Distribution Agreement

Signature:

________________________________ _______________

Jessica E. D. Alexander Date

Specificity and Generalization in Perceptual Adaptation to Accented Speech

By

Jessica E. D. Alexander
Bachelor of Arts, Hendrix College, 2003
Master of Arts, Emory University, 2005

Adviser: Lynne C. Nygaard, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the Graduate School
of Emory University in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
2009

Abstract

Specificity and Generalization in Perceptual Adaptation to Accented Speech
By Jessica E. D. Alexander

The present study investigated the process of perceptual learning of specific properties of foreign accented speech. Sidaras et al. (2009) found that listeners perceptually adapted to Spanish-accented speech when trained in a transcription task. The first experiment replicated and extended the findings to Korean-accented speech. Native English-speaking listeners were trained with Korean-accented English and tested with either the same speakers heard during training or novel Korean-accented speakers. Listeners trained with Korean-accented speech performed better than untrained controls when tested with novel words from either the same or different Korean-accented talkers, indicating that listeners perceptually adapt to Korean-accented speech. The second experiment investigated whether perceptual learning of accented speech was accent-specific or if learning one accent or multiple accents could generalize to other novel accents. Native English-speaking listeners were trained in the same transcription paradigm as the first experiment with Spanish-accented speech, with Korean-accented speech, or with a group of speakers from 6 different first language backgrounds. During training, all listeners transcribed accented words and received feedback. At test, listeners transcribed novel words produced by unfamiliar Korean- or Spanish-accented speakers and were not given feedback. Overall, there was evidence of specificity of learning, with listeners who were trained and tested on the same variety of accented-speech showing better transcription at test than untrained controls. However, patterns of transcription accuracy differed for listeners tested with Korean-accented speech and those tested with Spanish-accented speech, suggesting that accent-specific learning may be dependent on characteristics such as linguistic similarity, previous experience, or overall intelligibility. These results have implications for the kinds of mechanisms listeners employ during perceptual learning of accented speech and the changes that occur in the acoustic-phonetic categories and representations that subserve spoken language perception. That listeners show specificity of learning in the second experiment suggests that listeners' categories for speech sounds change in specific ways to reflect their experience with the accented talkers.

Specificity and Generalization in Perceptual Adaptation to Accented Speech

By

Jessica E. D. Alexander
Bachelor of Arts, Hendrix College, 2003
Master of Arts, Emory University, 2005

Adviser: Lynne C. Nygaard, Ph.D.

A dissertation submitted to the Faculty of the Graduate School
of Emory University in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
2009

Acknowledgements

I would like to thank my adviser, Lynne Nygaard for her excellent guidance in all steps of this project as well as her wisdom and support throughout my graduate career. I feel extremely fortunate to have had such a knowledgeable and caring mentor, and I hope that I have been half as valuable to her as she has been to me. I would also like to thank my dissertation committee members, Laura Namy, Harold Gouzoules, Donald Tuten, and Stella Lourenco, for their advice and suggestions as I began the project, and their enthusiasm and support as I finished it. I also owe a great deal to Sabrina Sidaras for her help and collaboration in this and other projects as well as her love and friendship both in and out of the lab. Many thanks are due to the researchers who helped in stimulus and data collection, especially Lauren Clepper, Melanie Hammet, Melanie Tumlin, Kathy Jernigan, Katie Bankieris, Hayley Heaton, Nihar Mathur, Elena Seiti, and countless other hardworking students. The love and support of my family and friends has been invaluable to me over the years, and I should especially thank my husband, Noel Alexander, for always being there when I needed someone.

Table of Contents

Korean-accented speech.

Figure 8. *Experiment 2: Transcription performance at test for listeners trained with Korean-, Spanish- or Mixed-accented speech and tested with Spanish-accented speech.*

Figure 9. *Diagram of possible acoustic-phonetic category shifts. Listeners may be forming subcategories for the accented speech sounds and/or they may be shifting native category boundaries to include accented items.*

When listeners hear speech, they encounter a wealth of both linguistic and non-linguistic information. The linguistic content, such as the syllables, words, and the relations between them, provide direct communication between the talker and the listener. The nonlinguistic properties of speech provide indexical information about talker identity (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985), sex (Monsen & Engebretson, 1977), status, health (Labov, 1972), regional dialect, and emotional state (Frick, 1985; Murray & Arnott, 1993; Scherer, Banse, Wallbott, & Goldbeck, 1991). All this nonlinguistic information present in the speech stream stems from variability in speech production across groups and individuals. Add to this the ways speech sounds change as they are used and combined in differing phonetic contexts, and spoken language is highly variable, with influences coming from linguistic variation as well as from individual talker and group factors, such as region of origin or social group membership.

One focus of research in spoken language processing has been investigating how listeners process spoken language in light of its highly variable nature. Even though listeners perceive speech in the presence of a great deal of variability, they achieve perceptual constancy, retrieving linguistic structure with little difficulty. For example, listeners know that the sound /t/ in the word *top* is the same when pronounced by different speakers or in different words, even though the acoustic realization of the sound may be unique in each situation. Traditional accounts have attempted to explain the ability to understand speech despite variability through the use of a normalization process (Green, Kuhl, Meltzoff, & Stevens, 1991; Halle, 1985; Joos, 1948; Obleser & Eisner, 2009; Pisoni, 1997; Summerfield & Haggard, 1973; Tenpenny, 1995). According to this abstractionist view, spoken language is normalized through a process in which information about a particular talker's voice is stripped away as the linguistic structure of speech is extracted. The nonlinguistic surface characteristics of speech are assumed to be perceived and encoded independently and separately from the abstract linguistic content. Thus, during perception, speech is stripped of any context-specific information, resulting in abstract linguistic representations that are processed and represented independently of the surface characteristics.

However, listeners do not appear to simply ignore or strip away the variability in speech produced by different talkers, but are sensitive to nonlinguistic or surface characteristics of language. Listeners use information about surface characteristics when processing linguistic information and retain surface characteristics in memory for spoken language (Bradlow, Nygaard, & Pisoni, 1999; McLennan & Luce, 2005; Nygaard, Burt, & Queen, 2000; Palmeri, Goldinger, & Pisoni, 1993). For example, spoken word processing is slower and less accurate (Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989), when lists are produced by multiple talkers than by a single talker, indicating that increased variability due to a multiple talkers' voices requires greater time and processing resources. Additionally, when listeners are presented with lists of words produced by multiple talkers, they are more accurate at recognizing previously presented items when they are repeated in the same, as opposed to different, talker's voice as during training (Bradlow, et al., 1999; Palmeri, et al., 1993), indicating that listeners retain perceptually-detailed information about talker's voice that affects later recognition memory.

Given these findings, an alternative to traditional accounts suggests that both the linguistic signal and the surface characteristics in speech are processed and retained together in perceptually detailed representations (Goldinger, 1998; Johnson, 1997; Jusczyk, 1993; Pisoni, 1993, 1997). According to this view, surface characteristics of spoken language influence the perception and processing of spoken language because both the linguistic and nonlinguistic properties of language are retained in memory together. Rather than simply constituting variability that would need to be stripped away, the surface characteristics of spoken language are an integral part of how we process and represent language. Sensitivity to changes in surface characteristics may shape linguistic processing, allowing listeners to infer information about the linguistic structure of an utterance. Thus, rather than presenting a perceptual problem for the listener, variability in the speech stream may contribute to the complexity and flexibility of the perception of spoken language.

The current study investigates how listeners cope with the variability inherent in spoken language by examining how listeners perceptually adapt to variation in the acoustic realization of spoken language. Perceptual learning may constitute one mechanism by which listeners perceive, categorize, and attribute variation in spoken language, allowing listeners to dynamically adjust their linguistic category structure to compensate for changes in surface form (Dupoux & Green, 1997; Francis, Baldwin, & Nusbaum, 2000; Francis, Nusbaum, & Fenn, 2007; Greenspan, Nusbaum, & Pisoni, 1988; Liss, Spitzer, Caviness, & Adler, 2002). In particular, the current study examines how listeners perceptually adapt to variation in accented speech. When listeners encounter non-native speakers of their language, they must learn to adapt not only to the normal range of variability due to talker's voice but also to the particular types of variability introduced by the foreign accent. Research on the perception of accented speech suggests that listeners can, in fact, become better at comprehending accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Weil, 2001), but little is known about *how* this adaptation is accomplished. This study addresses the question of what learning processes listeners might be using when they improve in their ability to understand accented speakers, which will speak to the issues of how listeners cope with variability in speech and use it to facilitate the perception of linguistic structure in spoken language. Examination of the mechanisms and processes involved in perceptual learning of accented speech can inform questions of phonological and linguistic representation, questions of perceptual learning more generally, as well as questions of the nature of behavioral and neurological plasticity.

