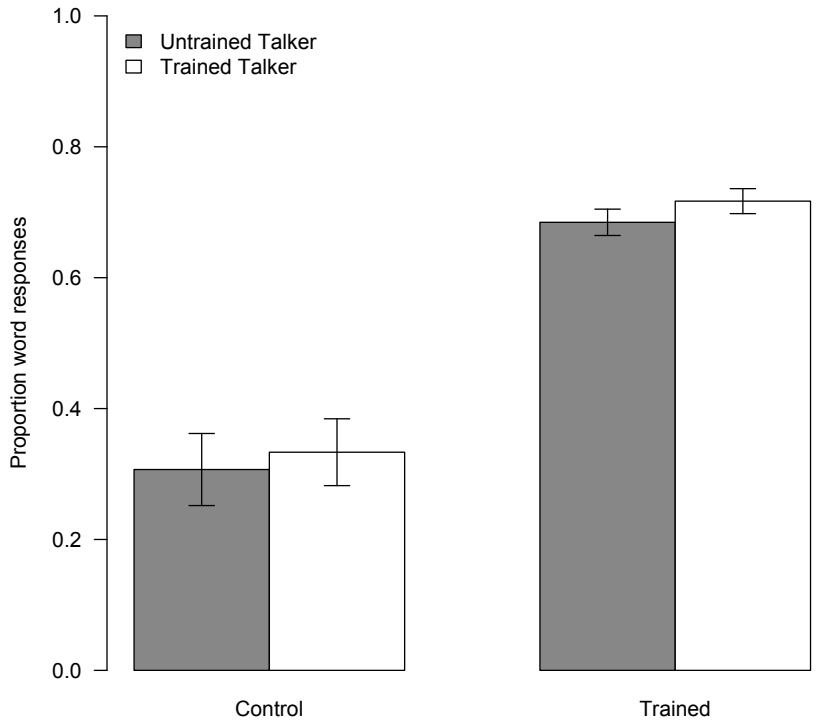


**Figure 3.3 Proportion of word responses to trained pattern items produced by untrained and trained talkers for control and trained listeners**

*3.2.3 Phoneme Type*

Finally, to examine the impact of phoneme type (vowel vs. consonant; Figure 3.4), an LMER model with responses to trained and untrained pattern items for both trained and untrained talkers was constructed. Fixed effects for Training (Control, Trained), Item Type (Trained vs. Untrained Pattern), Talker (Trained vs. Untrained), and Phoneme (Vowel vs. Consonant) were included along with their interactions. Random intercepts for participant and item were implemented, as well as by-participant random slopes for Item Type, Talker and Phoneme and by-item random slopes for Training. No significant main effect of Phoneme Type or interactions involving Phoneme Type were found ($\chi^2<2.38$, $p>0.12$),

indicating that phoneme type (vowel vs. consonant) did not play a significant role in listeners' adaptation to the novel accent.

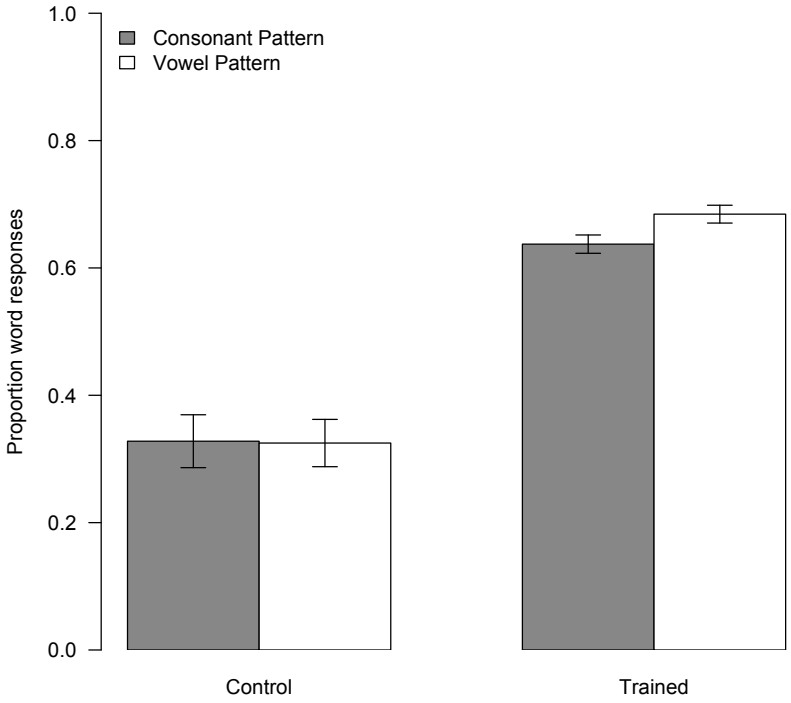


**Figure 3.4 Proportion of word responses to trained and untrained pattern items by the phoneme type of accent pattern (vowel vs. consonant) for control and trained listeners**

*3.3 Word Identification task*

Listeners provided responses to two types of stimuli, minimal pair change (where the item would be considered a word in a SAE accent and a different word in NSAE) and lexicality change (where the item would be considered a nonword in SAE but a word in NSAE). Responses to minimal pair items were coded in two ways: 1) identification accuracy in NSAE, termed NSAE Accuracy (listeners' accuracy identifying the item based on their knowledge of NSAE), and 2) identification accuracy in a SAE accent, termed SAE Accuracy

(listeners' accuracy identifying the item based on SAE). For example, the word "herd" [hɚ-d] would be pronounced "hurt" [hɚ-t] in NSAE. If listeners identified the item as "herd", it was scored as accurate by NSAE Accuracy (and inaccurate by SAE Accuracy). Conversely, if identified as "hurt", it was scored as accurate by SAE Accuracy (and inaccurate by NSAE Accuracy). All other responses were termed "other" and also tabulated.

For lexicality change items, identification accuracy in NSAE (NSAE Accuracy) was also calculated. For instance, the word "bleak" would be pronounced "blick" in NSAE, and thus, listeners were considered accurate by NSAE Accuracy if they were to transcribe it as such. The number of nonword responses (denoted by an 'X' by participants) was determined for both minimal pair and lexicality change items; however, the results here focus on the lexicality change data (as the proportion of minimal pair items identified as nonwords was very low).

*3.3.1 Trained vs. Untrained Accent Pattern*

Figure 3.5 depicts the proportions by response type (NSAE Accuracy, other, nonword responses) for lexicality change items produced by the trained. An LMER model was constructed with NSAE Accuracy for lexicality change items produced by the trained talker as the dependent variable. The same fixed effects for Training and Feedback conditions were implemented as in prior models, along with Item Type (Trained vs. Untrained Pattern) and all 2- and 3-way interactions. A significant main effect of Training ($\beta$=2.26, SE $\beta$=0.47, $\chi^2(1)$=21.74, $p<0.001$) as well as a Training x Item Type interaction ($\beta$=2.33, SE $\beta$=0.61, $\chi^2(1)$=11.504, p=0.0007) were found. All other effects and interactions were not significant ($\chi^2<0.64$, $p>0.42$). To investigate this 2-way interaction, separate LMER models for control

and trained listeners with Item Type as a fixed effect revealed that control listeners identified untrained pattern items as being words more often than trained pattern items (β=-1.66, SE β=0.70, $\chi^2(1)= 5.72$, p=0.017), whereas trained listeners NSAE Accuracy patterns did not differ as a function of Item Type ($\chi^2$=0.01, p=0.92). Moreover, subsequent models for each item type (Trained Pattern, Untrained Pattern) with Training as a fixed effect found that trained listeners identified items according to the NSAE accent significantly more than control listeners for both trained and untrained accent patterns ($\chi^2$>6.3, p<0.01).

For nonword responses, Training was found to be a significant factor (β=-1.8, SE β=0.5, $\chi^2(1)$=12.13, p<0.001). There was also a significant Training x Item Type interaction (β=-1.17, SE β=0.4, $\chi^2(1)$=8.15, p=0.004). Subsequent LMER models revealed no significant differences in Item Type (Trained Pattern, Untrained Pattern) for both control listeners and trained listeners ($\chi^2$<1.85, p>0.17). However, trained listeners had significantly fewer nonword responses to trained and untrained accent patterns relative to control listeners ($\chi^2$>5.73, p<0.017). This interaction arises from the magnitude of this difference being larger for trained patterns. These results indicate that trained listeners not only considered nonwords to be possible words in NSAE (lexicality change items) but also transcribed them according to the accent patterns to which they were exposed significantly more than control listeners.

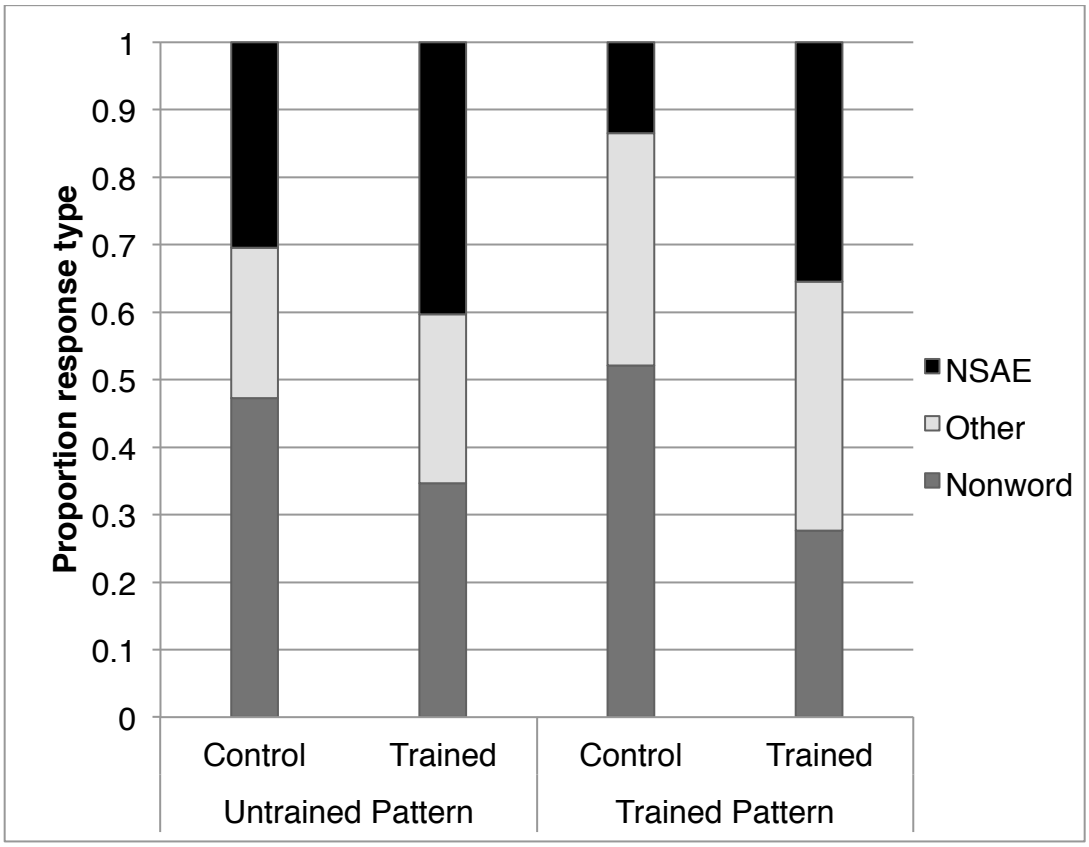


**Figure 3.5 Proportion of responses that were 1) accurate based on NSAE accent, 2) other responses, or 3) nonword responses for lexicality change items produced by a trained talker by group (Control, Trained) and Item Type (Trained pattern, Untrained Pattern).**

An identical model with NSAE Accuracy as the dependent variable (Figure 3.6) was constructed for the minimal pair items, which similarly yielded a significant effect of Training ($\beta$=2.5, SE $\beta$=0.84, $\chi^2(1)$=11.16, $p<0.001$). However, all other effects and interactions were not significant ($\chi^2<2.2$, $p>0.14$). The same model structure was applied to the SAE Accuracy data. A significant main effect of Training was found ($\beta$=2.5, SE $\beta$=0.84, $\chi^2(1)$=10.88, $p<0.001$), as control listeners identified a larger proportion of items according to SAE (i.e., consistent with their surface form) relative to trained listeners. A marginally significant Talker Variability x Feedback interaction was also obtained ($\beta$=-0.99, SE $\beta$=0.51, $\chi^2(1)$=3.78, $p=0.052$), stemming from the MT-Semantic condition obtaining higher SAE

Accuracy than the MT-Lexical condition ($\chi^2$=3.78, p=0.052). No other effects or interactions were significant ($\chi^2$<1.47, p>0.23).

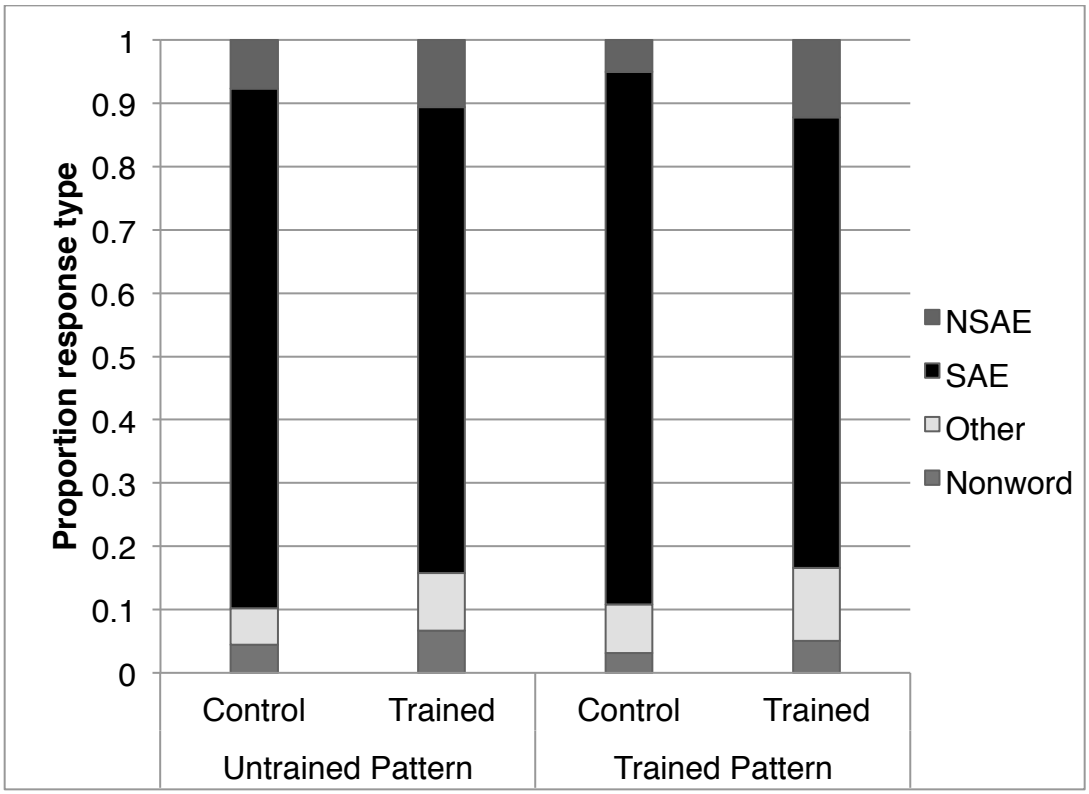


**Figure 3.6 Proportion of responses that were 1) accurate based on NSAE accent, 2) accurate based on SAE accent , 3) other responses, or 4) nonword responses for minimal pair change items produced by a trained talker by group (Control, Trained) and Item Type (Trained pattern, Untrained Pattern).**

*3.3.2 Trained and Untrained Talker*

To examine whether talker had a significant influence on word identification accuracy, an LMER model was constructed with NSAE Accuracy as the dependent variable for lexicality change items (trained accent pattern) produced by both the trained and untrained talkers (Figure 3.7). Fixed effects of Training and Feedback conditions were included, as in prior models, along with a fixed effect for Talker (Trained, Untrained) and their interactions. Random intercepts for participant and item were included. A significant

main effect of Training (β=2.96, SE β=0.49, $\chi^2(1)=33.58$, $p<0.001$) was found, along with a significant effect of Talker (β=-0.2, SE β=0.09, $\chi^2(1)=5.71$, $p=0.017$). No other significant effects or interactions were obtained ($\chi^2<3.3$, $p>0.07$).

For nonword responses, only a significant effect of Training was obtained (β=-2.5, SE β=0.5, $\chi^2(1)=22.46$, $p<0.001$), with a greater proportion of nonword responses by control relative to trained listeners. All other effects and interactions were not significant ($\chi^2<2.07$, $p>0.15$).

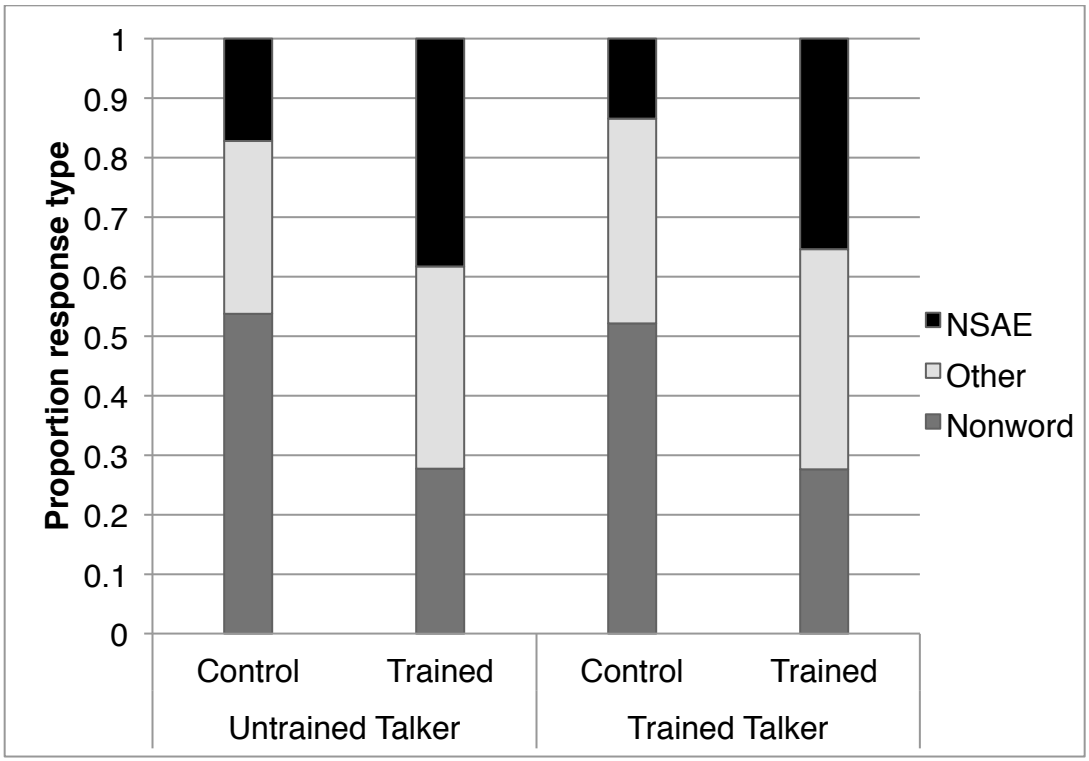


**Figure 3.7 Proportion of responses that were 1) accurate based on NSAE accent, 2) other responses, or 3) nonword responses for lexicality change items (trained pattern) by group (Control, Trained) and Talker (Trained, Untrained).**

An identical model was implemented for the minimal pair data (Figure 3.8), which similarly yielded significant effects of Training (β=3.08, SE β=0.73, $\chi^2(1)=20.17$, $p<0.001$)

and Talker (β=0.67, SE β=0.17, $\chi^2(1)=19.44$, $p<0.001$). The Training x Talker interaction was also significant (β=-2.76, SE β=1.06, $\chi^2(1)=9.002$, $p=0.003$). Follow-up LMER models revealed that both control and trained listeners had higher NSAE Accuracy for the trained talker relative to the untrained talker ($\chi^2>4.69$, $p<0.03$), with a larger trained vs. untrained talker difference for control relative to trained listeners. Critically, however, trained listeners demonstrated higher NSAE accuracy than control listeners for items produced by both the trained and untrained talkers ($\chi^2>14.48$, $p<0.001$). Additionally, a significant Talker x Talker Variability interaction was found (β=-0.71, SE β=0.27, $\chi^2(1)=7.13$, $p=0.008$). Subsequent LMER models investigating the 2-way interaction, fixing each variability condition, revealed no significant differences in NSAE Accuracy as a function of talker for the multi-talker conditions ($\chi^2=0.002$, $p=0.96$). However, for the single-talker conditions, a significant difference emerged based on talker (β=0.74, SE β=0.22, $\chi^2(1)=10.62$, $p=0.001$). NSAE Accuracy was significantly higher for the trained talker (*M*=13%) relative to the untrained talker (*M*=8%).