Perceptual Learning

Perceptual learning is typically defined as learning that takes place implicitly (Hall, 1991) and that produces long-lasting changes in the ways in which perceptual features are perceived and categorized (Goldstone, 1998; Helson, 1948). Goldstone (1998) stated that perceptual learning is, by definition, a process that produces benefits by using information from experience to tailor processes for future information gathering. By changing as a function of experience, perceptual systems stay attuned to the kinds of input received and allow for plasticity and flexibility in the ways in which organisms respond to their environments. Additionally, Goldstone outlines several mechanisms assumed to be involved in perceptual learning in general, and which may be involved in perceptual learning for surface characteristics of spoken language. Potential mechanisms include *differentiation*, or the development of the ability to distinguish previously indistinguishable percepts; *unitization*, which refers to the process of integrating parts of a stimulus into whole units to facilitate processing; *stimulus imprinting*, which is the development of processes to facilitate stimulus processing in whole or in part; and *attentional weighting*, or the increase or decrease of attention to aspects of a percept. These mechanisms have important implications for issues of how new stimuli are processed, represented, and categorized.

Processes of differentiation and unitization play important roles in spoken language processing and specifically in the processing of surface characteristics of speech. Perceptual learning of phoneme categories can rely heavily on differentiation, especially in situations in which listeners must learn to distinguish two previously confused phonological categories. For example, Japanese speakers do not typically distinguish the English /r/ and /l/ minimal pair, but through perceptual learning, they can improve their ability to make this distinction (Iverson, Hazan, & Bannister, 2005; J. S. Logan, Lively, & Pisoni, 1991). Differentiation may also play a role in the perceptual learning of accented speech as listeners learn to differentiate previously confused linguistic categories. Sidaras et al. (2009) found that listeners were less likely to confuse Spanish-accented /I/ and /i/ after training, suggesting that listeners used information available from spectral and temporal regularities in the accented speech to differentiate the phonemes.

Unitization functions almost as a reverse of the differentiation process, in which listeners learn to associate perceptual properties or units that co-occur into a whole that is processed as a single unit. For example, linguistic stimuli are, by nature, hierarchical, building sentences from words, which are in turn built from phonemes strung together in particular ways. It is possible that listeners could process and represent spoken language at any of these levels; however, evidence from studies of word perception indicates that listeners respond more quickly to words than nonwords (Salasoo, Shiffrin, & Feustel, 1985). Since listeners are more familiar with words, they can be expected to respond more quickly to the more familiar stimuli. Interestingly, Salasoo et al. (1985) found that when listeners are repeatedly presented with nonwords, they perceptually adapt to these new items and process them as quickly as words, showing evidence of unitization occurring during perceptual learning. Similarly, as adults acquire new vocabulary, unitization allows novel words to become processed as whole units rather than a series of phonemes (Leach & Samuel, 2007; Tamminen & Gaskell, 2006).

Stimulus imprinting refers to the development of specialized processes for the perception of particular items or perceptual features, producing a benefit for perception of repeated stimuli. Stimulus imprinting can occur for whole stimuli, perceptual features, or complex relations of stimulus properties. In spoken language, stimulus imprinting can account for many of the effects of perceptual learning of talker specific characteristics of speech. For example, listeners have better recognition memory for items repeated in the same voices rather than different voices (Bradlow, et al., 1999; Nygaard, et al., 2000; Palmeri, et al., 1993) and are better able to identify words presented in noise when they are presented with familiar rather than unfamiliar voices (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994), indicating that listeners retain perceptually-detailed information in memory. The kind of whole stimulus imprinting seen in listeners' abilities to form perceptually detailed linguistic representations is frequently cited as evidence for exemplar-based, or instance-based, models of speech perception (Goldinger, 1996, 1998;

Johnson, 1997). In an exemplar-based theory, instances of a percept are retained in memory such that increased experience with similar instances facilitates retrieval of the information and decreases the time needed for processing (G. D. Logan, 1988).

Attentional weighting, or the change in amounts of attentional resources devoted to a particular perceptual property, may also play a role in the perceptual learning of spoken language. Research on perceptual weighting indicates that some degree of cognitive load must be present for selective attention to be employed (Lavie, 1995), and the presence of cognitive load can change attentional weighting, resulting in perceptual learning. As mentioned previously, listeners are inundated with variability in the speech stream originating from numerous sources. For listeners to readily recover the linguistic structure of a word given finite cognitive resources, they must be able to devote attention to relevant sources of variation, such as talker's voice or speaking rate, while ignoring or attending to a lesser degree to irrelevant sources of variation, such as overall amplitude (Bradlow, et al., 1999; Nygaard, et al., 2000). Perceptual learning may allow listeners to assess the range and structure of variation in a highly multidimensional space in order to facilitate speech processing. Attentional processes almost certainly play an important role in perceptual learning of accented speech as well. Sidaras et al. (2009) found that training with Spanish-accented speech led to more accurate word identification for Spanish-accented speech, but there was no difference in identification for familiar or unfamiliar accented talkers, indicating that listeners may be attending to the more relevant variation due to accent while devoting less attention to idiosyncratic talker-specific variation.

Changes in attention can also alter the nature of the categories that underlie perception by making members of a category seem more like one another and members of different categories seem more distinct (Goldstone, 1994, 1995, 1998). Listeners may adapt their phonological categories based on experience with variability in the speech signal, which informs questions of language representation and language learning in particular and of perceptual learning in a complex, multidimensional space more generally. As listeners gain experience with the variability present in speech, perceptual learning may change listeners' categories, representations, or processing abilities, allowing for improvement in speech comprehension.

Perceptual Learning for Surface Characteristics of Spoken Language

Much of the relevant evidence for perceptual learning of surface characteristics of speech arises from research on how listeners adapt to differences among talkers. With all the information that can possibly be conveyed through nonlinguistic information in speech, no two talkers speak in quite the same way. Properties of voices, such as F0 range and speaking rate, show idiosyncratic variation from person to person, resulting in highly multidimensional differences among talkers. Listeners are sensitive to the properties that distinguish one talker's voice from another (Allen & Miller, 2004; Bradlow, Torretta, & Pisoni, 1996; Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985) and appear to retain surface characteristics of talker's voice in memory (Goldinger, 1996; Goldinger, Pisoni, & Logan, 1991; Palmeri, et al., 1993). For example, in a continuous recognition memory task, Palmeri et al., (1993) found that listeners more accurately identified repeated items when words were presented in the same versus a different voice. In an implicit memory task, Goldinger (1996) found that words mixed with noise were transcribed more accurately when repeated in the same as opposed to a different talker's voice. Similar results have been obtained for talker variability, as well as other forms of surface from, in other recognition memory tasks (Bradlow, et al., 1999; Church & Schacter, 1994; Goldinger, 1996; Nygaard, et al., 2000).

In addition to retaining surface characteristics of talker's voice in memory, listeners perceptually adapt to the voice-specific properties of spoken language, showing facilitation in processing for speech produced by familiar talkers. For instance, after listeners were familiarized with a set of talker's voices in a talker identification task, they were better able to identify words mixed with noise when presented in familiar than unfamiliar voices (Nygaard & Pisoni, 1998; Nygaard, et al., 1994). That listeners use information about talker's voice to facilitate processing of words presented in noise indicates that attending to and adapting to surface characteristics of spoken language helps listeners contend with the variability present in natural speech.

Although there is a great deal of variability from one talker to the next, there are sources of variation that crosscut groups of talkers as well. This leads to systematic variation in which groups of speakers share particular characteristics that differ from other groups. For example, regional dialects result in differences in speech for groups such as speakers of British and American English. In many dialects of British English, the /r/ sound following a vowel is routinely not pronounced, but most speakers of American English dialects do pronounce the postvocalic /r/. This characteristic difference in pronunciation crosscuts the individual variation due to particular talkers' voices. Listeners are sensitive to systematic variation between groups of speakers, such as the differences in pronunciation resulting from dialectal variation (Floccia, Goslin, Girard, & Konopczynski, 2006). Floccia et al. (2006) examined the differences in spoken word processing for familiar and unfamiliar dialects and found that listeners had faster reaction times on a lexical decision task for familiar regional dialects than for unfamiliar dialects, indicating that familiarity with a regional dialect facilitates speech processing.

Systematic variation can also be produced through artificially altering speech in specific ways from natural speech, and perceptual adaptation has been found for artificially altered speech in addition to natural speech (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Dupoux & Green, 1997; Fenn, Nusbaum, & Margoliash, 2003; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998; Schwab, Nusbaum, & Pisoni, 1985; Sebastian-Galles, Dupoux, Costa, & Mehler, 2000). Davis et al. (2005) examined perceptual learning of noise-vocoded speech, which is computer-altered speech that is created by applying the averaged amplitude envelope of six frequency bands in natural speech to band-limited noise. This technique produces speech that resembles the signal characteristics of cochlear implant processors and preserves little detailed spectral content. Davis et al. (2005) found over the course of training, that listeners perceptually adapted to the noise-vocoded speech particularly in training paradigms in which lexical, syntactic, and semantic constraints were provided. In addition, listeners were better able to recover the linguistic content when hearing noise-vocoded speech derived from speech produced by familiar than unfamiliar talkers. Listeners also perceptually adapt to computer-generated, or synthetic, speech (Francis, et al., 2007; Greenspan, et al., 1988; Schwab, et al., 1985). Schwab et al. (1985) found that listeners initially have great difficulty understanding synthetic speech but that they show improved ability to transcribe words in synthetic speech with increased exposure. Francis et al. (2007)

suggest that, in the case of synthetic speech, listeners learn to use reliable cues in the synthetic speech stream to facilitate processing, even though those cues may be very different from the types of cues they typically use when processing natural spoken language. Although noise-vocoded and synthetic speech are different from natural speech in that the variation listeners are accommodating is extremely systematic and, in some ways, less complex than that of natural speech, that listeners show perceptual adaptation to cues that differ so greatly from those they typically use suggests that there is remarkable flexibility in listeners' acoustic-phonetic representational system. Listeners are able to accommodate many changes in the speech system, both natural and artificially created, in order to reduce processing time and effort and to maximize comprehension and other communicative goals.