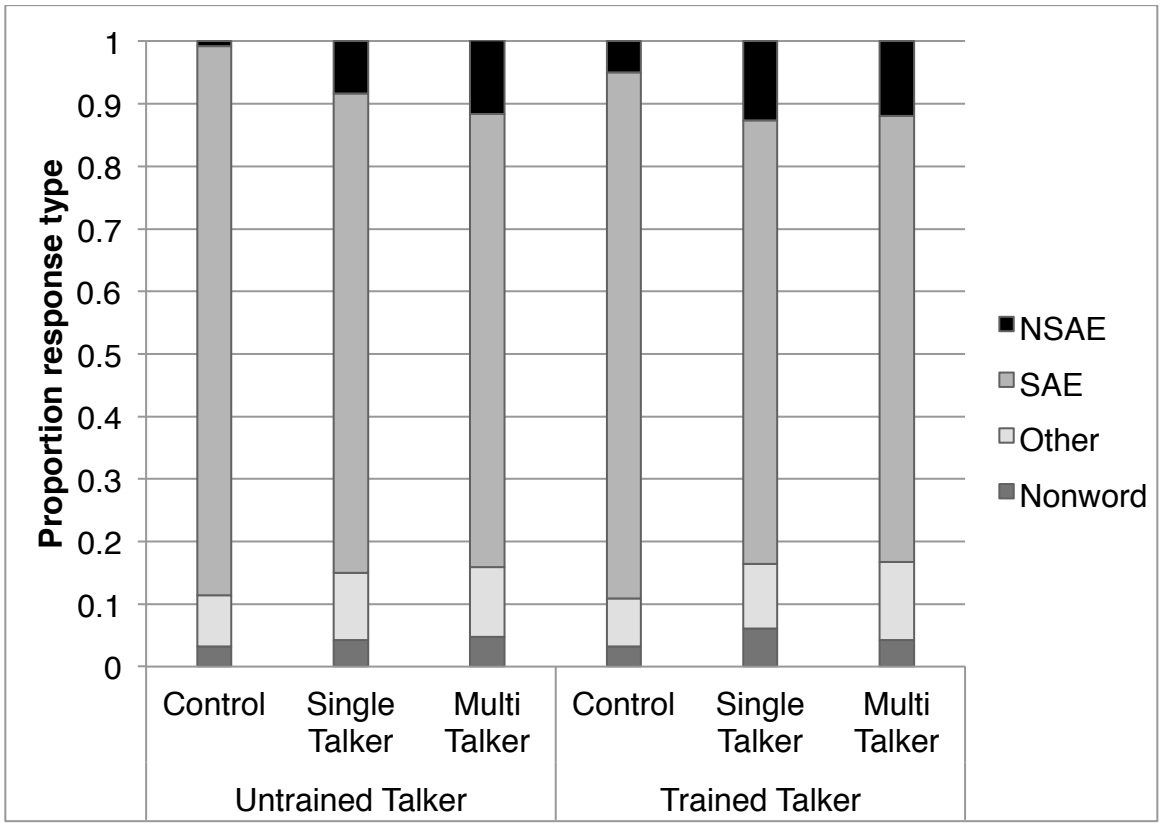


**Figure 3.8 Proportion of responses that were 1) accurate based on NSAE accent, 2) accurate based on SAE accent, 3) other responses, or 4) nonword responses for minimal pair change items (trained pattern) by group (Control, Trained) and Talker (Trained, Untrained).**

The same model was constructed with SAE Accuracy, which similarly yielded significant main effects of Training ($\beta$=-1.73, SE $\beta$=0.5, $\chi^2(1)$=11.59, p<0.001) and Talker ($\beta$=-0.29, SE $\beta$=0.09, $\chi^2(1)$=10.4, p=0.001), reflecting higher SAE Accuracy scores by control relative to trained listeners and higher scores for the untrained talker relative to the trained talker. The other effects and interactions did not reach significance ($\chi^2$<2.97, p>0.08).

*3.3.3 Phoneme Type*

LMER models were constructed with fixed effects of Training (Control, Trained), Item Type (Trained Pattern vs. Untrained Pattern), Talker (Trained, Untrained), and Phoneme (Vowel vs. Consonant; Figures 3.9 and 3.10). The only significant effect involving

Phoneme as a factor was a 4-way Training x Item Type x Talker x Phoneme interaction in models with NSAE Accuracy (lexicality change items; β=-3.78, SE β=1.2, $\chi^2(1)=9.98$, p=0.0016) and SAE Accuracy data (minimal pair items; β=-3.4, SE β=1.3, $\chi^2(1)=6.35$, p=0.004). For NSAE Accuracy, follow-up LMERs ultimately revealed that control listeners had higher NSAE Accuracy for items produced by the trained talker with trained consonantal patterns relative to trained vowel patterns ($\chi^2=4.28$, p=0.04). This difference did not reach significance for trained listeners ($\chi^2=1.77$, p=0.183).

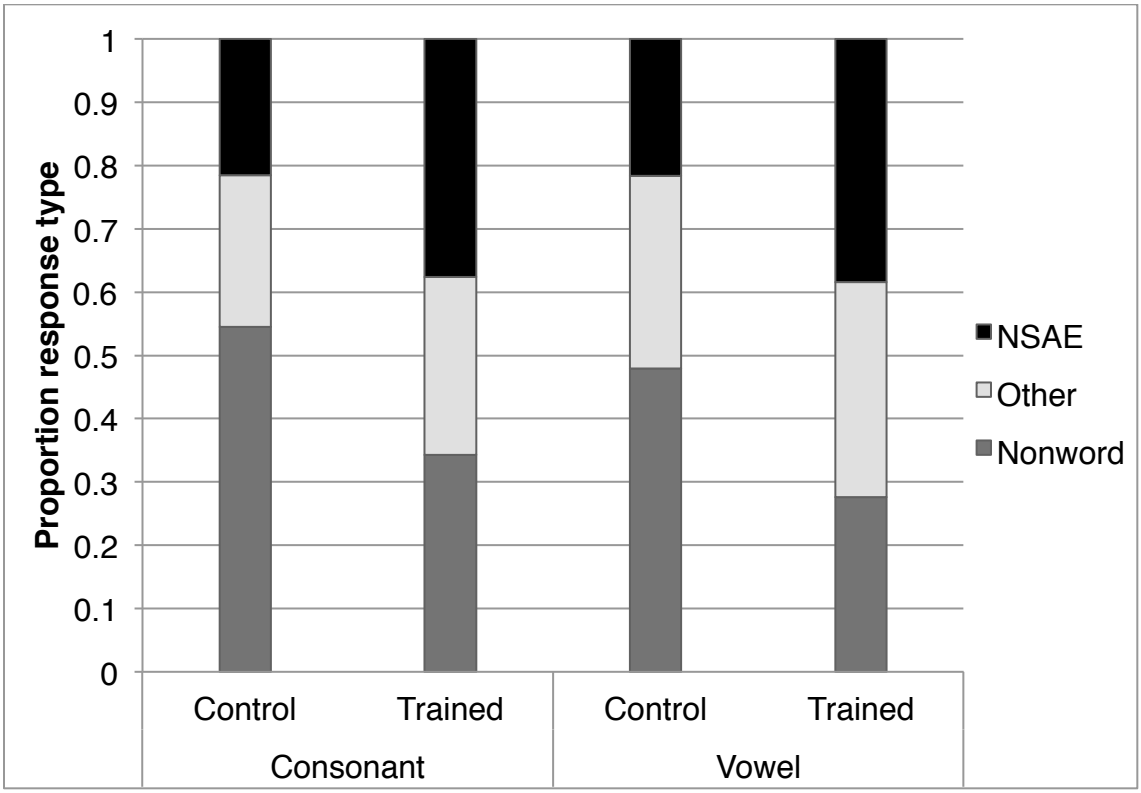


**Figure 3.9 Proportion of responses that were 1) accurate based on NSAE accent, 2) accurate based on SAE accent, 3) other responses, or 4) nonword responses for minimal pair change items by group (Control, Trained) and Phoneme Type (Consonant, Vowel).**

For SAE Accuracy (minimal pair items), subsequent LMERs indicated that both control ($\chi^2=8.11$, p=0.004) and trained ($\chi^2=8.09$, p=0.004) participants were more likely to

transcribe items produced by the trained talker based on SAE when they contained trained consonantal contrasts rather than trained vowel contrasts. The difference in SAE Accuracy between items containing consonantal versus vowel patterns was significantly larger for control (consonant patterns: 93%, vowel patterns: 75%) relative to trained participants (consonant patterns: 79%, vowel patterns: 63%).

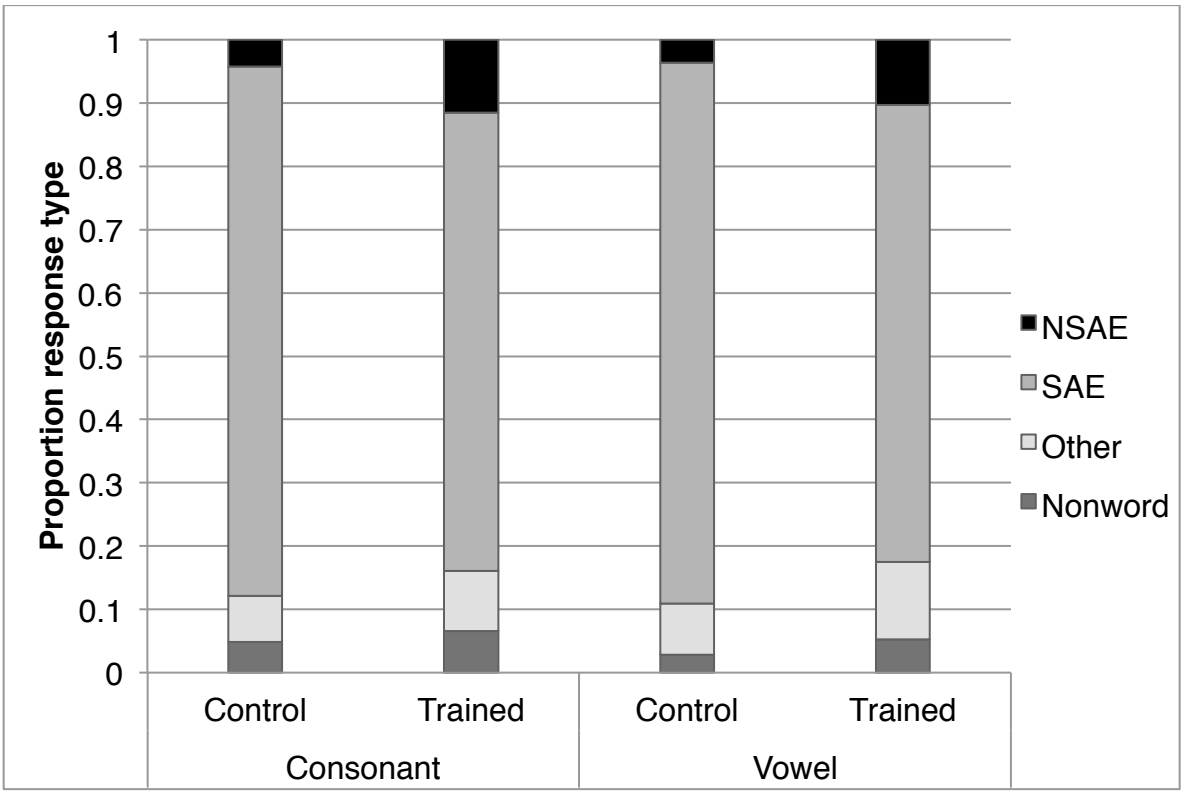


**Figure 3.10 Proportion of responses that were 1) accurate based on NSAE accent, 2) other responses, or 3) nonword responses for lexicality change items by group (Control, Trained) and Phoneme Type (Consonant, Vowel).**

*3.4 Summary*

The results of this experiment revealed that, in the probe task, trained listeners demonstrated increased lexical endorsement rates to both trained pattern and nonword items from Block 1 (prior to training) to Block 2. Crucially, however, there was a larger increase in word responses to trained pattern items relative to nonwords, indicating that after the first

training phase, listeners were learning the trained accent patterns and applying them to items containing the appropriate segments. Listeners from single- and multi-talker conditions receiving either lexical or semantic contextual feedback achieved similar levels of learning.

Following training, there was an overall higher rate of lexical endorsements in the lexical decision task for trained listeners relative to control listeners who had not received any training, as evidenced by a higher proportion of word responses to nonword items. However, there was also a significantly higher endorsement rate to trained pattern items relative to untrained pattern and nonword items as compared to control listeners, which was not modulated by phoneme type (vowel vs. consonant) or talker (trained vs. untrained). Trained listeners did not appear to generalize their learning to untrained accent patterns, as the difference in endorsement rates to untrained pattern and nonword items was not significantly different from control listeners, though there was a numerical trend towards higher endorsement rates for untrained pattern relative to nonword items. Similar to the probe task, all four trained conditions (ST-Lexical, ST-Semantic context, MT-Lexical, MT-Semantic context) demonstrated comparable endorsement rate patterns.

Finally, in the word identification task, for items that would be considered nonwords in SAE (lexicality change items), trained listeners identified them as nonwords significantly less often than control listeners and less often when the item contained a trained accent pattern relative to an untrained one. Moreover, when a word response was provided, trained listeners correctly identified these items based on what they had learned about the NSAE accent significantly more than control listeners. No interaction was found with Item Type or Phoneme, indicating that trained listeners were more accurate for both trained and untrained

accent patterns, regardless of whether they involved vowels or consonants, as compared to control listeners. For items that were actually real words in SAE (minimal pair items), trained listeners identified these items as different words, accurately based on NSAE, significantly more than control listeners, regardless of whether the items had trained or untrained accent patterns, vowels or consonants. There was also evidence that Talker variability was found to interact with Talker for these minimal pair items, with single talker conditions demonstrating higher NSAE Accuracy for the trained talker relative to the untrained talker, with no difference in Talker found for multi-talker conditions.

**4. Discussion**

The present work demonstrated systematic retuning of vowel and consonant categories as a product of exposure to NSAE-accented speech in conjunction with disambiguating linguistic information. We initially expected that the predictive strength of the feedback provided during training, that is the ability of the feedback to narrow the space of possible options and reduce uncertainty as to how the acoustic observations should be categorized, would modulate the efficacy of adaptation. As such, the current study provided two different types of feedback to trained listeners, either the exact lexical item (which would be 100% predictive) or a moderately-predictive semantic context for that item; however, no significant difference between feedback types was found. Indeed, both lexical and semantic context conditions demonstrated significant perceptual learning relative to the control condition. This is consistent with prior work providing sentence feedback that either lexically-matched or mismatched the NSAE-accented target and where both types of feedback were found to be equally facilitative during adaptation (Chapter 2). This provides

further evidence that, at least for native listeners, the perceptual system is remarkably flexible, able to draw upon varied types of higher-level disambiguating information equally effectively to adapt to the systematic phonetic deviations from native-accented norms present in foreign-accented speech. Lexical information (that completely disambiguates the item) or semantic context (that serves to narrow the possible options for that item) both provided sufficient information for the perceptual system to update listeners' beliefs about the distributions of the relevant categories for the accent.

In addition to manipulating the content of the feedback, the present work also manipulated talker variability during the training phase, with listeners being trained with either a single talker or four different talkers. Talker variability during training did impact generalization to a novel talker in the word identification task for what is arguably the more challenging item type (minimal pair change), where listeners heard a real word and could identify it based on their knowledge of SAE or NSAE. Listeners who had been trained on multiple talkers were more likely to identify the real word productions of a novel talker based on NSAE relative to listeners who had received single-talker training. Consistent with prior research (Bradlow & Bent, 2008; Sidaras et al., 2009), listeners in the multi-talker conditions appeared to have abstracted over the multiple talkers they were exposed to during training to make more generalized updates to their beliefs, building a set of beliefs about a generative model for the accent rather than a specific talker and allowing them to deploy their knowledge of NSAE to a novel talker in this context.

In the lexical decision task and for lexicality change items in the word identification task, listeners from both single and multi-talker conditions saw significant adaptation to both

trained and untrained talkers relative to control listeners. The role of acoustic similarity between trained and untrained talkers has been implicated in prior work as a significant contributing factor in cross-talker generalization (e.g., Reinisch & Holt, 2014), so it could be the case that the two male talkers in the current experiment were sufficiently acoustically similar to facilitate generalization, even for single-talker conditions, in these relatively easier conditions. However, recent work revealed robust generalization after single talker exposure, with generalization of learning to an acoustically dissimilar untrained talker (Weatherholtz, 2015). Given that, the present findings seem to reflect a highly flexible perceptual system, relatively tolerant of individual talker variation when adapting to substantive accent deviations, particularly when listeners are explicitly aware of the fact that both trained and untrained talkers have the same accent. As such, listeners likely formed beliefs about the speech statistics of the particular context to which they were exposed (in this case, a novel accent). When instructed that the untrained talker produced the same accent that they were exposed to during training, they determined that this recent experience and their beliefs about the generative model for the trained talker would be most relevant (rather than their prior experience with the speech statistics accrued from numerous SAE talkers) and employed that knowledge accordingly (Kleinschmidt & Jaeger, 2015).

Moreover, the present findings revealed that exposure to NSAE, containing an array of both vowel and consonantal deviations from SAE, yielded an overall increase in listeners' willingness to accept nonword forms as being possible English words, as indicated by the significantly higher lexical endorsement rate of nonword items (not containing trained or untrained accent patterns) for trained as compared to control listeners. This could have arisen

as a result of a general relaxing of criteria for what counts as an acceptable match between stored lexical representations and the incoming speech input, which has been found in cases where listeners find themselves in more adverse listening conditions, such as noise or reduced speech (Brouwer, Mitterer, & Huettig, 2012; McQueen & Huettig, 2012). In this case, listeners consistently exposed to productions that deviated strongly from SAE norms increased their overall tolerance for mismatches between input and representation.

However, their mismatch tolerance only extended so far, as lexical endorsement rates and word identification NSAE accuracy were still significantly higher for trained and, in some cases, untrained pattern items relative to nonwords. It is important to note that the nonwords employed in the present work were minimally different from real words (in that they differed by a single phoneme), in the same way as the trained and untrained pattern items. Prior work has utilized "maximal nonwords" (Weatherholtz, 2015), which differed from real words on multiple segments and features. However, if items were distinctly nonword-like, listeners would be less likely to false-alarm and consider them to be nonwords, even if their criteria for word status has been somewhat relaxed. We would argue that the "minimal nonwords" used in this study provide a strong test of the system's ability to differentiate exposed or structurally-related accent patterns from a general relaxing of criteria for nonword items. The present findings indicate that while the perceptual system was generally increasing its tolerance for atypical speech input, it was still constraining this tolerance to focus on adjusting distributions for specific as well as structurally-related categories. For example, listeners adjusted not only their voiced alveolar fricative category to accommodate voiceless alveolar fricatives (e.g., such that [saɪs] would be identified as

"size") but also their labiodental voiced fricative, to which they were not exposed (e.g., [foɪs] would be identified as "voice"). This is consistent with Kraljic and Samuel (2006), who found cross-contrast generalization (exposed to /d/-/t/ but generalized to /b/-/p/). Listeners in the present work did not, however, loosen their criteria to accept, to the same degree as the trained and untrained pattern items, minimal nonwords such as "spaish" [spaɪʃ] or "spum" [spʌm], which, if listeners had completely relaxed their lexical criteria, could have potentially been recognized as "spice" or "spun", respectively.

NSAE-accented exposure actually introduced a degree of ambiguity into word recognition, as trained listeners began to not only activate the lexical items (as pronounced) but also their possible NSAE variants, as evidenced by their identifying minimal pair items as different words than pronounced (e.g., the item /pɑt/ identified as "pod" following training on the NSAE accent which contains final devoicing). Interestingly, this resembles non-native listeners, who experience this ambiguity as a product of perceptual difficulties with second language phonemes (Cutler & Broersma, 2005; Weber, Broersma, & Aoyagi, 2011). In the case of native listeners, ambiguity can arise as a product of sound category boundaries shifting or expanding to accommodate atypical exemplars. In the context of the ideal adapter framework, listeners who had not received NSAE-accented exposure would possess prior beliefs about category distributions based on their experience with SAE-accented English speakers that would have led them to, for example, strongly activate the word "pot" and send relatively less activation to "pod" when presented with the production /pɑt/. However, trained listeners, as a result of bottom-up exposure and top-down information (lexical or semantic contextual feedback), will have updated their beliefs about the distributions of the

relevant categories when encountering this accent, which in this example actually collapses two categories, and will thus receive strong activation from both "pot" as well as "pod". This increased activation of multiple lexical possibilities does introduce ambiguity that is potentially detrimental to individual word recognition (i.e., homophony). However, in more natural communicative contexts, given that native listeners are proficient at drawing upon higher-level contextual information to facilitate lexical access, the benefit gained from having constructed an accent-specific model of cue distributions as a result of accent exposure should outweigh this seeming disadvantage.

In sum, the current study demonstrated the remarkable flexibility of the perceptual system when confronted with speech deviating from native-accented norms. Lexical and semantic contextual information were found to be equally informative in updating listeners' beliefs about relevant category distributions and generating predictions about future speech input. An influence of talker variability during exposure manifested on one of the more stringent tests (minimal pair items in word identification), with multi-talker exposure enabling listeners to extract systematic commonalities across talkers and update their beliefs about the category distributions of talkers from this accent more generally. Other test conditions revealed robust learning not only for the trained talker but also for an untrained talker, regardless of whether listeners experienced talker variability during training, which may have arisen as a result of providing explicit information about the talker's membership in the exposed accent group. Moreover, such information was utilized to update beliefs not only about category-specific distributions but also to make inferences about structurally-related category distributions. Critically, such performance cannot be completely accounted

for by an overall relaxing of criteria for lexical status, as evidenced by higher endorsement rates for trained pattern items relative to nonwords.

The observed flexibility in the perceptual system stemmed from listeners efficiently updating their beliefs about the contextually-appropriate generative linguistic model. What drives this belief updating? The present findings suggest that when the system encounters a mismatch between predicted and surface input, it requires some form of constraining information with which to evaluate the acoustic observations. While we initially anticipated that varying the predictive strength or the certainty with which listeners could label the incoming input, both lexical and semantic contextual feedback in the current experiment disambiguated the input sufficiently for listeners to be able to label the acoustic observations with relatively high certainty. Information that is substantially lower in its predictive strength may be necessary for its impact on the system to manifest behaviourally. This constraining information was coupled with listeners' explicit knowledge of the similarity between trained and untrained talkers (with respect to their being from the same accent group), illustrated by talker-general learning, as well as their implicit knowledge of phonetic structural similarity, shown by their generalization to novel accent patterns. Just as listeners are posited to form groups of talker-specific generative models based on some combination of top-down knowledge (e.g., awareness of talkers sharing the same language background) and bottom-up, phonetic information (Kleinschmidt & Jaeger, 2015), they likely also make inferences about how individual distributions for specific sound categories group together based on their structural similarity (e.g., sharing place or manner of articulation). Upon recognition of shared phonetic features (for a set of accent patterns) or shared membership within an accent

group (for a set of talkers), listeners can then deploy the corresponding accent-specific prior beliefs to the novel situation. This might then predict less robust learning and generalization under conditions of enhanced uncertainty, such as if the constraining information was not consistent or reliable or if listeners encountered accent patterns that were structurally dissimilar from or incongruent with anything to which they had previously been exposed (e.g., the production of /ɹ/ as /l/ or an instance of vowel fronting rather than lowering). It remains for future work to investigate how much uncertainty the perceptual system is willing accommodate before opting to remain stable; however, the present study demonstrates that the perceptual system will leverage reliable, constraining linguistic knowledge and signal-based information to adapt to pronunciation variation by familiar and unfamiliar talkers and to familiar and unfamiliar accent patterns.

# CHAPTER 4: PERCEPTUAL LEARNING OF NOVEL ACCENTED SPEECH BY SECOND LANGUAGE LISTENERS

## 1. Introduction

### *1.1 Native language speech perception*

Native (L1) listeners of a language possess a remarkably flexible perceptual system that enables them to extract information in adverse listening conditions, to identify different talkers with a high degree of accuracy, and to adapt to variability that arises as a function of differences in talker or accent (Cutler, 2012). Adaptation to talker-specific characteristics such as a foreign accent occurs rapidly, within a few minutes of exposure (Clarke & Garrett, 2004) and can subsist for several days without any intervening experience (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). These adaptation processes draw upon pre-existing linguistic knowledge or information extant in the signal to interpret the atypical sounds, adjusting phoneme category boundaries as needed to accommodate these new or unusual exemplars (e.g., Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2003; Samuel & Kraljic, 2009). L1 listeners have been found to effectively leverage a variety of different types of disambiguating information, including phonemic (Hervais-Adelman et al., 2008), lexical (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2007; Zhang & Samuel, 2014), phonotactic (Cutler et al., 2008) and visual cues (e.g., Bertelson, Vroomen, & De Gelder, 2003).

A large body of research examining perceptual adaptation for L1 listeners has utilized a category re-tuning paradigm, whereby listeners are first exposed to a sound ambiguous between two categories (e.g., /s/-/f/) in contexts that bias them to perceive the sound as one category or the other (e.g., Norris et al., 2003). Following exposure, a phonetic categorization

task on a continuum of the two relevant categories reveals that listeners will provide more responses to the sound category to which they were biased during exposure, indicating that listeners will adjust their category boundary in response to disambiguating information when exposed to ambiguous sounds. Following exposure, L1 listeners have been found to be able to generalize their knowledge to novel lexical items, indicating a sub-lexical locus for perceptual adjustments, novel (structurally-related) contrasts (Kraljic & Samuel, 2006) and novel talkers (e.g., van der Zande, Jesse, & Cutler, 2014). In addition to achieving a degree of plasticity that enables listeners to accommodate new and unfamiliar speech information, the perceptual system also must sustain stability in order to maintain existing categories and representations. For example, the system resists adapting to pronunciation variants if their atypicality can be attributed to an external source (e.g., a pen in the speaker's mouth; Kraljic, Samuel, & Brennan, 2008) or to phonetically-conditioned variation (Kraljic, Brennan, & Samuel, 2008). Such findings illustrate the fine balance between flexibility and stability within the native perceptual system that allows for relatively efficient and accurate spoken word recognition to take place.

*1.2 Non-native speech perception*

Despite the precision and flexibility characteristic of native language listening, speech perception in one's second language can be a challenging task, arising from difficulties at multiple levels of linguistic processing. With respect to low-level perception, L2 listeners, at least initially, perceive second language speech through the lens of their established L1 phonemic system, which has been attuned to extract the most relevant and reliable dimensions of information to differentiate L1 contrasts (e.g., Iverson, Kuhl, Akahane-

Yamada, & Diesch, 2003; Strange, 2011). As such, native language categories have been shown to exert a strong influence on L2 speech perception (e.g., Best & Tyler, 2007; Flege, 1995). For example, when distinct L2 categories assimilate to a single category in the L1, as in the case of the English /ɹ/-/l/ contrast for Japanese listeners or the English /ɛ/-/æ/ contrast for Dutch listeners, acquisition of the L2 categories is markedly worse relative to cases where the contrasting L2 sounds assimilate to two distinct L1 categories.

As a result of inaccurate L2 phoneme perception arising from native language interference, L2 spoken word recognition can become problematic. Words such as *peck* and *pack* for Dutch listeners or *rice* and *lice* for Japanese listeners are often indistinguishable from each other. Moreover, L2 listeners have the added challenge of contending with more lexical competitors during word recognition than L1 listeners, as a product of the activation of "phantom" competitors (Broersma & Cutler, 2008). For instance, in a lexical decision task, near-word items such as "flide" or "shib" would be designated non-words by native English listeners but more often as real words by native Dutch listeners (perceiving them as *flight* and *ship*, respectively), as a result of voicing not being distinctive word-finally in Dutch. Indeed, in a cross-modal priming paradigm, Dutch listeners' perception of the targets were significantly facilitated both when the prime item matched (*flight*-FLIGHT) as well as when the prime item was a near-word (*flide*-FLIGHT), whereas this facilitation was only found in the matched condition for English listeners. This indicates that perceptual phonetic confusions and phantom activation can lead to an increase in the amount of lexical competition with which an L2 listener has to contend, and a greater degree of lexical

competition has been demonstrated to yield slower word recognition (e.g., Norris, McQueen, & Cutler, 1995).

Furthermore, the challenges of non-native perception also extend to listening in sub-optimal or adverse conditions (e.g., background noise). Mayo, Florentine, and Buus (1997) tested native, proficient early- and late-bilingual listeners in a sentence-final word recognition task in multi-talker babble noise, at varying signal-to-noise ratios (SNR), where the final word was either highly predictable or not from the preceding context. The signal-to-noise levels that early-bilingual and native listeners were able to tolerate were greater in high-predictability relative to low-predictability contexts; however, this difference did not emerge for late-bilinguals. These findings indicate that late-bilinguals were less able to leverage higher-level contextual information relative to early-bilinguals and native listeners. Bradlow and Alexander (2007) followed up on this work by revealing that second language listeners were capable of utilizing contextual information during word recognition as long as this information was clearly specified within the acoustic input. That is, a contextual benefit emerged for L2 listeners only when it was produced in a clear speaking style (relative to a plain/conversational style), highlighting an interaction between low-level acoustic and higher-level knowledge-driven information during L2 speech perception.

While the majority of prior work has focused on the perception of L2 speech produced by native speakers, a growing body of research has also examined how L2 listeners perceive and adapt to L2 speech produced by non-native speakers. Despite the disadvantages discussed above that listeners typically face when listening in their second language, there is evidence to suggest that L2 listeners can achieve comparable performance (or even surpass)

native listeners when listening to foreign-accented speech, particularly when the talker and listener share an L1 background (e.g., Bent & Bradlow, 2003; Imai, Walley, & Flege, 2005; van Wijngaarden, 2001; Xie & Fowler, 2013). Bent and Bradlow (2003) found evidence in support of an "interlanguage speech intelligibility benefit", whereby high-proficiency non-native talkers were found to be as intelligible as native talkers for non-native listeners who shared the same language background as the talkers. For native listeners, there was an asymmetry in intelligibility, with native talkers being more intelligible than non-native talkers. Foreign-accented speech arises from segmental and suprasegmental deviations from native-accented norms as a product of interactions between the native and second language sound systems. As such, knowledge or familiarity with the particular deviation patterns that result from a specific L1-L2 language pair may better equip listeners to interpret the speech produced by a speaker with this particular language background.

Consistent with this view, Imai et al. (2005) posited the "phonological mismatch hypothesis", asserting that the phonological representations developed by L2 listeners are not "optimally matched" to the L2 speech input produced by native speakers (p. 897). Ultimately they may be better matched to speech produced by L2 speakers that share the same language background, as the particular acoustic patterns produced by these speakers would more likely align with the listeners' own productions. Indeed, Spanish-English bilinguals were found to actually outperform native English listeners at identifying Spanish-accented English words (from dense lexical neighbourhoods).

Using a cross-modal priming paradigm, Weber, Broersma, and Aoyagi (2011) similarly found facilitated processing when the acoustic manifestation of the foreign-

accented prime item aligned with the accent of the listeners (Japanese vs. Dutch). For example, the item /ɛkt/ primed the English word *act* for Dutch listeners, whereas /ˈakto/ primed *act* for Japanese listeners (and did not for Dutch listeners). Moreover, Dutch-accented items also primed words for Japanese listeners; however, the authors noted that this might have arisen as a result of the perceptual confusability of the Dutch-accented and native English pronunciations for Japanese listeners. Hanulíková and Weber (2012) also reported the influence of linguistic experience on the perception of foreign-accented pronunciation variants. The eye movements of German and Dutch learners of English were tracked when perceiving words containing three variants (/s/, /f/, /t/) of the pronunciation of the English interdental fricative /θ/. Despite /f/ being the most perceptually confusable with /θ/, listeners' looking preferences aligned with the pronunciation variant most frequently produced by listeners of the different language groups (German-accented English is characterized by /s/-substitutions, while Dutch-accented English is marked by /t/-substitutions).

As noted above, one of the hallmarks of native listening is the ability to flexibly adapt to variable speech input (Cutler, 2012), and recent research has begun to examine whether and how perceptual adaptation works for L2 listeners (Mitterer & McQueen, 2009; Reinisch, Weber, & Mitterer, 2012; Schertz, Cho, Lotto, & Warner, 2015; Weber, Betta, & McQueen, 2014). Mitterer and McQueen (2009) reported that Dutch listeners exposed to Scottish or Australian English accented speech on television were able to leverage English language subtitles to improve their comprehension relative to control listeners who either received Dutch subtitles or no subtitles. Listeners who received English subtitles were significantly more accurate at repeating back phrases produced by regionally-accented

speakers as compared to the control groups. As these L2 listeners demonstrated an ability to utilize lexical (as well as audio-visual and other lower-level cues) to adapt to the unfamiliar English accent, this indicates that adaptation in a second language draws upon similar resources as first language adaptation.

Reinisch et al. (2012) investigated whether L2 adaptation can occur in a more narrow sense, that is, whether listeners can adjust a specific phoneme boundary (as demonstrated by native listeners in the studies discussed above). German learners of Dutch completed a similar category re-tuning paradigm as Norris et al. (2003), where either word-final /s/ or /f/ was replaced with an ambiguous sound during exposure. It is important to note that both German and Dutch languages employ these fricatives. German L2 listeners demonstrated comparable category boundary retuning at test as Dutch L1 listeners following exposure to Dutch speech input containing the ambiguous sound, such that listeners who were exposed to the ambiguous sound in /f/-final words provided more f-responses than those who heard the ambiguous sound in /s/-final words. Phoneme recalibration effects were found even for Dutch L1 listeners who were exposed to the ambiguous sound in Dutch-accented English but tested in Dutch, indicating that retuned category boundaries can transfer across languages, at least in cases where the talker is the same and the categories align relatively well across the L1 and L2 languages.

*1.3 Current research*

Recent work has posited a formal model to capture perceptual learning effects (Kleinschmidt & Jaeger, 2015), which asserts that speech perception is a "problem of inference under uncertainty" (p. 4), whereby listeners maintain uncertain beliefs about the

distributions of speech cues they encounter. These beliefs are formed as a result of accrued knowledge of the language over the lifespan, through encounters with potentially hundreds of different talkers, and inform generative models constructed for individual talkers (or groups of talkers). Perceptual adaptation then is posited to be a process of belief updating, where listeners adjust their beliefs about cue distributions in response to novel observations, taking into account their prior knowledge and any assumptions they have about the context in which the speech occurred. Uncertainty can potentially permeate multiple levels of the speech communication experience—including not only uncertainty about the nature of a talker's generative model but also about the talker's identity or group membership and what prior experience is relevant. For instance, if a listener has had some prior experience with Mandarin-accented English and encounters a novel foreign-accented talker, the listener may have uncertainty about the identity of their accent, and whether or not he should apply his knowledge of the cue distributions of Mandarin-accented English to the speaker.

The present research sought to further pursue this notion of uncertainty during perceptual adaptation (Kleinschmidt & Jaeger, 2015) by examining adaptation processes in L2 listeners. They provide a natural testing ground for uncertainty in adaptation, as their relatively more impoverished L2 linguistic system may yield a relatively higher level of uncertainty about the language to which they are listening (as compared to native listeners). Moreover, prior work on perceptual learning with L2 listeners has either focused on a single contrast that exists in both L1 and L2 languages (e.g., Reinisch et al., 2014) or not controlled for the segmental content of the regionally-accented speech (e.g., Mitterer & McQueen, 2009). In the current study, in addition to examining the impact of general uncertainty as a

result of more limited exposure to the target language, we also examine the effect of specific uncertainty on adaptation by providing L2 listeners with items containing contrasts that exist in their L1 as well as items that contain contrasts that do not exist in their L1 or are neutralized in certain contexts. This will address the issue of how listeners' specific L1 experience mediate the adaptation process. For items with contrasts that are perceptually challenging or typically neutralized by the L1 filter, L2 listeners' degree of uncertainty about how to categorize the contrasts will likely increase relative to items with contrasts that they can more readily categorize. Heightened uncertainty would likely inhibit adaptation, as listeners would need more information (e.g., of the nature of the cue distributions, of what prior experience is relevant, etc.) in order to be confident that a perceptual adjustment is appropriate and what that adjustment should be. Listeners' linguistic uncertainty may stem in part from having accumulated relatively less experience with the target language, which might result in each encounter with the language having a more substantial influence on the listeners' beliefs about the language's generative model (akin in principle to how listeners are more strongly influenced by situation-specific exemplars of low frequency versus high frequency words; Goldinger, 1998). By virtue of having encountered fewer instances of the target language, L2 listeners may have developed more specific beliefs, which are less flexible (Kleinschmidt & Jaeger, 2015). Holding specific language beliefs coupled with general uncertainty about the language may constrain adaptation.

Moreover, we also investigate how uncertainty about the language interacts with the certainty with which the feedback information disambiguates the speech input. For instance, lexical feedback that is an exact match for the speech input provides a high degree of

certainty about the appropriate cue distributions in the input relative to semantic contextual information (which provides relatively less certainty, as it is not 100% predictive as is the case for lexical feedback). Given that prior work has previously found that L2 listeners have a more difficult time leveraging semantic context in speech-in-noise environments, (Bradlow & Alexander, 2007; Mayo et al., 1997), the question remains as to whether L2 listeners are able to utilize the same information as L1 listeners during perceptual adaptation or whether their relatively less robust L2 linguistic knowledge base requires them to rely on only certain dimensions of information.