Researchers have also examined perceptual learning of dysarthric speech, the disordered speech of individuals with neurologically based motor impairments, to investigate how listeners might adapt to the complexities of natural speech that contains both overarching and idiosyncratic forms of variation. Liss et al. (2002) found that there was systematic variation based on the particular motor deficits of patients with different types of dysarthria as well as variation among individuals with a single deficit type, and even inconsistencies within the speech of a single individual. Listeners appeared able to parse these multiple sources of variation when perceptually adapting to dysarthric speech, generalizing to novel utterances and talkers (Spitzer, Liss, Caviness, & Adler, 2000), but only for a particular type of motor deficit (Liss, et al., 2002). That listeners could generalize to novel talkers who shared a deficit type indicates that perceptual learning can occur for groups of accented talkers who have similar properties of speech production.

The adaptation to the deficit-specific properties in the presence of multiple levels of variation that are present in dysarthric speech provide evidence that perceptual learning can occur for groups of talkers who share systematic properties of spoken language.

Although the presence of multiple kinds of variability in natural speech make the process of investigating perceptual learning more difficult to control, it adds a great deal of ecological validity. The interleaved nature of variability in the natural system may have important implications for the types of mechanisms that may be used in the perceptual learning process. As listeners encounter multiple levels of variability in natural speech, listeners may differentially attend to properties that provide the greatest gain in terms of minimizing processing costs and facilitating comprehension for the group of talkers. Although evidence from synthetic and altered speech illustrates the great deal of flexibility in the perception of spoken language, it remains unclear how listeners contend with the multiple levels of variability found in natural speech. To investigate what mechanisms may be involved in learning multiple levels or types of systematic variability in the speech stream and what changes in underlying categorization processes may result, it is necessary to observe perceptual learning in listeners as they encounter variability present in natural speech.

Accented Speech

Accented speech provides an example of natural variation in speech, which has multiple levels of variability and that listeners encounter routinely in a variety of communicative settings. In its most general sense, accented speech refers to the differences between two groups of people in the pronunciations of sounds and words within a language (Yule, 2006). Accented speech reflects phonological properties (but not vocabulary) of a particular regional dialect or a particular first language (L1). Although adults can become fully fluent in a second language (L2), they tend to have more heavily accented speech than those who learn a second language during childhood (Flege, Yeni-Komshian, & Liu, 1999; Guion, Flege, Liu, & Yeni-Komshian, 2000). Adult speakers of an L2 have a difficult time learning new phonological categories for their second language and use close approximants from their L1 instead. For example, because Spanish and English consonants differ in voice onset time (VOT), a Spanish speaker might use Spanish VOTs when speaking English, instead of native English VOTs. Thus, even fluent speakers of a second language may have speech that is distinctly colored by the phonology of their first language.

For those speakers using an L2 as adults, the differences in phonological properties in their accented speech can produce barriers to communication with native listeners. Lane (1963) found that listeners were less accurate in a lexical transcription task when presented with foreign-accented speech than when presented with native speech, at various signal to noise ratios. Other research has shown deficits in intelligibility of accented speech in both transcription tasks (in noise and in the clear) and intelligibility ratings (Burda & Hageman, 2005; Munro & Derwing, 1995, 1998, 1999; Sidaras, Alexander, & Nygaard, 2009; van Wijngaarden, 2001). Increased processing time relative to native speech, another indication of difficulty in processing accented speech, has been found during the transcription of accented speech (Munro & Derwing, 1995) and when listeners made judgments about mispronunciations in accented sentences (Schmid & Yeni-Komshian, 1999). Processing accented speech seems to be a difficult task for listeners due in part to the phonological variation of accented speech, especially when placed in context with other variation from talker's voice and situational factors. My research seeks to investigate what processes listeners use to compensate for the phonological differences present in the speech of non-native speakers and how experience with non-native speech affects language processing and representation.

Like dysarthric speech, accented speech includes both the idiosyncratic talker characteristics that differ among individuals as well as the phonological characteristics common to the group of speakers who share the accent. Research on perceptual adaptation to accented speech offers a unique scenario to understand how people alter their representations of language to incorporate new experience with speech that has systematic variability in the phonological structure of the language. Although theories of perceptual learning in speech must account for a listener's ability to learn both types of information, the information necessary for learning accents and the mechanisms involved may be different from those necessary for talker learning. If listeners are able to engage in perceptual learning to facilitate the processing of accented speech, they would need to correctly apportion variability to accent and to talker's voice, maximizing comprehension *across* accented talkers. Little research has directly addressed the question of how native listeners learn to understand accented speech and the extent to which they generalize across talkers and words. Research on accented speech has mainly investigated what characteristics of the speakers and properties of speech itself make it less intelligible to native listeners (Burda & Hageman, 2005; Burda, Scherz, Hageman, & Edwards, 2003; Derwing, Rossiter, Munro, & Thomson, 2004; Munro & Derwing, 1995, 1998, 1999; van Wijngaarden, 2001), focusing less on the systematic properties that listeners may use to improve intelligibility and speech comprehension through perceptual learning.

More recent research has examined how listeners' perception of accented speech changes with experience. Clarke and Garrett (2004) examined rapid adaptation to foreign-accented speech. They presented listeners with accented sentences and then asked them to judge whether a word presented visually matched the last word of the accented sentence. They found that with experience, listeners' responses became faster and that this rapid adaptation occurred with exposure to as few as two to four accented sentence-length utterances. In an early study of learning and accented speech, Gass and Varonis (1984) found that listeners improved in their ability to transcribe accented sentences after listening to a brief paragraph in the same accent. Other research has found similar results of training in different paradigms, such as talker-identification training (Clarke, 2000) and transcription training (Bent & Bradlow, 2003; Bradlow & Bent, 2008; Sidaras, et al., 2009; Weil, 2001). Bradlow and Bent (2008) trained listeners on sentence-length utterances produced by multiple accented talkers. They found that when tested with a single accented talker, listeners generalized their learning to a novel talker with the same accent, but not to a novel talker with an unfamiliar accent. This pattern of generalization parallels findings that listeners generalize within but not across types of dysarthric speech (Liss, et al., 2002).

Taken together, these studies suggest that listeners adapt to accented speech and become better able to recognize the linguistic structure of accented words and sentences with training. Although Bradlow and Bent (2008) found evidence for generalization of accent learning, they only assessed generalization to a single novel talker. Clarke (2000) tested trained listeners on multiple accented talkers and found that training with accented voices did not generalize to a group of new talkers with the same accent. As mentioned previously, to adapt to the systematic properties of accented speech, listeners must learn which properties of an accent can be attributed to the accent group as a whole and which are due to idiosyncratic properties of talker's voice. If listeners are actually learning the properties of the accent, and not simply learning idiosyncratic properties of the talker's voice, they should generalize both to novel words and to multiple novel voices after training with a specific accent. Sidaras et al. (2009) trained listeners in a relatively short transcription-training paradigm with words or sentences spoken by accented speakers. At test, listeners transcribed utterances produced by either multiple same or different accented talkers. Listeners in control conditions either heard native English speakers during training or received no training at all. At test, listeners generalized to both novel utterances and multiple novel voices for both words and sentences. If trained listeners had shown training benefits for familiar but not unfamiliar accented talkers, it would have suggested that they were learning specific information about the idiosyncratic properties of the talkers' voices and not necessarily the accent-general properties shared by the talkers. However, that listeners generalized to novel unfamiliar voices with a familiar accent suggests that they learned information about the shared characteristics of the accent on which they were trained.

Although limited, the research on perceptual learning of accented speech has also varied the kinds of stimuli presented during learning. Some studies have trained listeners with sentence-length utterances (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Gass & Varonis, 1984; Weil, 2001) and found consistent evidence for perceptual learning, while other studies have trained listeners with isolated words with mixed results. Weil (2001) tested listeners with sentences and with words and found evidence of learning for sentence-length utterances, but not for word-length utterances. However, Sidaras et al. (2009) trained and tested listeners with both sentences and words, and found evidence for learning with both types of stimuli. When listeners are trained with sentence-length utterances, they may be using information in accented speech gleaned from multiple levels of linguistic and non-linguistic structure. In addition to lower-level phonological information, sentences contain prosodic contours that vary as a function of accent as well as contextual semantic and syntactic information that would constrain the recovery of the linguistic structure of the utterance. Thus, training listeners with sentences makes it difficult to ascertain what aspects of the accented speech listeners may be learning. However, that listeners can learn accented properties present in isolated words suggests that learning occurs at the acoustic-phonetic level. Sidaras et al. found that learning generalized to novel words, indicating that listeners learned something about the properties of individual phonological categories that could be applied to novel words.

Perceptual adaptation at the acoustic-phonetic level would facilitate adaptation to systematic variability in the speech stream originating from multiple sources. Not only would such a system be sensitive to large-scale regularities such as those present in accented or time-compressed speech, but it would also be sensitive to lawful variation due to idiosyncratic properties of talker's voice or situation-dependent factors. Since spoken language contains hierarchical forms of variation, with the idiosyncratic properties of talker's voice placed within the overarching properties that apply to many speakers with the same accent, any system of perceptual learning for spoken language must be able to contend with the variability in particular speech sounds originating from multiple sources. Perceptual learning at the acoustic-phonetic level would allow listeners to compare the accented acoustic-phonetic realizations of preexisting phonological categories across utterances to allow for learning of a particular talker's voice in different phonetic contexts as well as the similarities among utterances from a group of accented talkers.

Research on perceptual learning of accented speech using highly variable natural stimuli suggests learning can occur at the acoustic phonetic level. Converging evidence comes from studies that have begun to examine perceptual learning at the acoustic-phonetic level with highly controlled stimuli. Using highly constrained learning paradigms, researchers have examined whether changes occur in category boundaries and/or structure (Kraljic & Samuel, 2006, 2007; Norris, McQueen, & Cutler, 2003) as the result of perceptual learning. As mentioned previously, the potential consequences of perceptual learning include changes in perceptual category structures for spoken language, with category members seeming more similar to one another and members of different categories beginning to seem more distinct (Goldstone, 1994, 1995, 1998). Thus, novel pronunciations could be incorporated into existing categories, even though their inclusion may produce changes in the category structure or boundaries as more novel instances are learned.

Some recent evidence for perceptual adaptation at the acoustic-phonetic level comes from studies in which listeners were presented with words which differed from native speech along specific acoustic-phonetic properties and showed learning specific to the particular properties and phonetic context (Dahan, Drucker, & Scarborough, 2008; Maye, Aslin, & Tanenhaus, 2008; Sidaras, et al., 2009). For example, Maye et al. (2008) familiarized listeners with an artificial accent created by lowering the front vowels by one vowel unit, a change consistent with the types of shifts in vowel production seen historically. Listeners showed specificity of learning, accepting items with lowered vowels as words that previously would have been considered nonwords. Additionally, there was evidence of generalization in that trained listeners also endorsed items with lowered back vowels, which is consistent with the properties of the training materials even though no lowered back vowels were presented during training. However, when tested with items containing raised back vowels, participants did not endorse them as words, perhaps indicating that listeners learned specific information about the acoustic-phonetic properties of the accent. Similarly, Dahan et al. (2008) found evidence that listeners perceptually adapted to a series of words in which the vowel /æ/ was raised when presented before /g/, but not /k/. Listeners not only learned to identify the accented vowel as a valid pronunciation of the native sound, but they also learned that the accented speech sound occurred in a particular context. These findings indicate that listeners presented with accented speech learn something about the systematic properties that vary for that particular accent or dialect as well as how the accented speech maps on to familiar acoustic-phonetic categories.

Evidence for adaptation at the acoustic-phonetic level has been shown in research from my own lab using high variability natural stimuli. Sidaras et al. (2009) found that listeners readily adapted to certain accented vowels after brief training but showed little learning for other vowel categories. In this case, listeners may have learned those acoustic-phonetic cues that were most reliable, since accented speech was found to be generally more variable than native speech, training was brief, and it included a variety of consonant environments for the vowels. In the case of natural accented speech, unlike the synthetic and highly regular accent constructed by Maye et al. the cues available in natural accented speech are not necessarily perfectly consistent across groups of vowels or even across all instantiations of a particular vowel. The vowel-specific learning found by Sidaras et al. (2009) indicates that with limited experience, listeners may have learned to attend to the most reliable acoustic-phonetic cues, neglecting other, more widely variable, information. Of course, listeners should be able to adapt to more variable vowel categories with increased experience, as long as the variation is systematic for the group of accented talkers.

If listeners are perceptually adapting to accented speech at the acoustic-phonetic level, then learning should produce changes in the categorization process and the underlying representations for the acoustic-phonetic properties. Changes due to perceptual learning in how listeners categorize speech sounds have been found with ambiguous phonemes presented in words (Clarke-Davidson, Luce, & Sawusch, 2008; Kraljic & Samuel, 2006, 2007; Norris, et al., 2003). Norris et al. (2003) familiarized listeners with ambiguous phonemes halfway between an /f/ and an /s/ embedded in lexically constraining contexts such that the ambiguous phonetic segments either appeared in words biasing toward /f/ (e.g., chef) or toward /s/ (e.g., pass). In a subsequent phonetic categorization task, the context in which the ambiguous sound was presented during familiarization induced changes in the /f/ and /s/ category boundaries to include a sound that would normally be fully ambiguous. Kraljic and Samuel (2006) found similar alterations in categorization despite a change in talker's voice between familiarization and test, indicating that the information learned about the ambiguous phoneme generalized to a new talker. However, Kraljic and Samuel (2007) found that listeners' categorization of different phoneme contrasts interacted with effects of talker's voice. Listeners showed less talker-to-talker generalization for fricatives, indicating that they formed talker-specific representations, and they shower a greater degree of generalization for stop consonants, suggesting that they may have formed more general or abstract representations in that case. Changes in the categorization of phonetic segments as the result of perceptual learning may be based on the acoustic-phonetic properties of individual sounds and to what degree those properties vary as a function of individual or group pronunciation differences. In the case of the specificity and generalization of talker-specific effects found by Kraljic and Samuel (2007), listeners may have shown more specificity when contrasts included talker-specific information such as spectral cues in fricatives and more generalization when there were fewer talker-specific cues as with stops.

If perceptual learning of accented speech produces specific learning for a particular accent, it should be reflected in how listeners' linguistic categories change as a function of learning. Specificity of learning could result from a number of underlying changes in categorization processes. One process that might be driving category changes is that listeners may be learning to map acoustic-phonetic features of the accented phonemes onto pre-existing linguistic categories in a process similar to that proposed by Francis et al. (2007) for synthetic speech and Liss et al. (2002) for dysarthric speech. Learning to make the mapping from accented speech onto existing acoustic-phonetic categories would allow the accented phonemes to be considered members of those categories in future encounters. In this case, listeners would be forming representations of the accented speech, adding new accented exemplars to existing phonological categories.

Clarke-Davidson et al. (2008) found that perceptual learning of an acoustic-phonetic property not only changed boundaries between categories, but also changed the category structure as a whole, suggesting that listeners incorporate the new exemplars into existing categories. In the case of accented speech, it may be that a mapping of accented items into native categories might produce a kind of accent-specific filter for processing a particular accent. The change in categories would result in a fine-tuning of the perceptual system to allow only those properties specific to the particular accent to be considered category or subcategory members, which would support previous research showing generalization to novel talkers with shared characteristics, but not between groups of talkers with different accents or speaking styles (Bradlow & Bent, 2008; Liss, et al., 2002). Additionally, a change based on perceptual learning of accent-specific information is similar to the findings of Maye et al. (2008) in which listeners showed generalization to previously unlearned lowered back vowels which was consistent with the changes learned for front vowels, but they did not generalize to raised vowels, which was not consistent with any learned information. If perceptual learning of speech results in accent-specific categorization such that only tokens with the specific properties of the learned accent are accepted as category members, then perceptual learning should be accent-specific, only generalizing to speakers with the same or similar accents.

Alternatively, listeners may broaden their category boundaries to include examples of speech that are generally different from native speech. Research has shown that accented talkers produce more variable utterances at the acoustic-phonetic level than do native speakers (Lee, Guion, & Harada, 2006; Nygaard, Sidaras, Duke, & Rasmussen, 2006; Schmidt & Flege, 1996). If listeners preserve this non-systematic variability present in accented speech in their acoustic-phonetic representations, then they may show evidence of generalization to non-related accents because the preservation of non-systematic variation could potentially overlap with similar non-systematic variation found in the utterances of speakers from many different L1 groups. According to this view, listeners might still be mapping the acoustic-phonetic properties of accented speech onto pre-existing categories, but the preservation of non-systematic variation might result in the expansion of the category boundaries, rather than the formation of an accent-specific filter or subcategory. Category broadening could occur through the mapping of extremely varied tokens of accented speech onto an existing phonological category through supervised learning. This would cause the boundaries of the category to expand to include the many, varied properties encountered during learning. If this is the case, listeners should be expected to show more of this type of category broadening when the information learned has less systematic variation. When highly systematic information is presented, listeners might attend to only the most diagnostic criteria; however, when the input is more variable, listeners might preserve information including a greater number of cues that, in turn, might prove useful in facilitating processing of accented speech. Thus, listeners might broaden their categories such that experience with a variety of different kinds of accented speech would facilitate processing of virtually any L1 group of accented speakers.