As such, the current study sought to address these issues by administering the same training tasks as in Chapter 3 to a group of Dutch-English bilinguals, with trained groups exposed to either single or multiple talkers and receiving either lexical or semantic contextual feedback. L2 listeners also completed the same test tasks (lexical decision and word identification) along with an additional phonetic assessment word identification task. The assessment task enabled us to determine whether listeners were capable of accurately identifying items containing phonemes involved in the accent to which they would be exposed during training. The accent deviation patterns present in Non-Standard American English (NSAE; described in Table 4.1) are well-suited to the present aims of the study. Three patterns involve phonemes that are contrastive in Dutch, typically assimilating to distinct categories, which are indicated in the table as “Dutch and English contrasts”. The “English only” contrasts are ones that are fully distinctive in English and not Dutch, such that they are either not contrastive or are neutralized in Dutch production. Dutch does not differentiate /ɛ/ - /æ/, and as a result, these segments are perceptually confusable for L2

listeners (Cutler, Weber, Smits, & Cooper, 2004). Additionally, /θ/ is not a sound that exists in Dutch. Perceptually, /θ/ is most frequently misidentified as /s/ by Dutch listeners (Cutler et al., 2004); however, /t/ is the most common substitution in production by Dutch speakers (Wester, Gilbers, & Lowie, 2007). Finally, Dutch de-voices obstruents word-finally, and as such, Dutch speakers would normally not distinguish /d/ - /t/ in word-final position and often de-voice these segments when speaking English (Booij, 1995).

**Table 4.1 Trained NSAE accent deviation patterns**

| **NSAE-accented segments** | | |
|---|---|---|
| /i/ ➔ /ɪ/ | 'cream' [krim] ➔ 'crim' [krɪm] | Dutch and English contrasts |
| /eɪ/ ➔ /ɛ/ | 'cake' [keɪk] ➔ 'kek' [kɛk] | |
| /z/ ➔ /s/ | 'guzzle' [gʌzl] ➔ 'gussle' [gʌsl] | |
| /ɛ/ ➔ /æ/ | 'west' [wɛst] ➔ 'waest' [wæst] | English only contrasts |
| /θ/ ➔ /t/ | 'thirst' [θɝst] ➔ 'turst' [tɝst] | |
| /d/ ➔ /t/ (word-finally) | 'word' [wɝd] ➔ 'wert' [wɝt] | |

We hypothesize that listeners' overall uncertainty about the language to which they are listening will modulate their perceptual adaptation processes. The enhanced uncertainty that L2 listeners would have when listening to English might result in less learning relative to L1 listeners. If listeners maintain higher levels of uncertainty, then they might require more evidence to update their beliefs about the relevant cue distributions (Kleinschmidt & Jaeger, 2015). This would predict a smaller increase in lexical endorsement rates in the probe lexical

decision task as well as a smaller difference in endorsement rates between control and trained listeners in the test lexical decision task for L2 listeners. It would also result in lower NSAE identification accuracy for L2 relative to L1 listeners. Alternatively, L2 listeners have more experience and familiarity with variant pronunciations in L2 speech and as a result may be more flexible in their mappings of speech input onto stored representations (Weber et al., 2014), which may mean that they would require less evidence to update their beliefs about the distributions. As such, this would result in a larger increase in endorsement rates during training for L2 versus L1 listeners. At test, the difference between control and trained listeners may not differ as a product of language background, as this enhanced flexibility may result in L2 control listeners having an overall higher rate of lexical endorsements as compared to L1 control listeners. Additionally, as a product of a more impoverished L2 linguistic knowledge base, L2 listeners may also have a harder time leveraging certain types of linguistic information (e.g., Bradlow & Alexander, 2007). We might then predict that L2 listeners in the lexical feedback conditions would show greater adaptation relative to those in the semantic context conditions.

The type of contrast in the NSAE-accented item (Dutch vs. English only) was also predicted to have a significant impact on adaptation for L2 listeners. If items containing contrasts that are not distinctive or neutralized in the L1 result in higher uncertainty for L2 listeners, then we would predict that that the Dutch-English bilinguals would be slower to adapt to items containing English only relative to Dutch contrasts. An alternative prediction would be that their familiarity with hearing variable pronunciations of English only contrasts, as a product of hearing Dutch-accented English (their own productions as well as other L2

speakers), would result in comparable or even greater adaptation to English only items relative to Dutch items. Indeed, the items containing English only contrasts would be produced in much the same ways as Dutch-accented English speakers would produce them. Therefore, their prior experience with these specific cue distributions could be leveraged to more efficiently update their beliefs about the distributions of NSAE.

## 2. Methods

### *2.1 Participants*

Ninety-four Dutch-English bilinguals, which included 41 participants (Male=11; *M age*=20.4 years) tested in the lab and 53 participants (Male=8; *M age*=21.3 years) tested online, were included in the study and were paid 8 Euros for their participation. These participants were native speakers of Dutch studying at Radboud University Nijmegen and were not pursuing a Master's in English or following their Bachelor's or Master's degrees in English, such that their dominant language of instruction at the university was not English. Dutch was the language of instruction during their primary and secondary education. Participants reported acquiring English after the age of 7 and had been learning English for over 4 years (participants from this population typically have 7-8 years of English education; Mitterer & McQueen, 2009). They would be considered intermediate to advanced proficiency English learners. In-lab participants were tested in a sound-attenuated booth at the Max Planck Institute for Psycholinguistics. They self-reported no hearing impairments at the time of testing. Listeners were randomly assigned to one of the 5 conditions.

### *2.2 Stimuli*

Stimulus materials for training, lexical decision and word identification tasks were identical to the experiment outlined in Chapter 2. In addition to these tasks, the phonetic

assessment task included 30 real words divided evenly between the 6 accent deviation patterns. All words possessed a minimal pair item containing the other segment in the accent pattern (e.g., presenting “bed”, which is a minimal pair with “bad”, the other segment in the /ɛ/ ➔ /æ/ pattern). This enables us to examine whether these L2 listeners were able to correctly identify items containing these particular sound patterns in SAE. All items were produced by the trained talker used in the single-talker conditions.

*2.3 Procedure*

Participants in training conditions completed two blocks of training, each preceded by a probe lexical decision task. Following training, they completed three test tasks: 1) lexical decision, 2) word identification, and 3) phonetic assessment. Participants in the Control condition completed only the test tasks.

The training, lexical decision and word identification task procedures were identical to the experiment outlined in Chapter 2. The phonetic assessment task, which was always completed at the end of the experiment, consisted of a total of 30 randomized trials. Each item was presented to listeners individually, and they were asked to transcribe the item they heard. Listeners were informed that the talker would be speaking with an SAE accent and that all presented items were real words. There was no limit on response time.

**3. Results**

*3.1 Lexical Decision Probe task*

*3.1.1 L2 listeners*

Lexical endorsement rates were calculated in each of the probe blocks and submitted to logistic linear mixed effects regression (LMER) models (Baayen et al., 2008). First, a model examining trained pattern items was constructed, with contrast-coded fixed effects for

Probe Block (1, 2), Talker Variability (Single, Multiple), Feedback (Lexical, Semantic Context) and Contrast Type (Dutch, English only) along with their interactions. The maximal random effects structure that would converge was implemented, including random intercepts for participants and items, random slopes for Probe Block and Contrast Type by participant and random slopes for Talker Variability and Feedback by item. To determine which factors made a significant contribution, model comparisons were conducted. A significant effect of Probe Block was obtained ($\beta$=0.54, SE $\beta$=0.13, $\chi^2(1)$=15.058, p=0.0001), indicating that lexical endorsement rates for trained pattern items significantly increased from Block 1 (prior to training) to Block 2. Additionally, a marginally significant Block x Contrast Type interaction was found ($\beta$=0.45, SE $\beta$=0.23, $\chi^2(1)$=3.75, p=0.053). Follow-up LMER models investigating this interaction revealed no significant block difference in endorsement rate for English only contrasts ($\chi^2$=2.5, p=0.11); however, there was a significant increase in word responses from Block 1 to 2 for items containing contrasts that also exist in Dutch ($\chi^2$=19.43, p<0.001). No other effects or interactions reached significance ($\chi^2$<3.57, p>0.06). A similar model was constructed examining responses to the nonword items (removing the fixed effect for Contrast Type, as that was not manipulated in the nonwords). A significant effect of Probe Block was found ($\beta$=0.37, SE $\beta$=0.16, $\chi^2(1)$=5.28, p=0.02), indicating that listeners were also increasing their word responses to nonword items following exposure to NSAE-accented speech. An additional Block x Talker Variability interaction was yielded ($\beta$=-1.67, SE $\beta$=0.64, $\chi^2(1)$=6.75, p=0.009), with follow-up analyses revealing that while listeners in multi-talker conditions did not show a significant increase from Block 1 to 2 in endorsement rates to nonwords ($\chi^2$=0.01, p=0.92), listeners in single-talker conditions did demonstrate an

increased endorsement rate ($\chi^2$=9.47, p=0.002). All other main effects and interactions did not reach significance ($\chi^2$<3.6, p>0.06).

To examine whether listeners' lexical endorsement rates differed between blocks as a function of item type (trained pattern vs. nonword; Figure 4.1), an additional model was constructed with fixed effects of Probe Block (1, 2) and Helmert-contrasted coded fixed effects for Item Type (A: Nonword + Trained Pattern-English only vs. Trained Pattern-Dutch; B: Nonword vs. Trained Pattern-English only). To make additional comparisons within Item Type, an additional model was run where Item Type was coded as C: Nonword vs. Trained Pattern-English only + Trained Pattern-Dutch and D: Trained Pattern-English only vs. Trained Pattern-Dutch. To adjust for multiple comparisons, the critical *p* value was set to 0.025. Random intercepts were included for participant and item and by-participant random slopes for Block and Item Types. Significant main effects of Block (β=0.47, SE β=0.10, $\chi^2$(1)=18.97, p<0.001) and Item Type B (β=2.04, SE β=0.54, $\chi^2$(1)=12.12, p=0.0005) and Item Type C (β=2.0, SE β=0.63, $\chi^2$(1)=9.1, p=0.003) were found. Crucially, a significant Block x Item Type A interaction was obtained (β=0.59, SE β=0.25, $\chi^2$(1)=5.38, p=0.02), indicating that listeners increased their lexical endorsement rates between blocks for trained pattern items containing contrasts that exist in Dutch to a greater extent than to nonwords and trained pattern items that do not exist in Dutch. No Block x Item Type B (Nonwords vs. Trained Pattern-English only) was found ($\chi^2$=0.002, p=0.96). The remaining effects did not reach significance ($\chi^2$<3.85, p>0.05).

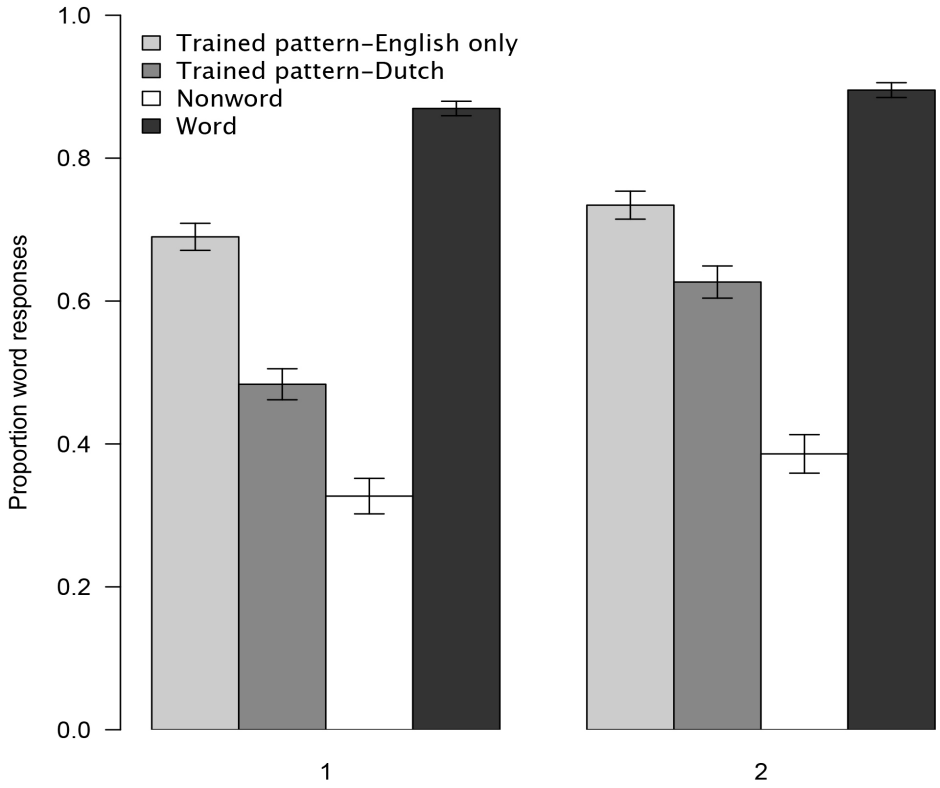


**Figure 4.1 Mean proportion of word responses to trained pattern (English only contrasts), trained pattern (Dutch contrasts), nonword and word items in Probe Block 1 and Probe Block 2. Error bars denote +/- 1 standard error.**

*3.1.2 Comparison of L1 and L2 listeners*

A combined analysis of lexical endorsement rates for the L1 listeners (Chapter 2) and the current L2 listeners was conducted (Figure 4.2). A logistic LMER model was constructed with contrast-coded fixed effects of Language Background (Dutch, English), Block (1, 2), and Helmert contrast-coded fixed effects for Item Type (A: Nonword + Trained Pattern-English only vs. Trained Pattern-Dutch; B: Nonword vs. Trained Pattern-English only). For additional comparisons within Item Type, models containing Item Type (C: Nonword vs. Trained Pattern-English only + Trained Pattern-Dutch, D: Trained Pattern-English only vs. Trained Pattern-Dutch) were also constructed. Accordingly, the critical *p* value was set to

0.025. Random intercepts for participant and item were included, along with random slopes for Language Background by item and Block and Item Type by participant. Significant main effects of Block ($\beta$=1.07, SE $\beta$=0.08, $\chi^2(1)$=117.92, p<0.001) and Language Background ($\beta$=-0.73, SE $\beta$=0.2, $\chi^2(1)$=12.61, p<0.001) were obtained, with a significant increase in lexical endorsement rates overall from Block 1 to Block 2 as well as significantly higher endorsement rates across blocks by L2 listeners relative to L1 listeners. Additionally, an effect of Item Type B (Nonword vs. Trained Pattern-English only) was also significant ($\beta$=1.82, SE $\beta$=0.43, $\chi^2(1)$=14.75, p<0.001).

In addition to these significant main effects, several 2-way interactions were also significant, including Language Background x Block ($\beta$=1.19, SE $\beta$=0.17, $\chi^2(1)$=45.19, p<0.001), Language Background x Item Type A ($\beta$=1.2, SE $\beta$=0.46, $\chi^2(1)$=6.34, p=0.01), Language Background x Item Type D ($\beta$=1.13, SE $\beta$=0.4, $\chi^2(1)$=6.34, p=0.01) and Block x Item Type B ($\beta$=0.43, SE $\beta$=0.16, $\chi^2(1)$=7.37, p=0.007). Crucially, multiple 3-way Language Background x Block x Item Type interactions[2] were also significant ($\chi^2$>5.42, p<0.02). Subsequent LMER analyses revealed significantly higher lexical endorsement rates for Trained Pattern-English only and Nonword item types in Block 1 by Dutch listeners relative to English listeners ($\chi^2$>12.93, p<0.0003), though no group differences between Trained Pattern-Dutch. By Block 2, English and Dutch listeners had comparable lexical endorsement rates by item type ($\chi^2$<1.23, p>0.27).

[2] These interactions included interactions with Item Type A (Nonword + Trained Pattern-English only vs. Trained Pattern-Dutch; $\beta$=-0.83, SE $\beta$=0.35, $\chi^2(1)$=5.42, p=0.02), Item Type B (Nonword vs. Trained Pattern-English only; $\beta$=0.86, SE $\beta$=0.33, $\chi^2(1)$=6.7, p=0.009) and Item Type D (Trained Pattern-English only vs. Trained Pattern-Dutch; $\beta$=-1.05, SE $\beta$=0.31, $\chi^2(1)$=11.19, p=0.0008)

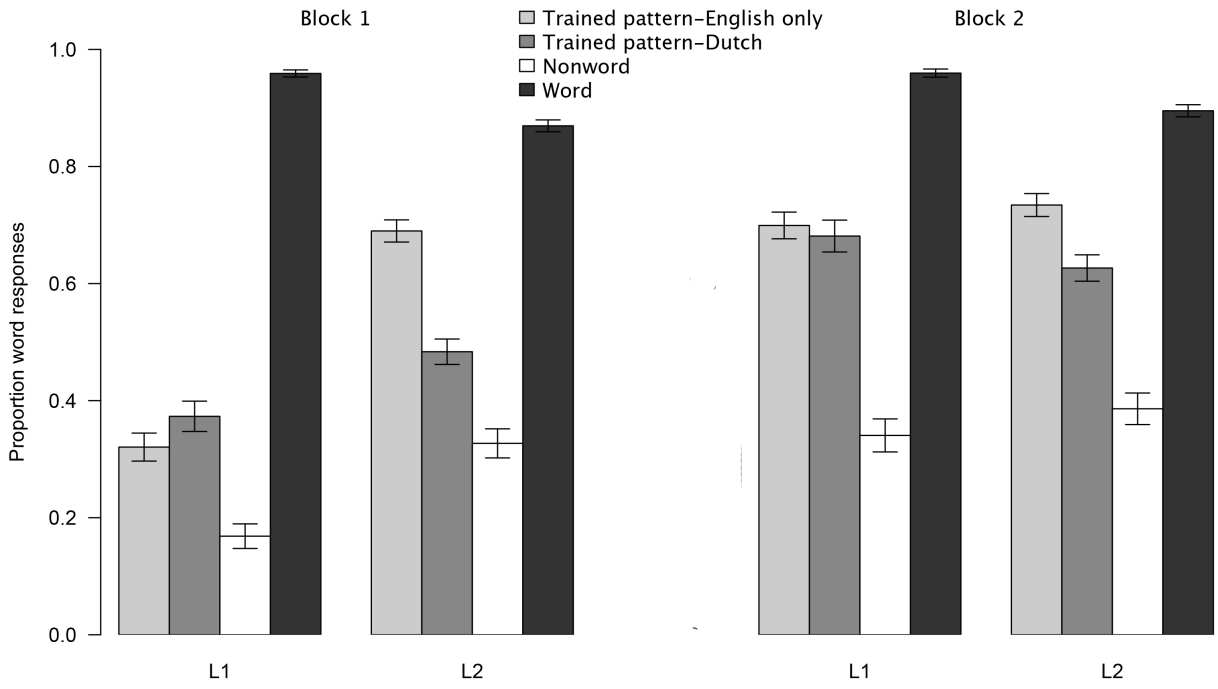


**Figure 4.2 Mean proportion of word responses to trained pattern-English only contrasts, trained pattern=Dutch contrasts, nonword and word items for L1 and L2 listeners in Probe Block 1 (left panel) and Probe Block 2 (right panel). Error bars denote +/- 1 standard error.**

*3.2 Lexical Decision Test task*

*3.2.1 Trained and Untrained Accent Pattern*

Word responses were tabulated for word, nonword and NSAE-accented (trained pattern and untrained pattern) items (Figure 4.3). Endorsement rates for nonword and NSAE-accented items produced by the trained talker were submitted to an LMER model containing contrast-coded fixed effects for Training (Control vs. Trained groups), Talker Variability (Single, Multiple), and Feedback (Lexical, Semantic Context), and Helmert contrast-coded fixed effects for Item Type (A: Nonword vs. Trained Pattern + Untrained Pattern; B: Trained Pattern vs. Untrained Pattern) along with their interactions. To further compare within Item Type, an additional model was constructed where Item Type was Helmert contrast-coded as Item Type C: Nonword + Untrained Pattern vs. Trained Pattern and Item Type D: Nonword

vs. Untrained Pattern. To adjust for multiple comparisons, the critical *p* value was set to 0.025. Random intercepts were included for participants and items, as well as by-participant random slopes for Item Type, and by-item random slopes for Training, Talker Variability and Feedback by item. Model comparisons revealed a marginally significant effect of Training ($\beta$=0.71, SE $\beta$=0.33, $\chi^2(1)$=4.46, p=0.03), whereby trained participants had overall higher lexical endorsement rates than control participants across item types. Significant effects of Item Type A (Nonword vs. Trained Pattern + Untrained Pattern; $\beta$=2.27, SE $\beta$=0.37, $\chi^2(1)$=33.12, p<0.001), Item Type C (Nonword + Untrained Pattern vs. Trained Pattern; $\beta$=1.39, SE $\beta$=0.32, $\chi^2(1)$=17.6, p<0.001) and Item Type D (Nonword vs. Untrained Pattern; $\beta$=1.56, SE $\beta$=0.33, $\chi^2(1)$=19.82, p<0.001) were found. Moreover, significant Training x Item Type B (Trained Pattern vs. Untrained Pattern; $\beta$=1.39, SE $\beta$=0.36, $\chi^2(1)$=14.48, p<0.001) and Training x Item Type C (Nonword + Untrained Pattern vs. Trained Pattern) interactions also obtained ($\beta$=1.58, SE $\beta$=0.40, $\chi^2(1)$=14.26, p<0.001); however, the Training x Item Type D (Nonword vs. Untrained Pattern) did not reach significance ($\chi^2$=1.04, p=0.31). These findings indicate that trained listeners endorsed trained pattern items as being words significantly more than nonwords and items with untrained accent patterns relative to control listeners.

In addition, the Feedback x Item Type C (Nonword + Untrained Pattern vs. Trained Pattern) interaction was marginally significant ($\beta$=-1.0, SE $\beta$=0.45, $\chi^2(1)$=4.75, p=0.029), with follow-up LMER analyses revealing that listeners who received semantic contextual feedback had a marginally larger difference in endorsement rates between trained pattern items and nonword + untrained pattern items relative to lexical feedback groups. Main effects

of Talker Variability and Feedback and other interactions with Item Type did not reach significance ($\chi^2$<2.52, p>0.62).

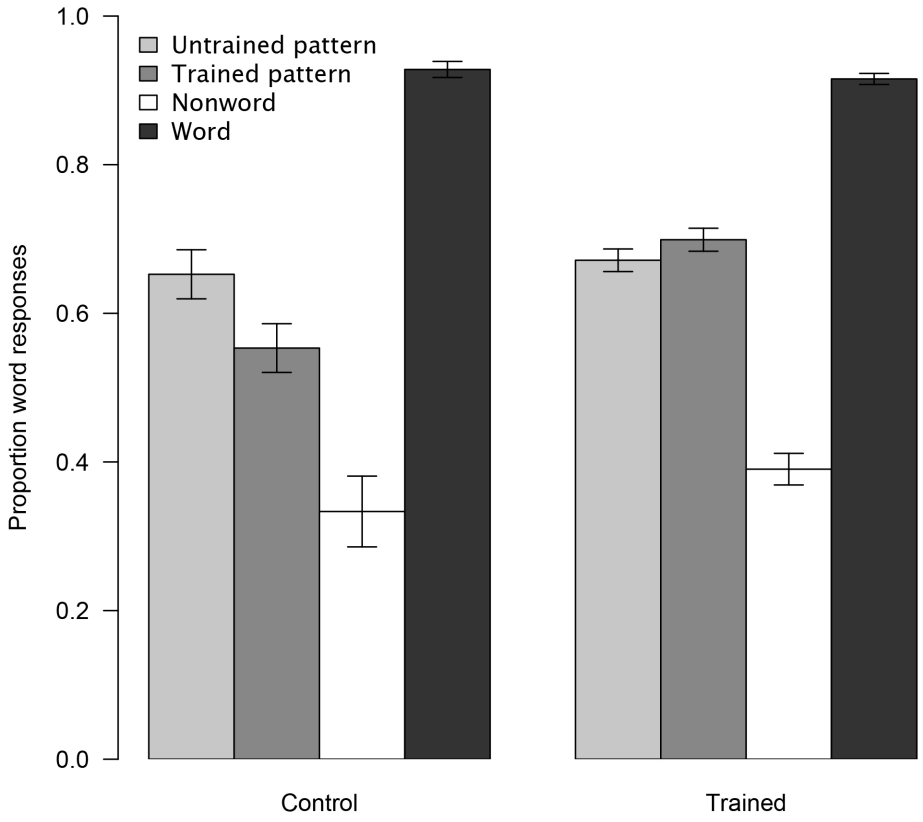


**Figure 4.3 Mean proportion of word responses by Item Type for control and trained listeners**

*3.2.2 Contrast Type*

Finally, to examine the impact of contrast type (English only vs. Dutch), an LMER model with responses to trained pattern items for both trained and untrained talkers was constructed (Figure 4.4), where fixed effects for Training, Talker Variability, Feedback, Contrast Type (English only vs. Dutch) were included along with their interactions. Random intercepts for participant and item, as well as a by-participant random slope for Contrast

Type and by-item random slopes for Training, Talker Variability and Feedback, were also included. A significant main effect of Training was found ($\beta$=1.45, SE $\beta$=0.36, $\chi^2(1)$=15.295, $p<0.001$) along with a significant effect of Contrast Type ($\beta$=1.31, SE $\beta$=0.31, $\chi^2(1)$=16.043, $p<0.001$). No other effects or interactions reached significance ($\chi^2<2.78$, $p>0.1$). Across groups, listeners were more willing to endorse items with English-only contrasts as being words relative to those containing Dutch contrasts.

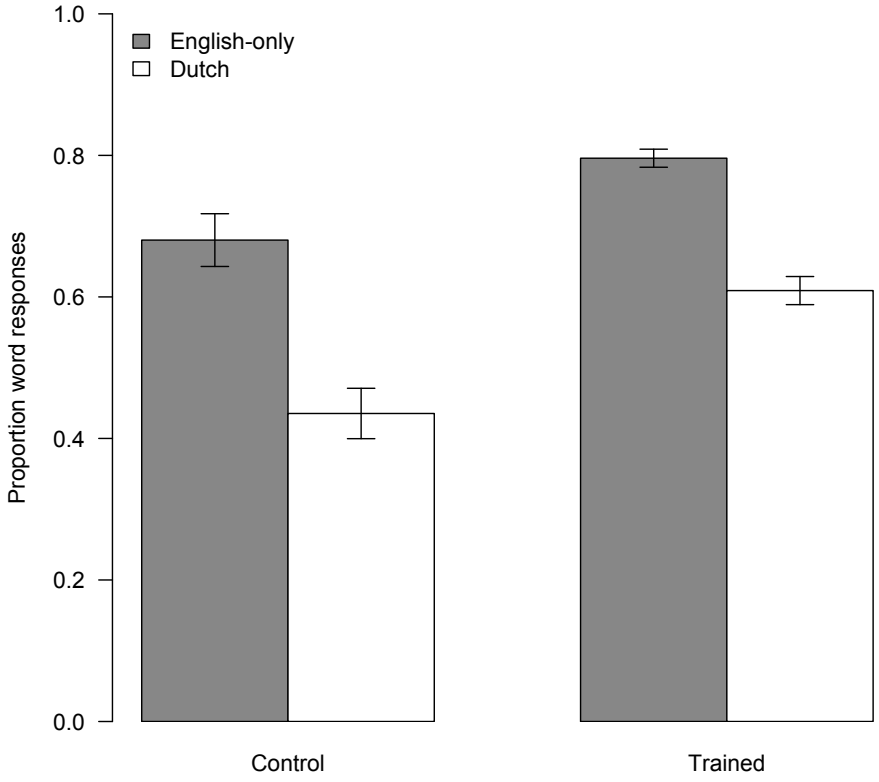


**Figure 4.4 Proportion of word responses to trained pattern items by Contrast Type (English-only, Dutch) and Training (Control, Trained)**

*3.2.3 L1 vs. L2 Comparison*

To examine how L2 listeners' endorsement rates compared to L1 listeners, a combined analysis of lexical endorsement rates for the L1 listeners (Chapter 2) and the L2 listeners was conducted (Figure 4.5). A logistic LMER model with responses to trained

pattern and nonword items was constructed with fixed effects of Language Background (Dutch, English), Training (Control, Trained), and Helmert contrast-coded fixed effects for Item Type (A: Nonword + Trained Pattern-English only vs. Trained Pattern-Dutch; B: Nonword vs. Trained Pattern-English only). Additional models were constructed for further within-Item Type comparisons (C: Nonword vs. Trained Pattern-English only + Trained Pattern-Dutch, D: Trained Pattern-English only vs. Trained Pattern-Dutch), with the critical *p* value set to 0.025 to account for multiple comparisons. Random intercepts for participant and item were included, along with random slopes for Language Background and Training by item and Item Type by participant. Significant main effects of Training ($\beta$=2.1, SE $\beta$=0.29, $\chi^2(1)$=50.02, p<0.001), Item Type B (Nonword vs. Trained Pattern-English only; $\beta$=2.0, SE $\beta$=0.26, $\chi^2(1)$=46.59, p<0.001), Item Type C (Nonword vs. Trained Pattern-English only + Trained Pattern-Dutch; $\beta$=2.09, SE $\beta$=0.29, $\chi^2(1)$=40.89, p<0.001) and Item Type D (Trained Pattern-English only vs. Trained Pattern-Dutch; $\beta$=-0.86, SE $\beta$=0.25, $\chi^2(1)$=10.79, p=0.001) were found.

A significant Training x Language Background was obtained ($\beta$=1.82, SE $\beta$=0.55, $\chi^2(1)$=10.70, p=0.001), and follow-up LMER models revealed significantly higher lexical endorsement rates by L2 control listeners relative to L1 control listeners (p=0.004), but no significant difference between L1 and L2 trained listeners (p=0.996). Training x Item Type interactions[3] ($\chi^2$>10.1, p<0.001) reflect significantly higher endorsement rates by trained listeners than control listeners to all item types, including nonwords, but a larger group

[3] Interactions with Type A (Nonword + Trained Pattern-English only vs. Trained Pattern-Dutch; $\beta$=1.1, SE $\beta$=0.34, $\chi^2(1)$=10.1, p=0.001), Type B (Nonword vs. Trained Pattern-English only; $\beta$=1.18, SE $\beta$=0.36, $\chi^2(1)$=10.20, p=0.001) and Type C (Nonword vs. Trained Pattern-English only + Trained Pattern-Dutch; $\beta$=1.68, SE $\beta$=0.41, $\chi^2(1)$=15.34, p<0.001).

difference for Trained Pattern-English only and Trained Pattern-Dutch items relative to nonwords.

There were also significant Language Background x Item Type interactions[4] ($\chi^2$>8.29, p<0.004). The remaining main effects and 3-way interactions did not reach significance ($\chi^2$<2.96, p>0.09). Follow-up analyses revealed that L2 listeners produced significantly higher endorsement rates to English only items relative to Dutch items (p=0.001), with no difference in endorsement rates to these items for L1 listeners (p=0.22). Both L1 and L2 listeners endorsed English only items as being words significantly more than nonwords (p<0.001); however, the magnitude of this difference was larger for L2 listeners as compared to L1 listeners.

[4] Interactions with Type A (Nonword + Trained Pattern-English only vs. Trained Pattern-Dutch; β=0.66, SE β=0.22, $\chi^2$(1)=8.29, p=0.004), Type B (Nonword vs. Trained Pattern-English only; β=-0.66, SE β=0.22, $\chi^2$(1)=8.67, p=0.003) and Type D (Trained Pattern-English only vs. Trained Pattern-Dutch; β=0.81, SE β=0.19, $\chi^2$(1)=16.19, p<0.001)

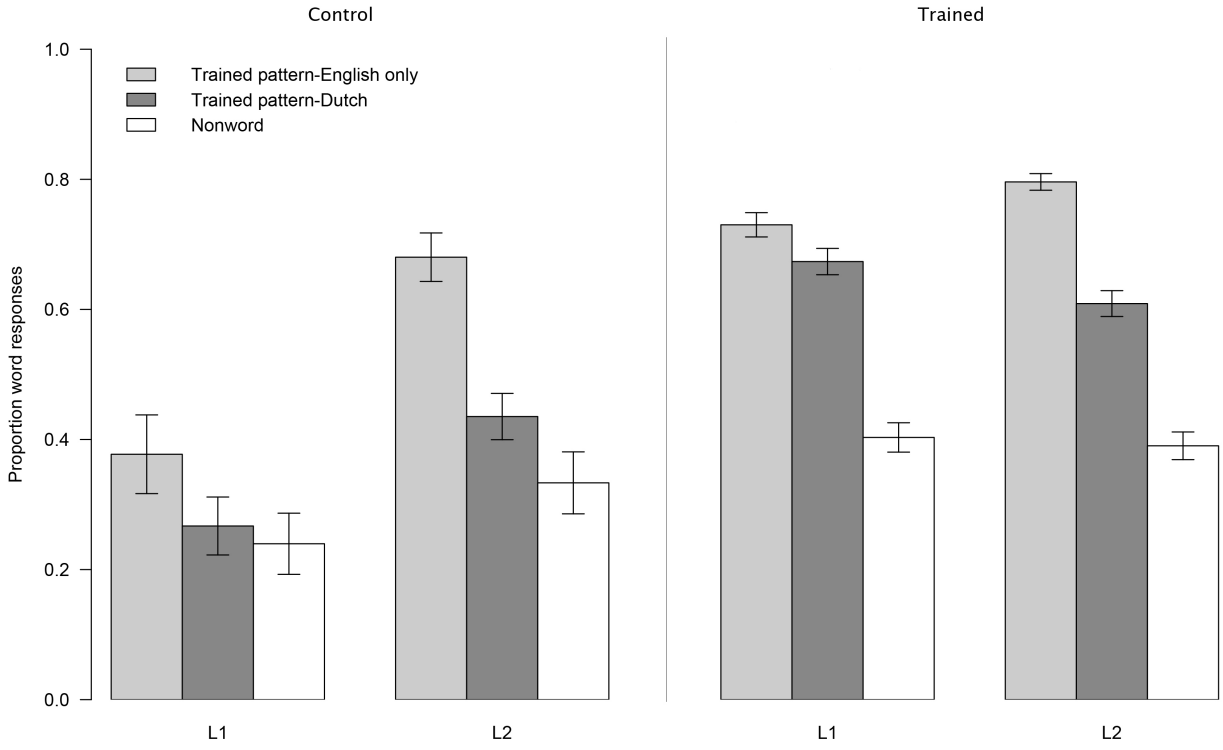


**Figure 4.5 Mean proportion of word responses to trained pattern-English only contrasts, trained pattern-Dutch contrasts and nonword items for L1 and L2 listeners by Training (control listeners: left panel; trained listeners; right panel). Error bars denote +/- 1 standard error.**

*3.3 Word Identification task*

Two types of stimuli were included in the word identification task: minimal pair (where the item would be considered a word in a SAE accent and a different word in NSAE) and lexicality change (where the item would be considered a nonword in SAE but a word in NSAE). Responses to minimal pair items were coded in two ways: 1) identification accuracy in NSAE, termed NSAE Accuracy, and 2) identification accuracy in a SAE accent, termed SAE Accuracy. For instance, the word "pod" would be produced as [pɑt] ("pot") in NSAE. If listeners transcribed it as "pot", then this would be correct in terms of SAE Accuracy but incorrect for NSAE Accuracy. Conversely, if they transcribed "pod", then their NSAE Accuracy would increase but not their SAE Accuracy.

For lexicality change items, identification accuracy in NSAE (NSAE Accuracy) was also calculated. This would involve being presented with item produced as [ʃælf] and accurately transcribing it as "shelf". The number of nonword responses (denoted by an 'X' by participants) was determined for both minimal pair and lexicality change items; however, the results here focus on the lexicality change data (as the proportion of minimal pair items identified as nonwords was relatively low).

*3.3.1 Trained vs. Untrained Accent Pattern*

An LMER model was constructed, with NSAE Accuracy for lexicality change items produced by the trained talker as the dependent variable (Figure 4.6). The same fixed effects for Training and Feedback conditions were implemented as in prior models, along with Item Type (Trained Pattern vs. Untrained Pattern) and all 2- and 3-way interactions. No significant main effects or interactions were found ($\chi^2<2.65$, $p>0.1$). An identical model was constructed for minimal pair items, which similarly did not yield any significant main effects or interactions ($\chi^2<1.17$, $p>0.28$).

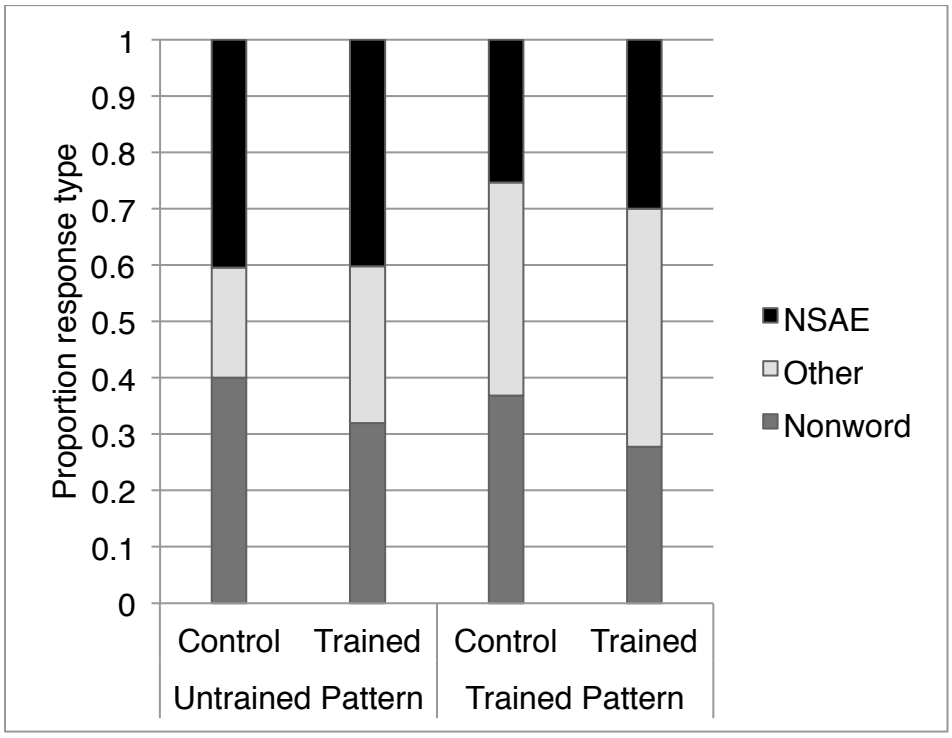


**Figure 4.6 Proportion of responses that were 1) accurate based on NSAE accent, 2) other responses, or 3) nonword responses for lexicality change items produced by a trained talker by group (Control, Trained) and Item Type (Trained pattern, Untrained Pattern).**

For nonword responses to lexicality change items, Training was the only significant factor obtained (β=-0.96, SE β=0.41, $\chi^2(1)=5.46$, p=0.02), with control listeners responding "nonword" significantly more than trained listeners (*M*=38% vs. 30%). All other effects and interactions were not significant ($\chi^2<1.58$, p>0.21).