In support of the idea that listeners may expand their category boundaries when learning accented speech, recent research has shown an intelligibility benefit for non-native speakers of English when listening to accented speech from a group of accented speakers from an L1 accent group different from their own (Bent & Bradlow, 2003). Bent and Bradlow found that non-native speakers of English transcribed accented speech produced by speakers with accents different from their own as accurately as they transcribed both speech produced by native English speakers and accented speech produced by other speakers of their native language. Although it is unclear what accounts for this interlanguage intelligibility benefit, these listeners, who were themselves accented speakers, may show greater intelligibility for accented speech regardless of whether they share an L1 with the speaker, because they are likely to have encountered a variety of accents and degrees of intelligibility while learning their L2 and immigrating to a different language culture. However, it is unclear to what degree the interlanguage intelligibility effect extends to different levels of proficiency or intelligibility and other language pairings (Stibbard & Lee, 2006). Stibbard and Lee (2006) found that although high-proficiency talkers were equally or more intelligible to some native and nonnative listeners, low-proficiency talkers were less intelligible than native speakers and had the lowest intelligibility for nonnative listeners from other L1 groups. While the finding that high-proficiency talkers are highly intelligible is consistent with those found by Bent and Bradlow (2003), Stibbard and Lee (2006) found that the effect changed based on the language pairing. In their study, native Korean listeners rated high proficiency Saudi Arabian-accented English as less intelligible than native English or Korean-accented English. Thus, not only does the interlanguage intelligibility effect not seem to extend past the most proficient nonnative talkers, it is not present for all language pairings. The findings of Bent and Bradlow (2003) and Stibbard and Lee (2006) indicate that the relationship between perception of accented speech and interlanguage generalization may be dependent on talker proficiency or intelligibility and similarities between languages or specific talkers.

Regardless of the exact nature of category change, listeners must attend to and encode temporal and spectral properties of speech that do not typically support native language processing, but that produce reliable cues to comprehension in accented speech. For example, Francis and Nusbaum (2002) trained English listeners on Korean stop consonants and found that listeners could learn to attend to an initial change in F0, a cue that does not produce reliable category differences in English. In this situation, listeners' phonological categories may not only shift along dimensions previously used to identify native speech sounds, but may shift to new dimensions which were previously non-diagnostic. Learning to use novel characteristics of speech sounds could result in greater effects of specificity of learning than learning information that differs only slightly from the bulk of past language experience. Given the great amount and many types of variability present in natural accented speech, listeners' acoustic-phonetic categories may be changing in multiple ways to incorporate necessary information. Depending on what kinds of acoustic-phonetic properties a listener encountered during learning and what mechanisms of perceptual learning were most useful, learning could show effects of accent-specificity and/or produce facilitation in speech processing beyond the specific accent learned. The current study investigates effects of specificity and generalization in perceptual learning of accented speech and what particular kinds of categorical changes may be taking place due to exposure to accented speech.

The current study builds on the research of Sidaras et al. (2009) to address the relative specificity of perceptual learning of accented speech. In previous research, we have used accented speech from native speakers of Spanish. The current study replicates past findings of perceptual learning using a database of accented speech collected from a group of native Korean speakers. Both Spanish and Korean differ from English in a number of phonological and phonetic characteristics. These L1 differences seem to carry over into the accented-speech of the L2. For example, Spanish-accented speech systematically differs from native English speech in temporal and spectral characteristics of vowels (Sidaras, et al., 2009), such as the high degree of similarity of first and second formant values between Spanish-accented vowels /i/ and /ɪ/ (Flege, Munro, & Skelton, 1992; Sidaras, et al., 2009). The overlap in those two vowels is likely due to the fact that the Spanish vowel space contains the vowels /i e a o u/, but does not have equivalents of the English vowels /æ/, /ʌ/, /ɔ/, /ʊ/, or /ɪ/ (Nash, 1977). Spanish also differs from English in consonantal characteristics such as voice onset time (VOT) (Nash, 1977). Native Spanish-speakers tend to produce shorter VOTs for stop consonants, such as /p t k/ (Flege, Schirru, & MacKay, 2003), but early bilinguals tend to produce more native-like VOTs than late bilinguals (Flege & Eefting, 1986). Late Spanish/English bilinguals tend to produce more variable VOTs for stop consonants, sometimes producing shorter, Spanish-like stops, and sometimes produce long English-like stops (Schmidt & Flege, 1996). Korean differs from both English and Spanish in consonant pronunciation, exhibiting differences in place and manner of articulation (Nissen, Dromey, & Wheeler, 2004; Rice, 2002), which is found in Korean-accented speech as well (Nissen, et al., 2004). Korean-accented speech contains fewer differences among vowels in terms of duration than native English, and with regard to unstressed vowels, Korean-accented late Korean/English bilinguals tend to produce more variable first and second formant values than early bilinguals (Lee, et al., 2006). Like Spanish, Korean also has a different set of vowels than English. Some Korean vowels map easily onto English vowels (/i e ɛ a ʌ o u/) in F1/F2 space, but there are English vowels, which do not have a clear Korean equivalent (e.g., /ɪ æ ʊ ɔ/) (Yang, 1996). Yang (1996) found that although some of the Korean vowels mapped onto English vowels, the F1/F2 values were less similar than for other Korean/English vowels (e.g., /u a o/). Korean-accented speech differs from native English in terms of consonants as well. For example, Korean-accented speakers do not produce releases on final voiceless stops, such as /t/ and /k/ the same way that native English speakers do (Tsukada, et al., 2004).

These differences found among native English, Korean-accented, and Spanish-accented speech produce the unique set of systematic regularities that listeners may use during perceptual adaptation. Based on the properties of the L1 as well as what is known about the particular accents of Spanish and Korean, the properties of each language shape the accented English in unique ways. That the L1 systematically changes the realization of English for accented speakers not only gives a group of accented talkers a common form of spoken language, but it provides the listener with systematic variation that can be learned through experience. One commonality in studies of Spanish- and Korean-accented English is that late bilinguals tend to produce more temporal and spectral variability than early bilinguals (Lee, et al., 2006; Nygaard, et al., 2006; Schmidt & Flege, 1996), which may change the degree to which listeners can learn systematic information.

Assuming listeners should generalize from one accent to another to the degree that the specific acoustic-phonetic properties are similar, two distinct accents were needed to test the hypothesis that perceptual learning for accented speech is accent specific. Spanish- and Korean-accented speech appear to have little acoustic-phonetic overlap, providing two perceptually distinct accents with which to train and test listeners. For two languages that are very similar (e.g., Spanish and Italian) in the acoustic-phonetic properties of the phonemic inventory, accented properties might overlap, potentially facilitating generalization of learning.

To examine the process of how listeners perceptually learn accented speech, Experiment 1 investigated perceptual learning for Korean-accented speech, and Experiment 2 used multiple accents to examine specificity in learning. In Experiment 1, participants were trained and tested on Korean-accented English words to replicate and extend the finding of perceptual learning to another, very different first language group. Using accented speech from both Spanish and Korean speakers, Experiment 2 investigated whether listeners are learning *accent-specific* properties of language or are simply becoming better at processing accented speech regardless of the native language of the speaker. Listeners were trained with a group of speakers from one of three language groups: Spanish-accented English, Korean-accented English, or English produced by a group with a variety of different accents. After training, each group was tested on their transcription performance for novel Spanish-accented or Korean-accented English words from unfamiliar talkers (see Figure 4). For listeners trained on Spanish-accented speech, the Spanish-accented English was the familiar accent, and Korean-accented English was an unfamiliar accent. For listeners trained with Korean-accented speech, the Korean accent was familiar and the Spanish accent was unfamiliar. For the Mixed accent training group and the no training control group, both Spanish and Korean accents were unfamiliar. If listeners alter their phonological categories to reflect the properties of the particular accent they heard at training, the listeners who received Spanish-accented training should show better transcription performance for Spanish- than for Korean-accented speech. Listeners who received Korean-accented training should show better transcription performance for Korean-accented speech than for Spanish-accented speech. The group who received Mixed accent training may perform relatively poorly when tested with an unfamiliar accent, especially if listeners are unable to transfer any learned properties to the accent presented at test. However, the Mixed training group should also gain experience with a large variety of different kinds of cues that may not normally be diagnostic in native speech. The varied experience might allow listeners receiving Mixed training to become sensitive to many new diagnostic dimensions, producing facilitation for other accents that share any subset of those properties. If participants broaden their phonological categories to include both changes in known dimensions and new diagnostic criteria, then listeners who received any kind of accentedness training (Spanish, Korean, or Mixed) may show better transcription performance than control listeners who received no accentedness training.

Experiment 1

Experiment 1 examined how listeners perceptually adapt to English spoken by native speakers of Korean. In past research using Spanish-accented English, listeners who received training performed better for both familiar and unfamiliar accented talkers than those listeners who did not receive training (Sidaras, et al., 2009). In the current experiment, perceptual adaptation to Korean accented speech was examined both in order to determine if listeners perceptually adapt to accented speech produced by non-native speakers of a language that is less familiar to English speaking participants than Spanish accented speech, as well as to establish perceptual learning with an accent other than Spanish for comparison in the second experiment. In previous studies, English-speaking listeners adapted to the systematic variation present in Spanish accented speech over the course of training. Similarly, when native-English speaking listeners are trained on Korean-accented English, they should perform better on novel tokens for both familiar and unfamiliar talkers than those who did not have training with accented speech. This finding would indicate that listeners perceptually adapt to accented speech from Korean-accented speakers, and that this adaptation similarly generalizes to novel speakers and novel utterances.