The same model structure was applied to the SAE Accuracy data (only calculated for minimal pair items; Figure 4.7). A significant Training x Item Type interaction was found (β=-0.97, SE β=0.47, $\chi^2(1)=4.03$, p=0.04) as well as a significant Talker Variability x Item Type interaction (β=1.03, SE β=0.45, $\chi^2(1)=5.06$, p=0.02). Subsequent LMER analyses to investigate these 2-way interactions revealed that control listeners produced marginally higher SAE Accuracy rates on items containing untrained accent patterns relative to trained

listeners (β=-0.59, SE β=0.34, $\chi^2(1)=2.97$, p=0.08). Furthermore, listeners from single-talker conditions were found to produce significantly higher SAE Accuracy rates (*M*=56%) than those from multi-talker conditions (*M*=50%) on items with untrained accent patterns (β=-0.97, SE β=0.44, $\chi^2(1)=4.3$, p=0.04).

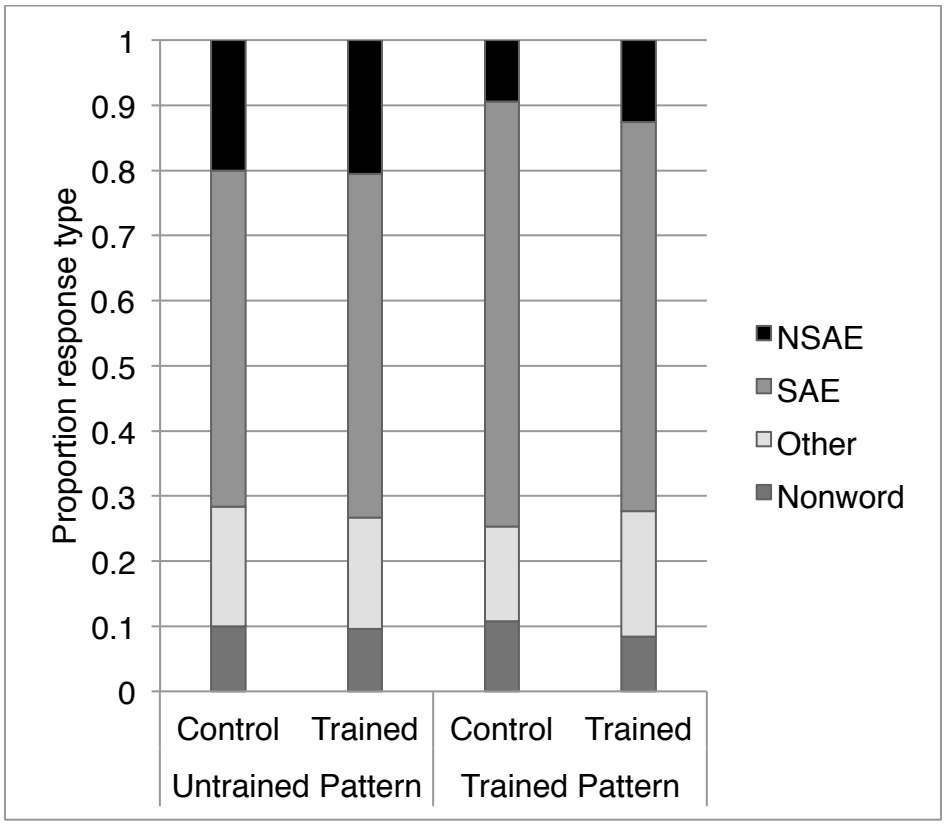


**Figure 4.7 Proportion of responses that were 1) accurate based on NSAE accent, 2) accurate based on SAE accent, 3) other responses, or 4) nonword responses for minimal pair change items produced by a trained talker by group (Control, Trained) and Item Type (Trained pattern, Untrained Pattern).**

*3.3.2 Contrast Type*

To investigate the impact of contrast type, LMER models were constructed on NSAE Accuracy for lexicality items (trained pattern) with fixed effects of Training, Talker Variability and Feedback conditions, as in prior models, as well as an effect of Contrast Type

(Dutch, English only) along with their interactions. Significant effects of Training (β=0.63, SE β=0.25, $\chi^2(1)=6.3$, p=0.01) and Contrast Type were obtained (β=-1.54, SE β=0.42, $\chi^2(1)=12.25$, p=0.0005), with higher NSAE Accuracy rates for trained versus control listeners, and higher accuracy for items containing contrasts that are English only relative to Dutch (Figure 4.8). Furthermore, a significant Training x Contrast Type was also found (β=1.23, SE β=0.4, $\chi^2(1)=9.23$, p=0.002). Control and trained listeners displayed comparable NSAE Accuracy rates for items that contained English only contrasts (*M*=44% for both groups); however, trained listeners had higher NSAE accuracy rates relative to control listeners for items containing Dutch contrasts (*M*=26% vs. 18%, respectively). Analysis of nonword responses yielded significant effects of Training (β=-0.98, SE β=0.39, $\chi^2(1)=6.24$, p=0.01) and Contrast Type (β=1.03, SE β=0.32, $\chi^2(1)=9.9$, p=0.002), with more nonword responses to Dutch versus English only items and by control relative to trained listeners.

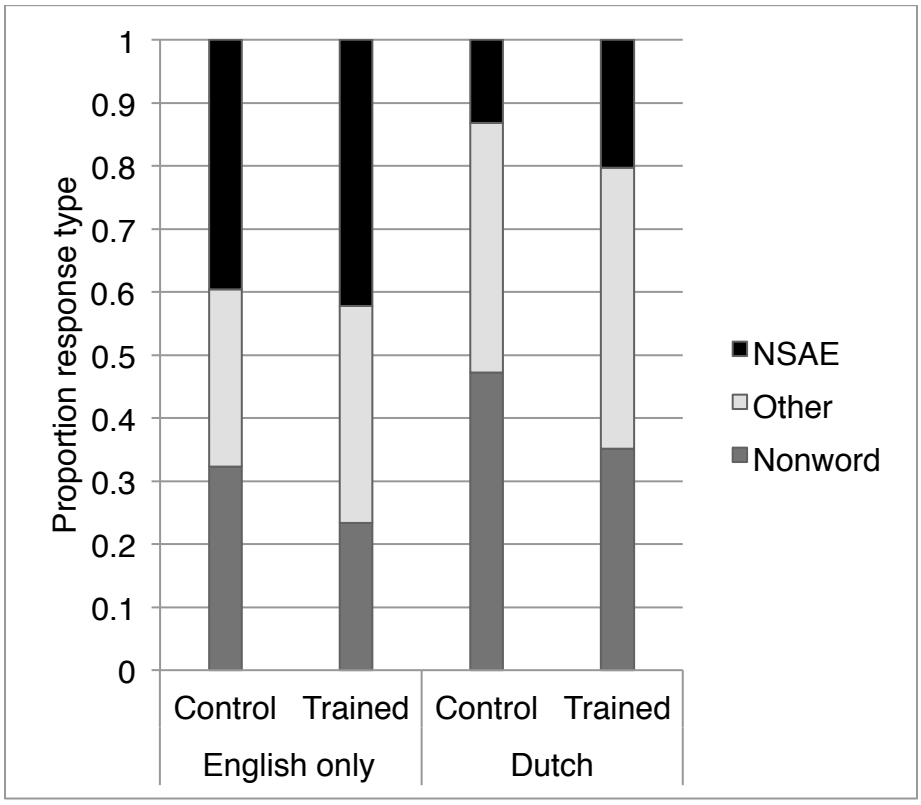


**Figure 4.8 Proportion of responses that were 1) accurate based on NSAE accent, 2) other responses, or 3) nonword responses for lexicality change items by group (Control, Trained) and Contrast Type (English only, Dutch).**

An identical model performed on NSAE Accuracy for minimal pair items (Figure 4.9) found only a significant effect of Contrast Type ($\beta$=-2.2, SE $\beta$=0.72, $\chi^2(1)$=12.25, p=0.0005), with overall higher NSAE accuracy rates for items with English only contrasts (*M*=19%) as compared to those with Dutch contrasts (9%). No significant effects or interactions were found for SAE Accuracy ($\chi^2$<3.25, p>0.07).

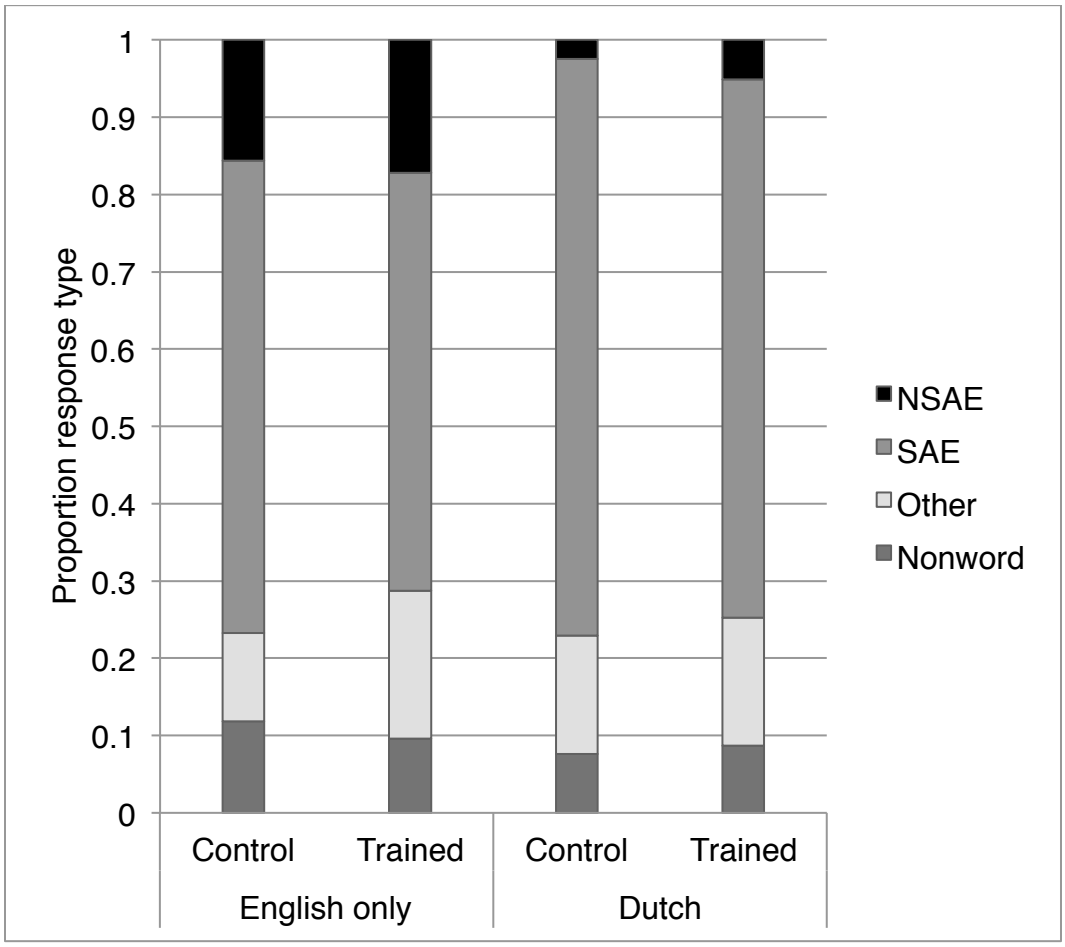


**Figure 4.9 Proportion of responses that were 1) accurate based on NSAE accent, 2) accurate based on SAE accent, 3) other responses, or 4) nonword responses for minimal pair change items produced by group (Control, Trained) and Contrast Type (English only, Dutch).**

*3.3.3 L1 vs. L2 comparison*

To compare word identification performance between L1 and L2 listeners, an LMER model for on NSAE Accuracy on lexicality change (trained pattern) items was constructed (Figure 4.10), with fixed effects of Language Background (Dutch, English), Training (Trained, Control) and Contrast Type (English only, Dutch), along with their interactions. Random intercepts for participant and item were included, as well as by-item random slopes for Language Background and Training, as well as a by-participant slope for Contrast Type. A significant effect of Training was obtained ($\beta$=2.0, SE $\beta$=0.32, $\chi^2(1)$=37.897, $p<0.001$),

indicating that across language backgrounds, participants who underwent training were more likely to accurately identify words based on the NSAE accent relative to control listeners. An effect of Contrast Type was also significant (β=-1.0, SE β=0.47, $\chi^2(1)=6.98$, p=0.008). Additionally, 2-way Training x Language Background (β=1.96, SE β=0.58, $\chi^2(1)=11.14$, p=0.0008) and Language Background x Contrast Type (β=1.58, SE β=0.35, $\chi^2(1)=16.41$, p<0.001) interactions were found. Moreover, a marginally 3-way Training x Language Background x Contrast type interaction was obtained (β=-1.34, SE β=0.69, $\chi^2(1)=3.64$, p=0.056). Subsequent analyses revealed that L2 control listeners had significantly higher NSAE Accuracy rates for English only items than L1 control listeners; however, L1 and L2 trained listeners performed comparably on these items. For items containing Dutch contrasts, L1 and L2 control listeners were not significantly different, but L1 trained listeners achieved higher NSAE Accuracy as compared to L2 trained listeners.

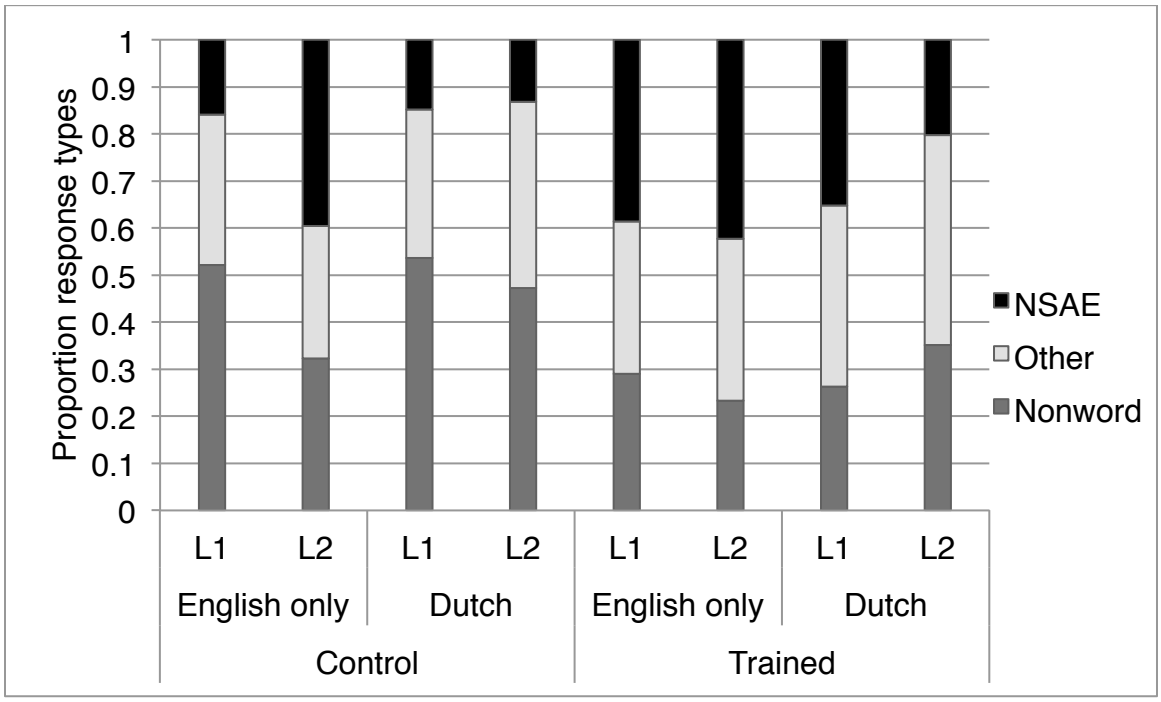


**Figure 4.10 Proportion of responses that were 1) accurate based on NSAE accent, 2) other responses, or 3) nonword responses for lexicality change items by language background (L1, L2), Contrast Type (English only, Dutch), and group (Control, Trained).**

An identical model on NSAE accuracy for minimal pair change items (Figure 4.11) revealed significant effects of Training ($\beta$=2.3, SE $\beta$=0.6, $\chi^2(1)$=20.06, p<0.001) and Contrast Type ($\beta$=-1.34, SE $\beta$=0.49, $\chi^2(1)$=7.16, p=0.007), along with a significant Language Background x Contrast Type interaction ($\beta$=2.0, SE $\beta$=0.32, $\chi^2(1)$=13.4, p=0.0003). The 3-way Training x Language Background x Contrast Type was also significant ($\beta$=-4.1, SE $\beta$=1.35, $\chi^2(1)$=10.23, p=0.001). Congruent with the analysis of lexicality change items, L2 control listeners produced significantly higher NSAE accuracy rates than L1 control listeners for minimal pair change items containing English only contrasts (p<0.001), whereas the L1 and L2 trained listeners did not differ significantly (p=0.43). In contrast, for items containing Dutch contrasts, L1 and L2 control listeners achieved comparable NSAE accuracy rates

(p=0.37), and the L1 trained listeners obtained higher accuracy rates than L2 trained listeners (p=0.01).

A model containing the same fixed and random effects structure was constructed with SAE Accuracy on minimal pair change items as the dependent variable. Results revealed significant main effects of Training (β=-1.0, SE β=0.33, $\chi^2(1)=9.21$, p=0.002) and Language Background (β=0.85, SE β=0.24, $\chi^2(1)=11.14$, p=0.0008), along with a significant Training x Language Background interaction (β=-1.4, SE β=0.62, $\chi^2(1)=6.98$, p=0.008). Subsequent LMER models indicate that L1 control listeners produced significantly higher SAE accuracy rates than L1 trained listeners (p=0.002); however, this effect of training did not reach significance for the L2 listeners (p=0.09), though there was a numerical trend in the appropriate direction.

Finally, an analysis of nonword response rates for L1 and L2 groups found a significant main effect of Training (β=-1.76, SE β=0.34, $\chi^2(1)=25.696$, p<0.001) along with significant 2-way interactions of Language Background x Contrast Type (β=0.85, SE β=0.24, $\chi^2(1)=8.46$, p=0.004) and Training x Language Background (β=0.85, SE β=0.24, $\chi^2(1)=4.94$, p=0.03). Follow-up analyses indicate that while the effect of training was significant for both L1 (p<0.001) and L2 groups (p=0.01), with control listeners producing higher nonword response rates relative to trained listeners, the magnitude of this difference was significantly larger for L1 listeners as compared to L2 listeners.

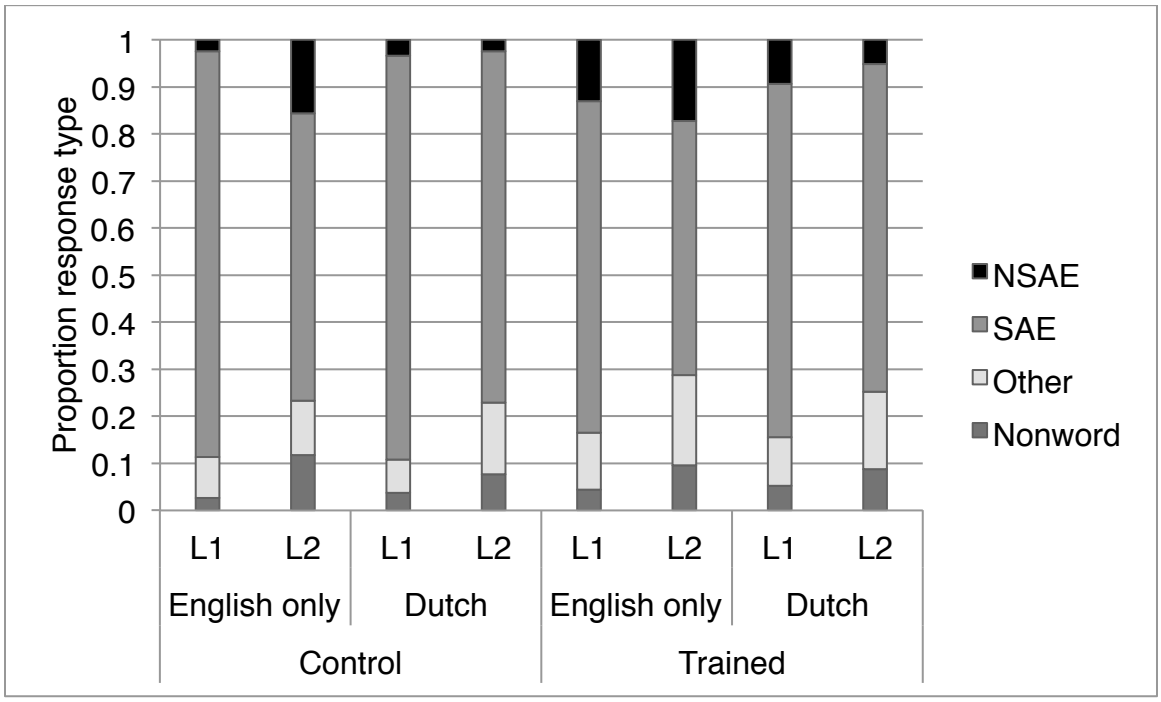


**Figure 4.11 Proportion of responses that were 1) accurate based on NSAE accent, 2) accurate based on SAE accent, 3) other responses, or 4) nonword responses for minimal pair change items by Language Background (L1, L2), Contrast Type (English only, Dutch), and group (Control, Trained).**

*3.4 Phonetic Assessment task*

To determine whether L2 listeners differed with respect to their ability to identify SAE-accented contrasts, word identification accuracy was tabulated for the phonetic assessment task (Figure 4.12), which they completed after training and test tasks. A logistic LMER was constructed with contrast-coded fixed effects of Training (Control, Trained) and Contrast Type (Dutch, English only), with participant and item as random intercepts, and a by-participant random slope for Contrast Type and a by-item random slope for Training. No significant effects or interactions were found ($\chi^2$<2.48, p>0.12). Though, there was a numerical trend for words containing Dutch contrasts to be more accurately identified (*M*=79%) relative to words with English only contrasts (*M*=69%). L2 listeners' accuracy rates were compared against a group of L1 listeners who completed this task on Amazon

Mechanical Turk (n=20), and a significant effect of language background was found ($p<0.001$), whereby native listeners were significantly more accurate at identifying these words (*M*=92%) relative to the Dutch-English bilinguals (*M*=75%).

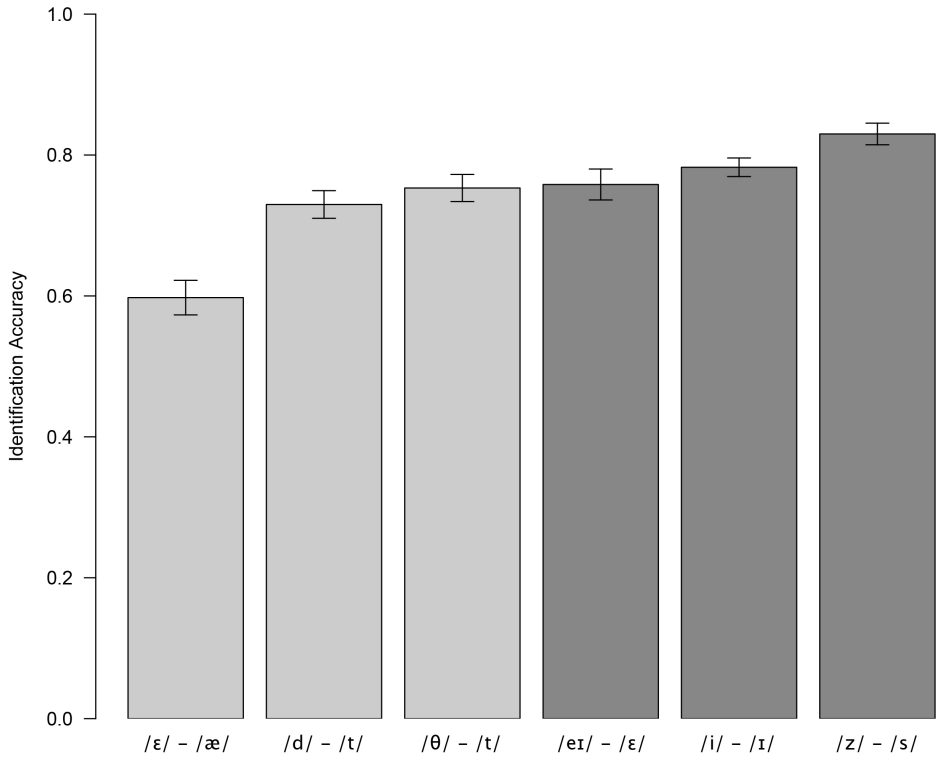


**Figure 4.12 Proportion word identification accuracy in assessment task by contrast. Light grey bars denote contrasts designated "English only"; dark grey bars "Dutch".**

*3.5 Summary*

The findings of the current experiment revealed that, in the probe task, L2 listeners saw a significant increase from Block 1 (prior to training) to Block 2 in lexical endorsement rates for items containing Dutch contrasts though not for English only items. Relatively higher endorsement rates for English only items in Block 1 may have contributed to this lack of increase. Moreover, endorsement rates to nonwords also increased, specifically for listeners in single-talker conditions. A comparison of L1 and L2 groups found that across

blocks, L2 listeners had higher endorsement rates overall relative to L1 listeners. Prior to training, Dutch listeners endorsed nonwords and items containing English only contrasts significantly more than English listeners (with comparable endorsement rates between the two groups for Dutch contrast items). By Block 2, both groups had comparable endorsement rates by item type.

Following training, L2 listeners completed a lexical decision task. Trained listeners endorsed items containing trained patterns significantly more than nonwords and untrained pattern items relative to control listeners. Across groups, listeners endorsed items that contained English only contrasts as being words more than those that contained Dutch contrasts. L2 control listeners had significantly higher endorsement rates relative to L1 control listeners; however, no significant difference emerged between L1 and L2 trained groups. Furthermore, the difference in endorsement rates to English only items relative to nonword items was significantly larger for L2 versus L1 groups.

Finally, in the word identification task, for items that would be considered nonwords in SAE (lexicality change items), L2 trained listeners identified these items as nonwords significantly less than L2 control listeners. Across groups, more nonword responses were provided when the item contained a Dutch contrast relative to an English only contrast. When a word response was provided, trained listeners correctly identified these items based on what they had learned about NSAE significantly more than control listeners when the item contained a trained Dutch contrast (control and trained listeners performed comparably on items with trained English only contrasts). For items that were actually real words in SAE (minimal pair items), control listeners identified items containing untrained patterns based on

SAE more often than trained listeners. Additionally, single-talker conditions had higher SAE accuracy rates than multi-talker conditions. In response to lexicality change items, there was a significantly larger difference in nonword responses between control and trained listeners for the L1 group compared to the L2 group. For both lexicality change and minimal pair items, L2 control listeners had higher NSAE accuracy rates for English only items relative to L1 control listeners, with comparable performance by trained listeners. For Dutch contrasts, control listeners from both language groups performed comparably, and L1 trained listeners had higher accuracy NSAE accuracy rates than L2 trained listeners. Moreover, SAE accuracy rates for minimal pair items was higher for L1 trained versus L1 control listeners; however, this difference was not significant for L2 listeners.

**4. Discussion**

The current study investigated the impact of uncertainty during perceptual learning. Using an adaptation paradigm, this was examined by comparing L1 and L2 listeners, who necessarily vary in their linguistic knowledge, as indexed by L2 listeners' tendency to be slower in processing their L2 (Cutler, 2012) and impaired access to L2 linguistic knowledge in non-optimal conditions (e.g., Bradlow & Alexander, 2007). As a result of their relatively more impoverished L2 linguistic system, L2 listeners were posited to maintain a higher degree of uncertainty when listening to their L2 as compared to native listeners. In the present work, this uncertainty may have resulted in enhanced flexibility within the perceptual system for L2 listeners, as evidenced by higher endorsement rates from L2 control listeners and trainees prior to training relative to L1 control listeners. That is, Dutch-English bilinguals, without any training or exposure to NSAE-accented English, were more likely to

consider nonword items as being words in English than native English listeners. It is conceivable that when listening in a second language, where listeners are aware that their knowledge is less robust or as a result of experience with juggling more than one language (Weber et al., 2014), that they may be more willing to be flexible in their phonetic-to-lexical mappings. Overall higher endorsement rates may have also resulted from phantom word activation (Broersma & Cutler, 2008), whereby nonword items containing perceptually confusable segments for L2 listeners may have activated real word neighbours and led them to consider the items to be real words.

Consistent with prior work (e.g., Mitterer & McQueen, 2009), L2 listeners demonstrated significant adaptation to an unfamiliar accent of English. However, L2 listeners did not appear to be affected by the uncertainty level of the provided feedback during training, with lexical and semantic context groups performing similarly in probe and test tasks, consistent with native listeners (Chapter 2). It could be the case that the Dutch-English bilinguals used in the present study were proficient enough to leverage semantic contextual information as effectively as lexical feedback, and it remains for future work to examine how L2 proficiency influences the efficacy of different types of linguistic information during adaptation. These findings indicate that the L2 perceptual system, at least ones at a moderate to high level of proficiency, can draw upon varied types of higher-level linguistic information, not only lexical as has been shown in prior studies (e.g., Reinisch et al., 2012), but also contextual, during adaptation in much the same way as L1 listeners.

One of the novel contributions of the current study was the finding that perceptual adaptation was modulated by the type of phonemic contrast employed in the accent deviation

pattern, that is, whether or not it was contrastive in the listeners' native language. Three of the deviation patterns involved segments that either assimilate to a single category (/ɛ/ - /æ/, /θ/ - /t/) or are neutralized in a particular context (/d/ - /t/ word-finally) in Dutch, which were termed "English only" contrasts, while the other three patterns were considered to be distinctive in Dutch (/i/ - /ɪ/; /e/ - ɛ/, /z/ - /s/), referred to as "Dutch" contrasts. Items containing Dutch contrasts were less likely to be considered words prior to training (and by control listeners) than those with English only contrasts. This likely resulted from listeners' higher certainty level regarding the Dutch contrasts (as a product of their perceptual precision) and also their being less familiar with pronunciation variants involving these contrasts, in much the same way as L1 listeners (whose perceptual precision with these English contrasts is much higher and who are less familiar with variant pronunciations) demonstrated lower endorsement rates initially. However, despite being more resistant to being considered a word at first, higher certainty about the contrasts enabled L2 listeners to adapt these categories to a greater extent during training relative to English only contrasts. Indeed, trained listeners showed an effect of training only for items with Dutch contrasts, with an increase in lexical endorsement rates in the probe tasks and higher NSAE Accuracy in the word identification task as compared to the control listeners.

Items with English only contrasts, on the other hand, seemed to be more resistant to adaptation. The starting lexical endorsement and identification accuracy rates for these types of items were significantly higher (as illustrated by the control listeners). Their being familiar with variant pronunciations involving these contrasts may have led to phantom activation (Broersma & Cutler, 2012), such that they not only considered them to be words but also

more accurately identified them based on the NSAE accent (both control and trained listeners). However, L2 trained listeners did not improve significantly beyond that point as a result of training, with no significant increase in endorsement rates during the probe task, and no control versus trained group differences on the test tasks. While one could imagine that because their starting performance was higher on these items, they could have reached a ceiling, thus making any effect of training difficult to detect. However, NSAE Accuracy (lexicality change items), for example, for control and trained listeners was at 41%, which would not appear to be a performance ceiling. These findings point to the interactive influences of prior knowledge and uncertainty on perceptual adaptation. Listeners had preliminary beliefs about the distributions of these categories, based on their own prior experience with pronunciation variants of English along with their native language influences on the shape of these distributions, that would have initially led them to categorize items with Dutch contrasts more often as being nonwords and English only contrasts as words. Listeners who underwent training were able to update their beliefs for contrasts about which they had a higher degree of certainty (as to the nature of the distributions), yielding an increase in accuracy and endorsement rates for items containing these contrasts. In contrast, it is conceivable that listeners' relative uncertainty about the distributions of English only contrasts led to their maintaining stability and prevented them from further updating their beliefs as a result of training, thereby yielding no significant performance difference from control listeners. One could argue that L2 listeners were not uncertain about the distributions of the English only contrasts but rather did not detect any difference between the NSAE and SAE pronunciations, and as such, made no adjustments. However, the phonetic assessment

task found no significant difference between listeners' ability to accurately identify SAE-accented words containing English only versus Dutch segments, indicating that these contrasts have been established, to a certain extent, for these listeners. Their performance, though, was not completely native-like in the assessment task, as native listeners were still found to outperform them at identifying the English words, suggesting that these L2 listeners may still possess a degree of linguistic uncertainty.

Moreover, despite being relatively more certain about Dutch versus English only contrasts, L1 trained listeners were more accurate at identifying words based on their knowledge of NSAE, namely for Dutch contrast items, relative to L2 trained listeners. While prior work (Reinisch et al., 2012) has shown a comparable degree of category boundary shifting between L1 and L2 listeners in phoneme categorization tasks following exposure to an ambiguous sound, the present findings reveal that L2 listeners may have a more difficult time utilizing this knowledge at a higher-level of processing (namely, involving lexical identification). If we consider these findings in the context of uncertainty, this performance discrepancy as a product of language background may have resulted from L2 listeners maintaining an overall higher level of uncertainty when listening to English relative to L1 listeners. Listeners with a higher level of uncertainty about the nature of the relevant generative model would require more evidence (either through additional training or more explicit training) in order to effectively update the beliefs about that model relative to those with lower levels of uncertainty. As a result, L2 trained listeners were more likely to identify a given item as a nonword (lexicality change items) or identify it based on how it was actually produced (SAE accent; minimal pair items) relative to L1 listeners.

The current research has considered this notion of uncertainty as reflecting listeners' confidence—confidence, for example, in the relevant cue distributions, the identity of the speaker, or whether a category adjustment should be made. Listeners' uncertainty level is a critical component of learning and may underlie many of the different studies that have found slowed or inhibited generalization and perceptual learning. For example, the lack of perceptual learning when listeners hear an ambiguous sound but see a speaker with a pen in their mouth may stem from listeners' having higher certainty that the pen is the source of this ambiguity than the speaker (Kraljic, Samuel, et al., 2008). A lack of generalization to a novel foreign-accented talker after single talker exposure (Bradlow & Bent, 2008) may arise from listeners having higher uncertainty as to whether the specific acoustic deviation patterns from the trained talker are applicable to the novel talker, in other words, whether or not they should leverage their prior experience with their cue distributions and apply that knowledge to the novel talker. To further investigate the issue of uncertainty in perceptual learning, future work could consider including some measure of listener confidence (e.g., rating task, where listeners rate their confidence on perhaps a number of different dimensions) or an online measure such as eye-tracking, which could give insight into how quickly listeners interpret a given item and the other hypotheses the listener is entertaining.

In sum, the findings of the present work highlight the influence of prior experience, namely linguistic experience, on listeners' ability to flexibly accommodate atypical pronunciations during speech perception. Dutch-English bilinguals were found to effectively utilize either lexical or semantic contextual information to adapt to a Non-Standard American English accent; however, this adaptation process was mediated by their knowledge of (or

certainty about) the particular phonemic contrasts employed in the items. That is, listeners' prior knowledge was found to intersect with uncertainty about specific cue distributions for contrasts that are neutralized in their native language. Prior experience with English pronunciation variants, similar to the ones employed in NSAE, enabled listeners to endorse items as words and identify them more accurately, even without training (English only contrasts) than items with contrasts that exist in their native language (Dutch contrasts), of which they would have relatively less experience with hearing variable pronunciations. However, their relative uncertainty about the particular cue distributions for contrasts that do not exist in their L1 inhibited their ability to further adapt to the NSAE accent for those contrasts, while they were able to demonstrate significant learning for contrasts that do exist natively. This would predict that as listeners' L2 categories sharpen and become more native-like, listeners would be better able to update their beliefs about the distributions of those categories when encountering atypical pronunciations and show effects of adaptation in their second language.

# CHAPTER 5: CONCLUSIONS

This dissertation sought to better understand the processes that underlie perceptual adaptation to speech variation. In particular, the current experiments examined how the perceptual system utilizes different types of linguistic knowledge and signal-based information during adaptation and how linguistic experience modulates this process. Notions of predictive strength, that is, the extent to which disambiguating information narrows the space of possible category options, as well as uncertainty provided a perspective within which to conceptualize these issues (Guediche et al., 2014; Kleinschmidt & Jaeger, 2015). Listeners are posited to construct a generative linguistic model for a given talker or listening situation, containing information about the specific cue distributions of linguistic categories for that talker or situation. However, they maintain uncertain beliefs as to the exact nature of these generative models, as listeners never truly have full access to this information. Adaptation results from an updating of these beliefs based on a mismatch between predictions derived from various sources of disambiguating information and the observed sensory input. The present work hypothesized that the predictive strength of the disambiguating information would mediate adaptation, with more predictive information (e.g., providing an exact lexical match for the input) increasing listeners' certainty about how to categorize the input, thereby facilitating adaptation, relative to less predictive information (e.g., jabberwocky sentence).

Chapter 2 addressed this issue by examining the relative contribution of different levels of linguistic information (e.g., phonemic, lexical, prosodic) in adaptation to Mandarin-accented English. Participants transcribed Mandarin-accented English sentences in speech-

shaped noise and were provided with feedback sentences by a native talker that either aligned with the target at 1) all linguistic levels, 2) syntactic and sub-lexical levels with real words or 3) syntactic and sub-lexical levels with non-words. Performance in these conditions was compared against conditions providing non-English "feedback" (Korean sentences) or accent only exposure. Contrary to our initial prediction that adaptation performance would vary as function of predictive strength, listeners who were provided with any kind of native English-accented feedback (matched or mismatched sentences) significantly outperformed those who were not presented with this kind of feedback. While mismatching sentences did not give information about the specific lexical items contained in the Mandarin-accented sentences, they did provide information about the syntactic structure and acoustic manifestation of native-accented phonemes and prosody for similar types of sentences, yielding points of alignment between the native- and Mandarin-accented sentences beyond the lexical. These findings suggest that native English listeners will leverage any linguistically relevant information during adaptation, even if it is not highly predictive of the input.