In addition to examining perceptual learning of accented speech produced by non-native speakers of a language that is less familiar to the listeners and includes different acoustic-phonetic properties than Spanish, Experiment 1 examines the role of lexical characteristics in perceptual learning. To that end, the stimuli consist of *easy* and *hard* words produced by Korean-accented speakers. Easy and hard words differ on characteristics of neighborhood density and word frequency (Luce, 1986; Vitevitch & Luce, 1998). Neighborhood density refers to the number of words that differ from a target word by a single phoneme substitution, deletion, or addition (e.g., *kiss* and *miss*). Word frequency is simply the frequency with which particular words can be found in normal usage (Kucera & Francis, 1967). Easy words are high frequency words with relatively few, low frequency neighbors, such as *work,* which has few low frequency neighbors (e.g., *shirk, lurk,* and *murk*.) Hard words are low frequency words with a great number of high frequency neighbors, such as *bead,* which has many high frequency neighbors (e.g., *beat, lead,* and *feed.)* Lexical characteristics such as word frequency and neighborhood density have been shown to affect the time course and accuracy of spoken word recognition (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Using easy and hard words during training and test allowed for evaluation of possible effects of lexical neighborhood or word frequency on learning. Examining the effects of lexical characteristics on the perceptual learning process can yield more information about the ways in which listeners represent phonological information. Words with many high frequency neighbors will be more difficult to distinguish from other words and may cause difficulty in making correct mappings from the accented speech to existing acoustic-phonetic categories. If this is the case, listeners may differentially benefit from learning for *easy* and *hard* words. The use of easy and hard words in this study may help address questions of how lexical properties interact with perceptual learning of accented speech.

Method

Participants

Seventy-two listeners between the ages of 18 and 30 participated in Experiment 1. They were either paid $15 or received partial course credit in an introductory psychology class for their participation. All participants were native speakers of American English and had no reported history of hearing or speech disorders. Listeners were also screened for prior experience with or exposure to Korean and for frequent exposure to Korean-accented speech, and all participants with prior experience were excluded from participation.

Materials

Twelve native Korean speakers (6 males, 6 females) living in the Atlanta area were recorded reading a list of 144 easy and hard words (72 easy words and 72 hard words; see Appendix). The mean word frequency for the easy words is 312.13 and for the hard words is 10.75, and the mean neighborhood density for the easy words is 13.54 and for the hard words is 26.75. Word frequency was calculated using frequency norms from the *Brown Corpus of Standard American English*, a database of words from a wide variety of print sources (Kucera & Francis, 1967). Frequency-weighted neighborhood density was calculated using the sum of the log frequencies of all the neighbors (Luce, 1986; Vitevitch & Luce, 1998). Both easy and hard words were rated as highly familiar, with a mean familiarity of 6.97 for the easy words and 6.80 for the hard words (Nusbaum, Pisoni, & Davis, 1984).

Speakers were between the ages of 20 and 37, with the mean age of 26.7 years at the time of recording. To ensure the native Korean speakers spoke a similar dialect, all 12 speakers were born in South Korea and lived in or near Seoul before coming to the US. They all began learning English at a mean age of 13 years, with a range of 10 to 15 years of age. Their mean age of arrival in the US was 24.3, and the range of age of arrival was 15 to 35 years of age. All stimuli were recorded in a sound-attenuated room with a SONY Digital Audio Tape-corder TCD-D7. The recordings were re-digitized on an iMac and edited for presentation using Sound Studio 3. Baseline intelligibility measures and accentedness ratings were obtained for all speakers. Participants were presented with 144 words and 100 sentences from one of the native Korean speakers. A total of 120 participants (10 for each speaker) transcribed the words and sentences they heard.

Intelligibility scores were calculated for the proportion correct for words and sentences. The native Korean speakers had a mean intelligibility of 57.47% for words and 88.07% for sentences (see Table 1 for each talker's mean scores). The male speakers were less intelligible, with a mean of .55 proportion words correct, than the female speakers, with a mean of .60 proportion words correct. A total of 10 participants rated the accentedness of a set of 240 sentences (20 from each native Korean speaker) on a Likert-type scale of 1-7, with lower scores indicating less accentedness and higher scores indicating greater accentedness. The mean accentedness rating was 4.14. The mean for male speakers was 4.38, and the mean for female speakers was 3.91. Two talker groups of six native Korean speakers (3 males, 3 females) were formed with intelligibility and accentedness equated across groups. Group 1 had a mean intelligibility of .55 and a mean accentedness of 4.27. Group 2 had a mean intelligibility of .59 and a mean accentedness of 4.01. There were no significant differences between the groups on intelligibility, t (10) = .024, p = .981, or on accentedness, t (10) = .191, p = .852.

Procedure

Training phase. Stimulus presentation and data collection were controlled using EPrime software (Schneider, Eschman, & Zuccolotto, 2002). Auditory stimuli were presented binaurally over Beyerdynamic DT100 headphones at approximately 75 dB (SPL). Participants were trained in a single training session on words spoken by six of the native Korean-speakers (three males, three females) and then tested on 48 novel words. A No Training control group received no training, and only participated in the test portion of the experiment. Training consisted of two types of training blocks: 4 Comparison blocks, which contained 6 words (3 *easy* words, 3 *hard* words), each spoken by all six speakers; and 3 Variability blocks, which contained 24 words spoken by the six speakers in random order (12 *easy* words, 12 *hard* words), then repeated with a different set of word-voice pairings (See Figure 1). The Comparison and Variability blocks were presented in alternation for a total of 192 tokens across training. During each Comparison block, listeners heard each accented word twice and were asked to rate how accented that item sounded to them on a scale of one to seven, from not at all accented to heavily accented. During each Variability block, participants heard an accented word and were asked to transcribe the word they thought the speaker intended to produce. After each word transcription, they saw the correct answer on the screen and heard the word again to provide feedback on their transcriptions. Half of the participants were trained on speaker group 1, and half were trained on speaker group 2. The order of the training materials within the training session was counterbalanced across participants.

Sidaras, et al. (2009) used Comparison and Variability blocks in training to allow participants to have training with highly variable stimuli presented randomly with both items and talkers mixed as well as training with stimuli grouped by words so that listeners might compare the voices of the different talkers as they all produced the same item. These two types of training led to robust learning effects for Sidaras, et al., and may each contribute a unique environment for perceptual learning. Other research in my lab has compared these two types of learning conditions directly and found that listeners may employ attentional weighting differently when presented with stimuli organized by item, talker's voice, or randomly (Duke & Sidaras, 2006). In the current study, both Variability and Comparison blocks were used to maximize learning in the brief training paradigm.

Generalization test. After training, the test was administered. All participants heard and transcribed 48 novel words presented randomly (24 *easy* words, 24 *hard* words). Half of the participants who received training were presented words spoken by the six familiar talkers heard during training, and half heard words spoken by six unfamiliar Korean-accented talkers. Thus, for those listeners trained with Talker group 1, half were tested with Talker group 1 (familiar talkers, familiar accent), and half were tested with Talker group 2 (unfamiliar talkers, familiar accent). Listeners did not receive feedback on their responses during the test phase. The No Training control group completed only the generalization test. Half of the listeners in the control group heard one group of Korean-accented talkers, and half heard the other group. For all No Training listeners, both accent and talkers were unfamiliar.

Results and Discussion

Listener transcriptions were scored for accuracy, with misspellings and homophone spellings counted as correct (e.g., the response *cheif* was counted as correct for the word *chief*, and the response *sighs* was counted as correct for the word *size*). Mean transcription accuracy was calculated for easy and hard words in each training block and at test for each participant. Percent correct transcription averaged across participants in each condition is reported throughout. Means and standard deviations for training and test are reported in Table 2.

Training phase. To determine to what degree listeners improved in their transcription performance across the training blocks, transcription performance was analyzed using a repeated measures analysis of variance (ANOVA) with block (Blocks 1-3) and lexical properties (*easy* and *hard* words) as within-subjects factors. A significant main effect of lexical properties was found, F (1, 47) = 381.57, $p < .001$, *partial* η^2 =

.890, with listeners performing more accurately for *easy* words than for *hard* words. There was no main effect of training block, F (2, 94) = .15, p = .864, *partial* η^2 = .003, and no interaction between training block and lexical properties, F (2, 94) = .22, p = .799, *partial* η^2 = .005. Transcription performance did not seem to improve across blocks for either easy or hard words, and lexical properties did not interact with the learning process (See Figure 2).

Generalization test. To examine the degree to which training with accented speech facilitated transcription performance of novel words and talkers at test, a mixed factorial ANOVA was performed with Training Condition (Same, Different, and No Training) as the between-subjects factor and easy/hard words as the within-subjects factor (See Figure 3). There was a significant main effect of easy/hard words, F (1, 69) = 205.13, p < .001, *partial* η^2 = .748, with transcription performance for easy words significantly better than for hard words. There was a significant main effect of Training condition, F (1, 69) = 4.50, p = .015, *partial* η^2 = .115. Planned comparisons revealed no difference between Same and Different Korean-accented training conditions F (1, 69) = .02, p = .445, but listeners who received training with Korean accented speech performed significantly better at test than those who received no training, F (1, 69) = 8.97, p = .007 (see Figure 3). No interaction was found between the two variables, F (2, 94) = .224, p = .799, *partial* η^2 = .005.

These findings indicate that very brief training with Korean-accented speech allowed listeners to perceptually adapt to the systematic variation present in Korean-accented English. This adaptation facilitated transcription at test for Korean-accented novel words and novel talkers, indicating that listeners learned systematic information in

Korean-accented English to better extract the linguistic information from the utterances of novel talkers with that same accent.

With regard to lexical properties, a main effect of word type was found for both training and test, but it did not interact with training condition or training block. Although easy words were more accurately transcribed than hard words, there was little evidence of an interaction between the lexical properties of words and the perceptual learning process, at least in this brief training paradigm. This may indicate that with such brief training, perceptual learning is occurring at a sublexical level. It may be the case that with increased experience with accented speech or greater repetitions of particular lexical items, listeners would have the opportunity to employ mechanisms of unitization. If listeners could accrue more experience with particular lexical items, they would be able to form lexical representations of the accented words that would be processed as quickly as the unaccented words with which they are already familiar. Once listeners unitized the accented speech into word-level representations, processing effects of lexical properties might be seen.

Similarly, little evidence was found for the presence of talker-specific learning during training. Listeners tested with the same talkers heard during training performed no differently than those tested with different talkers. The lack of talker-specific learning is not surprising given that listeners received relatively more experience with the Korean accent overall than with any particular talker. This type of highly variable training paradigm may have drawn attention to properties of the Korean-accent that spanned all the speakers rather than the particular characteristics of any one talker's voice.

These findings replicate the perceptual adaptation to Spanish-accented words found in previous studies (Sidaras, et al., 2009), illustrating that listeners can perceptually adapt to accented speech from both Spanish and Korean accented speakers. In Sidaras et al. (2009), listeners were not speakers of Spanish, but data were not collected on the familiarity of listeners with Spanish or Spanish-accented speech. In the current study, listeners were uniformly unfamiliar with Korean and Korean-accented speech. Controlling for familiarity insured that the learning condition manipulation was not a product of previous experience with the particular accent. Additionally, Korean has a different set of phonological characteristics than Spanish does, and is, in turn, uniquely different from English. Learning with both Spanish- and Korean-accented speech provides evidence that listeners can adapt to very different sets of cues or features for the instantiation of phonological categories of English (Alexander, Nygaard, & Sidaras, 2008; Sidaras, et al., 2009).

Experiment 2 directly addresses the question of what listeners are learning during training with accented speech. Given that listeners have been shown to perceptually adapt to both Korean- and Spanish-accented speech, Experiment 2 assesses generalization and specificity of learning within and across specific accents. Listeners were trained with either Spanish-accented speech, Korean-accented speech, or speech from a group of accented talkers from multiple first language groups. Listeners were tested with either Korean- or Spanish-accented speech to address whether learning with one accent generalizes to a novel accent or whether learning is specific to the accent presented during training.

Experiment 2

Previous research has shown that listeners learn properties of accented speech that allow them to better understand both familiar and unfamiliar talkers with the same accent they were exposed to during training. However, it is unclear to what degree listeners are learning something specific about the systematic variation present in a particular accent or are generally expanding their criteria for what is considered a valid English speech sound. Rather than making shifts in their phonological categories toward particular accented characteristics of speech, listeners may instead be broadening their categories to include many or any possible variations on their native phonemes. If listeners are broadening their categories then training with any accent should produce a similar degree of perceptual learning as training with the particular accent.

In Experiment 2, specificity and generalization of learning were examined by manipulating what kind of accented training listeners received. Listeners were trained on one of three groups of speakers, Spanish-accented, Korean-accented, or a group including speakers from multiple first language groups which did not include speakers of Spanish or Korean. They were then tested on Korean- or Spanish-accented speech. If listeners learn the specific properties of the L1 accent group they are trained on, then they should show accent-specific benefits of training that may have little transfer across L1 accent groups. This means that listeners trained with Korean-accented English should perform best on the Korean-accented test, and listeners trained with Spanish-accented English should perform best on the Spanish-accented test. Those listeners trained with the Mixed-accented group or given no training should not perform as well on either the Spanish- or Korean-accented speech at test if learning is accent-specific. However, if listeners are broadening their phonological categories or relaxing their categorization criteria to accept many different types of sounds that map onto the same linguistic category, Mixed-accented training should show training benefits relative to the other training conditions and untrained controls. If this is the case, Spanish- and Korean-accented training will still produce good performance at test for the same accent heard during training, but Mixed-accented training should produce similar or better performance. This benefit of Mixed-accented training would be due to the greater variability present in the speaker set at training, which should, in turn, increase range of variation that was learned to map onto existing linguistic categories.

Method

Participants

One hundred-sixty listeners between the ages of 18 and 30 participated. They were either paid $15 or given partial course credit in an introductory psychology course for their participation. All participants were native speakers of American English with no reported history of hearing or speech disorders.

In order to directly compare listener familiarity with Korean- and Spanish-accented speech, listeners were asked about their familiarity with accented speech in various contexts. They were asked to report their experience with close friends or family members who spoke with an accent from birth until high school, during high school, and during college. They were also asked to report their experience with teachers or roommates who spoke with an accent. Out of 160 participants, 105 reported learning Spanish at some point in the past. Of those, 96 reported beginning to learn Spanish in high school or before, and nine reported learning Spanish during college or later. The mean number of years of study was 4.3, with a range from less than a year to 11 years. Only two participants reported learning Korean, and both began studying it during college, one person for one year, and the other for two years. Listeners reported their contact with close friends or family members who spoke with some form of Hispanic accent: 17 reported exposure before high school, 31 during high school, and 40 during college. They also reported contact with teachers or roommates who spoke with a Hispanic accent: 15 reported exposure before high school, 36 reported exposure during high school, and 41 reported exposure during college. Listeners reported relatively fewer close friends or family members with a Korean accent: 2 reported exposure before high school, 6 during high school, and 16 during college. They also reported their exposure to Korean accented teachers or roommates: no participants reported exposure before high school, one reported exposure during high school, and 14 reported exposure during college. Across all time categories, 55 participants indicated they had close friends or family members with a Hispanic accent, 64 that they had Hispanic-accented teachers or roommates, 18 that they had Korean-accented friends or family, and 14 that they had Korean-accented teachers or roommates. Overall, the participants had a diverse range of experience with accented speakers, from many different language backgrounds. It seems that Spanish-accented speech is more familiar for the listeners in the current study than Korean-accented speech.

Materials

The Korean speaker stimuli used in Experiment 1 were also used in Experiment 2. Two other sets of accented stimuli were also used. The Spanish-accented stimuli from Sidaras et al. (2009) were used. Those stimuli consisted of twelve native Spanish speakers from Mexico City living in the Atlanta area. The Spanish speakers were recorded reading the list of 144 easy and hard words. The native Spanish speakers had a mean age of 32.75 years at the time of recording, with a range of 26-39 years of age. Their mean age of arrival in the US was 26.42 years, with a range of 21-34 years of age, and they had begun speaking English at approximately 16.67 years of age on average, with a range of 2-28 years of age. Spanish-accented stimuli were recorded in a sound-attenuated room with a SONY Digital Audio Tape-corder TCD-D7. The recordings were re-digitized on an iMac and edited for presentation using Sound Studio 3.

A third group of speakers with a variety of first languages (Mixed-accented speakers) were also recorded reading the 144 easy and hard words. These speakers were all members of the Emory University community and were living in Atlanta. The first languages of the speakers were from a variety of language families. They included speakers of Albanian, Dutch, Japanese, Romanian, Bengali, Hindi, French, German, Somali, Russian, Mandarin, and Turkish. They were also recorded in a sound-attenuated room with a SONY Digital Audio Tape-corder TCD-D7 and re-digitized on an iMac. The recordings were edited for presentation using Sound Studio 3. The Mixed accent speakers had a mean age of 25.09 years at the time of recording, with a range of 19-37 years of age. Their mean age of arrival in the US was 19.91 years, with a range of 3-31 years of age, and they had begun speaking English at approximately 10.18 years of age on average, with a range of 3-16 years of age.

Intelligibility measures and accentedness ratings were obtained for the Spanish accented speakers and the Mixed-accented speakers. The same procedures used with the Korean-accented speakers for obtaining these measures were also used for the Spanish and Mixed-accented speakers. The native Spanish speakers had a mean intelligibility of 49.8% for words and 82.7% for sentences, and the male speakers had a mean of 53.0%

for words, and the female speakers had a mean of 46.5% for words (see Table 1 for individual speakers). The mean accentedness rating for the Spanish speaker group was 4.23, the mean accentedness for male speakers was 3.77, and the mean for female speakers was 4.69. Two talker groups of six native Spanish speakers (3 males, 3 females) were formed based on intelligibility and accentedness. Group 1 had a mean intelligibility of 50.6% for words and a mean accentedness of 4.121. Group 2 had a mean intelligibility of 48.9% for words and a mean accentedness of 4.333. There were no significant differences between the groups on intelligibility, $t\ (10) = 2.23$, $p = .56$, or on accentedness, $t\ (10) = 2.01$, $p = .75$.