Chapter 3 further explored the issue of predictive strength by comparing the contribution and interaction of different types of information, namely linguistic knowledge (e.g., lexical vs. semantic contextual feedback) and signal-based information (single vs. multiple talker training) during exposure to a novel English accent (NSAE). Similar to Chapter 2, both types of linguistic feedback were found to be equally effective for perceptual learning, despite potential differences in their degree of predictiveness. Talker variability emerged as a significant factor only in one of the more challenging conditions (word identification of minimal pair items), with listeners in the multi-talker condition identifying

words from the untrained talker based on the NSAE accent significantly more than listeners in the single talker condition. However, in most other contexts, both single and multi-talker training yielded significant adaptation to both the trained and untrained talker. It was also not found to interact with any type of linguistic feedback. Native listeners were thus remarkably flexible and efficient in their adaptation, drawing upon a range of different types of information to rapidly adjust the relevant categories. Moreover, they were also found to apply the generative model that they had developed over the course of training to a novel talker and, to a certain degree, novel accent deviation patterns. These generalization patterns likely stemmed from listeners' explicit knowledge of the similarity between trained and untrained talkers (with respect to their being from the same accent group) as well as their implicit knowledge of phonetic structural similarity.

Following Kleinschmidt and Jaeger (2015), uncertainty is posited to exist for listeners on multiple levels—not just uncertainty about the particular cue distributions of a given generative model or the relevant prior experience that should be brought to bear during adaptation but also uncertainty about the linguistic system more generally. In particular, we hypothesized that second language listeners have an extra source of uncertainty as a result of their more impoverished linguistic knowledge, coupled with their awareness of that reduced knowledge relative to native listeners. Chapter 4 investigated this by testing Dutch-English bilinguals with the same paradigm as in Chapter 3 and comparing their performance to native English listeners. Furthermore, we also expected that L2 listeners would possess more uncertainty about the distribution of certain phonemic contrasts than others, namely those that are not contrastive in their native Dutch, which would potentially slow the adaptation

process. Consistent with prior work (e.g., Broersma & Cutler, 2008), L2 listeners generally endorsed more items as being real English words than L1 listeners. This may have arisen from their enhanced uncertainty about the L2 linguistic system—they are likely aware that their general linguistic knowledge (e.g., about the appropriate distribution of phonemes or vocabulary items) is more limited as compared to native listeners. They are thus not as certain about what is or is not an actual word, so they remain generally more flexible about what they are willing to consider a word of English. Similar to native listeners, the predictive strength of the feedback did not mediate adaptation, with listeners from both lexical and semantic context conditions demonstrating significant learning. Linguistic experience did however have an impact on adaptation with respect to which accent deviation patterns L2 listeners were found to adapt. Adaptation did not occur to the same extent as native listeners for the contrasts that do not exist or are neutralized in certain contexts in Dutch, but did occur for other contrasts, despite being capable of identifying these segments with relatively high accuracy. Given that, it appears that listeners' enhanced uncertainty about the cue distributions for these particular non-native contrasts may have inhibited adaptation.

Taken together, the present work provides support for the notion that listeners are "aggressively opportunistic" (p. 1217, Samuel & Kraljic, 2009) during perceptual learning, capable of leveraging any relevant information available in the context, even if it does not provide an exact match for the input they received. The different types of feedback provided in this dissertation varied in their predictive strength, which we initially expected to be tied to the certainty with which a listener would make a perceptual adjustment (that is, information that is highly predictive might increase listeners' certainty about how to categorize the input,

thereby facilitating adaptation, relative to less predictive information). However, the findings of the current research suggest a divergence between predictive strength and certainty, such that listeners were willing to draw upon varied types of information (regardless of their predictive strength) in order to enhance their certainty and make them willing to adjust the relevant categories. Listeners were able to do so rapidly; indeed, trained listeners (Chapters 3 and 4) only received between 16 and 24 items of total exposure (with feedback) for each accent deviation pattern. From that relatively limited exposure, they were even able to make inferences about structurally-related contrasts, as evidenced by generalization to untrained patterns in certain contexts. In most real-world communicative contexts, the kind of information available to listeners is not typically concurrently-presented, lexically-matching feedback, but rather sequentially-presented, discourse-building information. It is perhaps a necessity then for the perceptual system to have developed in such a way that it can capitalize on whatever kinds of information it can extract.

A much higher degree of uncertainty may be required for perceptual learning to be inhibited. It remains for future work to determine what level of uncertainty the system must experience before opting to remain stable rather than adapting. For example, how does the adversity of the listening situation (beyond the challenge of listening to foreign-accented speech) impact the perceptual system's capacity to utilize different types of information? In the present experiment, the noise conditions were not highly challenging (+5 SNR; Chapter 2), and listeners' training occurred in many contexts that were self-paced. In high cognitive load or more challenging noise or response conditions, where the perceptual system is being taxed, it is conceivable that the system will rely more heavily on information that may

facilitate learning more efficiently, such as highly predictive feedback. Listeners in more adverse listening contexts may not have the resources necessary to extrapolate what is necessary to make efficient perceptual adjustments from less predictive information.

Additionally, second language listeners were also found to be able to adapt to a novel accent of English, utilizing the same types of knowledge- and signal-based information as native listeners. This provides further evidence that similar adaptation processes are at work in the second language as in the native language (Mitterer & McQueen, 2009; Reinisch et al., 2012). However, consistent with Kleinschmidt and Jaeger (2015)'s proposition that listeners contend with uncertainty at multiple levels, the present work found that these learners, despite showing significant learning in certain contexts, likely maintained a higher uncertainty about the nature of their second language generative model. Linguistic experience was found to play a dynamic role in adaptation, demonstrating a combination of plasticity and stability within the system. In the case of contrasts about which listeners were perhaps more uncertain, the system appeared to be resistant to adapting, as indicated by a lack of significant difference between control and trained listeners. More evidence may be required for listeners to be confident enough to make a perceptual adjustment in such cases. For contrasts about which listeners were more certain, evidence of plasticity emerged, as shown by L2 trained listeners' significantly higher endorsement rates and identification accuracy than control listeners. This suggests that listeners who were exposed to the NSAE accent were more willing to make category adjustments for these specific contrasts. It is important to note that the adaptation paradigm in Chapters 3 and 4 was originally designed to include a range of different contrasts, to more closely approximate foreign-accented speech,

but which placed restrictions on the possible vocabulary items to be presented. Future work following up on the effects of linguistic experience on adaptation could restrict themselves to a smaller number of contrasts to open up the number of possible vocabulary items included and could control for their familiarity with L2 listeners.

Several open questions remain with respect to the influence of language background on adaptation. First, how does one generate learning for contrasts for which listeners have higher uncertainty? One explanation is to simply provide more training, as it would allow listeners more time to amass evidence about cue distributions for these uncertain contrasts. It could also be the case that listeners reached a plateau in terms of what they could extract from the input (as it was structured in the current experiment), which might indicate that different sources of information or a different training structure may be necessary to promote learning in these cases (e.g., being explicitly taught about the accent pattern). Indeed, prior work has found interactive effects of the perceptual abilities of the listener and the structure of the instructional paradigm on the acquisition of a non-native contrast (Perrachione, Lee, Ha, & Wong, 2011), whereby perceptually weak learners were actually inhibited by high-variability training relative to perceptually strong learners (as assessed by pre-training aptitude measures). Given that, different kinds of training may be necessary for certain L2 contrasts which are more perceptually challenging for L2 listeners. It is perhaps useful to remember that perceptual learning is a *learning* process, and individuals differ widely in how best they learn and how efficiently they do so (e.g., Golestani & Zatorre, 2009; Perrachione et al., 2011). Future work examining individual differences in adaptation would shed light on whether the factors that contribute to perceptual learning of speech variation are shared for

other types of speech learning (e.g., acquiring novel phonological contrasts) and with learning in other cognitive domains (e.g., improving on spatial orientation discrimination tasks). Moreover, such work could also have practical implications, as training paradigms could be better tailored to maximize improvement for particular types of learners.

Additionally, the Dutch-English bilinguals included in this research were intermediate to advanced proficiency. This relatively high proficiency level was chosen in order to ensure that some degree of adaptation would be obtained, particularly given that higher-level information (e.g., lexical, semantic contextual) was used in this paradigm to drive learning. However, relatively little is known about how proficiency factors into perceptual adaptation. One would imagine that lower proficiency listeners would have even higher levels of uncertainty about their second language, which would be predicted to yield slower adaptation relative to higher proficiency listeners. On the other hand, low proficiency learners might possess such unstable representations (at least of contrasts that do not exist or are neutralized in their native language), compared to high proficiency learners, that they might conversely be less resistant to modify these representations. The relative instability of newly-forming representations could potentially make them more malleable.

Linguistic experience provided a window into one possible way that uncertainty could influence perceptual learning. The question remains as to whether or not other sources of uncertainty operate in a similar manner. For example, pronunciations in foreign-accented speech are not always consistent, even within a single speaker. Hanulíková and Weber (2012) found that Dutch and German learners of English would often produce two or three different substitutions for the same English segment /θ/. This would likely result in increased

uncertainty about how best to make category adjustments. Indeed, Witteman, Weber, and McQueen (2014) provided an initial examination of this issue, comparing a "consistent accent" condition (German-accented Dutch items only) against an "inconsistent accent" condition (native-accented Dutch with certain items pronounced with a German accent) in a cross-modal priming experiment, finding slowed adaptation by the inconsistent accent condition. How do listeners accommodate foreign-accented speech when, for instance, a portion of the time they have evidence that one substitution occurs (e.g., /θ/ "think" ➔ /s/ "sink"), while perhaps a smaller proportion of the time a different substitution occurs (e.g., /θ/ "thin" ➔/t/ "tin")? If we assume that listeners are tracking the distributional information in the speech signal, then we must also assume that their generative model would reflect this distribution. If one substitution, for example, occurs 60% of the time and the other substitution 40% of the time, then we might expect listeners' responses to reflect that distribution.

In the present work, we have referred to second language learners as having a higher degree of linguistic uncertainty than native listeners. However, it is crucial to note that uncertainty about a language encompasses a multitude of potential sources, including knowledge of the linguistic structure (e.g., phoneme distributions, vocabulary), the frequency of exposure to second language exemplars, the nature of that exposure (e.g., native vs. foreign-accented talkers) and its variability (e.g., number of talkers, environmental conditions). For instance, research examining individual differences in adaptation by older native listeners found that vocabulary knowledge was a significant predictor of the magnitude of adaptation to foreign-accented speech (Janse & Adank, 2012), which we would

expect to similarly be the case for second language learners as well. As the current experiments do not allow us to pinpoint which sources that give rise to language uncertainty specifically influence the adaptation process, it is important for future work to tease apart these sources and determine their relative contribution during learning. As adaptation involves the integration of prior knowledge, beliefs about the relevant generative model and information extracted from the sensory input, such research would provide insight into how the nature of prior knowledge (e.g., vocabulary size, frequency of exposure) interacts with these other components of the adaptation process.

## REFERENCES

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–9. http://doi.org/10.1037/a0013552

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903–9. http://doi.org/10.1177/0956797610389192

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Baese-Berk, M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, *133*(3), EL174–EL180. Retrieved from http://faculty.wcas.northwestern.edu/ann-bradlow/publications/2013/2013_BaeseBerkBradlowWright.pdf

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. http://doi.org/10.1121/1.1603234

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*(6), 592–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14629691

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34).

Booij, G. (1995). *The Phonology of Dutch*. Oxford: Oxford University Press.

Bradlow, A. R., Ackerman, L., Burchfield, L. A., Hesterberg, L., Luque, J., & Mok, K. (2011). Language- and Talker-Dependent Variation in Global Features of Native and Non-Native Speech. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 356–359). Hong Kong. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3594809&tool=pmcentrez&rendertype=abstract

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. http://doi.org/10.1121/1.2642103

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

*106*(2), 707–29. http://doi.org/10.1016/j.cognition.2007.04.005

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, *36*(1), 22–34. http://doi.org/10.1016/j.system.2007.11.003

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, *27*(4), 539–571. http://doi.org/10.1080/01690965.2011.555268

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. http://doi.org/10.1121/1.1815131

Clopper, C. G. (2012). Effects of dialect variation on the semantic predictability benefit. *Language and Cognitive Processes*, *27*(7-8), 1002–1020. http://doi.org/10.1080/01690965.2011.558779

Clopper, C. G., & Pisoni, D. B. (2004). Effects of Talker Variability on Perceptual Learning of Dialects. *Language and Speech*, *47*(3), 207–238. http://doi.org/10.1177/00238309040470030101

Cooper, A., & Bradlow, A. R. (n.d.). Linguistically-guided adaptation to foreign-accented speech. *Psychonomic Bulletin & Review*.

Cutler, A. (2012). Native listening: The flexibility dimension. *Dutch Journal of Applied Linguistics*, *1*(2), 169–187. http://doi.org/10.1075/dujal.1.2.02cut

Cutler, A., & Broersma, M. (2005). Phonetic Precision in Listening. In W. J. Hardcastle & J. Mackenzie Beck (Eds.), *A Figure of Speech* (pp. 64–91). Mahwah, NJ: Lawrence Erlabum Associates.

Cutler, A., McQueen, J. M., Butterfield, S., Norris, D., & Planck, M. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. In J. Fletcher, D. Loakes, M. Wagner, & R. Goecke (Eds.), *Proceedings of Interspeech 2008*. Brisbane.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, *28*(5), 746–755.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. http://doi.org/10.1121/1.1810292

Dahan, D., & Magnuson, J. S. (2006). *Spoken Word Recognition*. (M. A. Gernsbacher & M. J. Traxler, Eds.)*Handbook of Psycholinguistics*. Amsterdam: Academic Press. http://doi.org/10.1016/B978-012369374-7/50009-2

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–41. http://doi.org/10.1037/0096-3445.134.2.222

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–38. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15971687

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–1953. http://doi.org/10.1121/1.2178721

Flege, J. E. (1995). Speech Language Speech Learning: Theory, Findings and Problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD.

Goldinger, S. D. (1998). Echoes of echoes? An Episodic Theory of Lexical Access. *Psychological Review*, *105*(2), 251–79. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9577239

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1826729

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612. http://doi.org/10.1146/annurev.psych.49.1.585

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*, 55–67.

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, *7*(January), 126. http://doi.org/10.3389/fnsys.2013.00126

Hanulíková, A., & Weber, A. (2012). Sink positive: linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception & Psychophysics*, *74*(3), 613–29. http://doi.org/10.3758/s13414-011-0259-7

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, 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*(2), 460–74. http://doi.org/10.1037/0096-1523.34.2.460

Hirata, Y. (2004). Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts. *Computer Assisted Language Learning*, *17*(December), 357–376. http://doi.org/10.1080/0958822042000319629

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*(2), 896–907. http://doi.org/10.1121/1.1823291

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., & Diesch, E. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, 47–57. http://doi.org/10.1016/S0

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*(8), 1563–85. http://doi.org/10.1080/17470218.2012.658822

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. http://doi.org/10.1037/a0038695

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*, 54–81. http://doi.org/10.1016/j.cognition.2007.07.013

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–78. http://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*(2), 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1–15. http://doi.org/10.1016/j.jml.2006.07.010

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First Impressions and Last Resorts: How Listeners Adjust to Speaker Variability. *Psychological Science*, *19*(4), 332–338. http://doi.org/10.1111/j.1467-9280.2008.02090.x

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–43. http://doi.org/10.1016/j.cogpsych.2009.04.001

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–62. http://doi.org/10.1080/03640210802035357

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686–693. http://doi.org/10.1044/jslhr.4003.686

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–21. http://doi.org/10.1037/0278-7393.31.2.306

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological Abstraction in the Mental Lexicon. *Cognitive Science*, *30*, 1113–1126.

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*(1), 509–17. http://doi.org/10.1121/1.3664087

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review*, *13*(6), 958–965. http://doi.org/10.3758/BF03213909

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: evidence from a learning paradigm. *Cognitive Science*, *35*(1), 184–97. http://doi.org/10.1111/j.1551-6709.2010.01140.x

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, *4*(11), e7785. http://doi.org/10.1371/journal.pone.0007785

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–78. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3515846&tool=pmcentrez&rendertype=abstract

Nespor, M., Peña, M., & Mehler, J. (2003). On the Different Roles of Vowels and Consonants in Speech Processing and Language Acquisition. *Lingue E Linguaggio*, (2), 203–230. http://doi.org/10.1418/10879

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, *35*(1), 85–103. http://doi.org/10.1016/j.wocn.2005.10.004

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209–28. http://doi.org/10.1037//0278-7393.21.5.1209

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. http://doi.org/10.1016/S0010-0285(03)00006-9

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–76. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9599989

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*(January), 620. http://doi.org/10.3389/fpsyg.2012.00620

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. http://doi.org/10.1121/1.3593366

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*(2), 539–555. http://doi.org/10.1037/a0034409

Reinisch, E., Weber, A., & Mitterer, H. (2012). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 75–86. http://doi.org/10.1037/a0027979

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105. http://doi.org/10.1016/j.wocn.2014.04.002

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–49. http://doi.org/10.1111/j.1467-7687.2008.00786.x

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception & Psychophysics*, *71*(6), 1207–1218. http://doi.org/10.3758/APP

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception & Psychophysics*, 355–367. http://doi.org/10.3758/s13414-015-0987-1

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306–3316. http://doi.org/10.1121/1.3101452

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195–211. http://doi.org/10.1037/a0016803

Smith, R., Holmes-Elliott, S., Pettinato, M., & Knight, R.-A. (2014). Cross-accent intelligibility of speech in noise: long-term familiarity and short-term familiarization. *Quarterly Journal of Experimental Psychology*, *67*(3), 590–608. http://doi.org/10.1080/17470218.2013.822009

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences* , 1–10. http://doi.org/10.1073/pnas.1523266113

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. http://doi.org/10.1016/j.wocn.2010.09.001

Thomson, R. I., & Derwing, T. M. (2015). The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. *Applied Linguistics*, *36*(3), 326–344. http://doi.org/10.1093/applin/amu076

van der Zande, P., Jesse, A., & Cutler, A. (2014). Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *Journal of Phonetics*, *43*, 38–46. http://doi.org/10.1016/j.wocn.2014.01.003

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, *35*(1-2), 103–113. http://doi.org/10.1016/S0167-6393(00)00098-4

Weatherholtz, K. (2015). *Perceptual Learning of Systemic Cross-Category Vowel Variation*. The Ohio State University.

Weber, A., Betta, A. M. Di, & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, *46*, 34–51. http://doi.org/10.1016/j.wocn.2014.05.002

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491. http://doi.org/10.1016/j.wocn.2010.12.004

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, *29*(2-3), 477–491. http://doi.org/10.1016/j.langsci.2006.12.029

Witteman, M. J., Bardhan, N. P., Weber, A., & McQueen, J. M. (2014). Automaticity and Stability of Adaptation to a Foreign-Accented Speaker. *Language and Speech*. http://doi.org/10.1177/0023830914528102

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*(3), 537–56. http://doi.org/10.3758/s13414-012-0404-y

Witteman, M. J., Weber, A., & McQueen, J. M. (2014). Tolerance for inconsistency in foreign-accented speech. *Psychonomic Bulletin & Review*, *21*(2), 512–9. http://doi.org/10.3758/s13423-013-0519-8

Xie, X., & Fowler, C. a. (2013). Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics*, *41*(5), 369–378. http://doi.org/10.1016/j.wocn.2013.06.003

Zhang, X., & Samuel, A. G. (2014). Perceptual Learning of Speech Under Optimal and Adverse Conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 200–217. http://doi.org/10.1037/a0033182