The Mixed-accent speakers had a mean intelligibility of 74.4% for words and 93.5% for sentences, the male speakers had a mean of 75.4% for words, and the female speakers had a mean of 73.4% for words (see Table 1 for individual speakers). The mean accentedness rating for the Mixed-accent group was 3.84, the mean accentedness for male speakers was 4.04, and the mean for female speakers was 3.64. Two talker groups of six Mixed-accent speakers (3 males, 3 females) were formed to include a diverse group of native languages, but the groups were also balanced for intelligibility and accentedness. Group 1 had a mean intelligibility of 72.9% for words and a mean accentedness of 4.18. Group 2 had a mean intelligibility of 76.0% for words and a mean accentedness of 3.49. There were no significant differences between the groups on intelligibility, $t\ (10) = .41$, $p = .69$, or on accentedness, $t\ (10) = .850$, $p = .415$.

Baseline intelligibility and accentedness across accent groups were compared using one-way ANOVAs. There were no significant differences among the groups for accentedness, $F\ (2, 35) = .326$, $p = .724$. There was a significant difference in intelligibility, $F(2, 35) = 8.69$, $p = .001$. Planned comparisons found that Spanish-accented speakers were significantly less intelligible than Korean-accented speakers $t(33) = 2.86$, $p = .007$, and than Mixed accent speakers, $t(33) = 4.06$, $p < .001$. There was no significant difference between Korean-accented speakers and Mixed accent speakers, $t(33) = 1.20$, $p = .238$. These systematic differences in intelligibility across accent groups may affect perceptual learning and the ways in which listeners process and represent accented speech for the different groups.

Procedure

Training phase. Stimulus presentation and data collection were controlled using EPrime software (Schneider, et al., 2002). The training paradigm was the same as that of Experiment 1, except that the first language of the talker group during training varied by condition, and the test contained either Spanish- or Korean-accented speech. Participants were divided into four conditions for training: Spanish training – with Spanish-accented talkers, Korean training – with Korean-accented talkers, Mixed training – with the group of accented speakers from various languages, and No Training controls (See Figure 4). Participants were trained on words spoken by six of the accented speakers and were tested immediately after training. Training was counterbalanced by talker group so that half of the Korean training participants trained with talker group 1, and half trained with talker group 2. The Spanish training groups and Mixed training groups were similarly counterbalanced. Figure 3 illustrates the experimental design of Experiment 2.

Generalization test. At test, all participants transcribed 48 novel words (24 easy, 24 hard) spoken by either the Spanish- or Korean-accented speakers (See Figure 4). The talkers heard at test were different from those heard during training for all groups. The use of unfamiliar talkers for all conditions served to avoid any potential effects of talker-specific learning for the groups who received Spanish- or Korean-accented training. During the test, participants heard each word and were asked to transcribe the word they thought the speaker intended to say. No feedback was given at test.

Results and Discussion

Listeners' transcriptions were scored for accuracy in the same manner as for Experiment 1. Means and standard deviations for training are shown in Table 3, and those for test are shown in Table 4.

Training phase.

Performance across training was analyzed for each training group using repeated-measures ANOVAs with training blocks (blocks 1-3) and lexical properties (*easy* and *hard* words) as the within-subjects factors (Figure 5). For listeners trained with Korean-accented speech, there was a main effect of lexical properties F (1, 39) = 322.39, $p <$ .001, *partial* η^2 = .892, with listeners transcribing *easy* words more accurately than *hard* words. There was also a main effect of training block for Korean-accented training that approached significance, F (2, 78) = 2.86, p = .063, *partial* η^2 = .068. Improvement occurred between blocks 1 and 2, p = .021. There was no interaction of lexical properties and training block, F (2, 78) = .227, p = .80, *partial* η^2 = .006. For listeners trained with Spanish-accented speech, there was a main effect of lexical properties F (1, 39) = 329.58, p < .001, *partial* η^2 = .894, with listeners transcribing *easy* words more accurately than *hard* words. There was no main effect of training block for Spanish-accented training, F (2, 78) = 1.14, p = .325, *partial* η^2 = .028, and no interaction of lexical properties and training block, F (2, 78) = 1.14, p = .327, *partial* η^2 = .028. For listeners trained with the

Mixed-accent group, there was a main effect of lexical properties $F(1, 39) = 331.98$, $p < .001$, *partial* $\eta^2 = .897$, with listeners transcribing *easy* words more accurately than *hard* words. There was no main effect of training block for Mixed-accented training, $F(2, 78) = .28$, $p = .754$, *partial* $\eta^2 = .007$. However, there was a significant interaction of lexical properties and training block, $F(2, 78) = 4.68$, $p = .012$, *partial* $\eta^2 = .110$. Follow-up analyses indicated that there was no interaction for blocks 1 and 2, $F(1, 39) = 2.20$, $p = .145$, *partial* $\eta^2 = .054$, but there was a significant interaction between blocks 2 and 3, $F(1, 39) = 10.07$, $p = .003$, *partial* $\eta^2 = .205$. Paired-samples *t*-tests were used to compare performance at blocks 2 and 3 for *easy* and *hard* words separately. There were significant differences between blocks 2 and 3 for both *easy* words, $t(40) = 2.04$, $p = .025$, and *hard* words, $t(40) = -1.81$, $p = .04$; however, transcription performance decreased between blocks 2 and 3 for *easy* words and increased for *hard* words.

The mixed results for transcription training in Experiment 2 mirrored Experiment 1 in that no clear evidence of learning was found across the training blocks. Since listeners received novel words in each block, the difficulty of the blocks should have been similar. Listener performance may have been similar across training blocks due to the extreme variability of the phonetic contexts and combinations they heard in each new set of items. That there was no clear effect of increased performance in the training blocks may indicate that listeners learned slowly, amassing experience with the accented speech over many trials, and that the learning process involved much trial and error such that learning was only apparent in comparison with untrained listeners.

Generalization test.

Transcription performance at test was analyzed with a mixed factorial ANOVA with Training Condition (Same accent, Different accent, Mixed accent, or No training) and Test accent (Korean- or Spanish-accented speech) as between-subjects variables and lexical properties (*easy* and *hard* words) as within-subjects factors. There was a main effect of lexical properties, F (1, 152) = 391.93, $p < .001$, *partial* $\eta^2 = .721$, with *easy* words transcribed more accurately than *hard* words. There was also a significant main effect of Test accent, F (1, 152) = 407.43, $p < .001$, *partial* $\eta^2 = .728$, with Korean-accented speech transcribed more accurately overall than Spanish-accented speech. There was a significant main effect of Training Condition, F (3, 152) = 4.53, $p = .005$, *partial* $\eta^2 = .082$ (See Figure 6). There were no significant interactions among any of the variables.

The main effect of training condition reflects different transcription performance as a function of training experience. Planned comparisons indicated that listeners trained and tested with the Same accent transcribed test words more accurately than listeners who received No Training, t (156) = 1.78, $p = .039$. Listeners who were trained and tested with the Same accent also performed better than listeners trained and tested with Different accents, with the difference approaching significance, t (156) = 1.78, $p = .066$. No significant differences were found among the other training conditions (Different accent and No training; Same and Mixed accent; or Mixed accent and No training). That listeners who were trained and tested with the same accent performed more accurately than listeners who received no training strongly suggests specificity of learning. However, examining the Test Accent Groups in more detail may reveal complexities in the patterns of listener performance. Although there was no interaction between Training Condition and Test Accent, the difference in overall transcription performance between listeners tested with Korean- and Spanish-accented speech was large. To explore potential differences in patterns in the Spanish- and Korean-accented data, listeners who were tested with Korean-accented speech were analyzed separately from data from listeners tested with Spanish-accented speech. Since there were no interactions with lexical properties, *easy* and *hard* words were combined for all following analyses.

For listeners tested with Korean-accented speech, there was a main effect of condition, $F\ (3, 80) = 2.82$, $p = .044$, *partial* $\eta^2 = .100$ (Figure 7). Planned comparisons revealed that listeners trained with Korean-accented speech transcribed novel Korean-accented words more accurately than listeners trained with either Spanish-accented speech, $t\ (76) = 2.68$, $p = .009$, or Mixed-accented speech, $t\ (76) = 1.97$, $p = .052$, or than listeners who received no training at all, $t\ (76) = 2.26$, $p = .027$. There were no significant differences among the other Training Conditions. That listeners trained and tested with Korean-accented speech performed better on the transcription task than listeners in other conditions suggests that was not learned through training with another particular accent or with the Mixed-accent speakers.

For listeners tested with Spanish-accented speech, there was a main effect of condition, $F\ (3, 80) = 3.19$, $p = .028$, *partial* $\eta^2 = .112$ (Figure 8). Planned comparisons indicated that listeners trained with Spanish-accented speech transcribed novel Spanish-accented words more accurately than untrained listeners, $t\ (76) = 2.70$, $p = .009$. Additionally, listeners trained with the Mixed-accent speech transcribed more accurately than untrained controls, $t\ (76) = 2.63$, $p = .010$. There were no significant differences among the other Training Conditions. While listeners trained with Spanish-accented speech performed better than controls, that was also true of listeners trained with the

Mixed-accent group of talkers.

Recall that for each condition in Experiment 2, two sets of training talkers were used. For both Korean- and Spanish-accented speech, those groups were balanced in terms of type of accent, intelligibility, and accentedness. For the Mixed accent groups, speakers were assigned to groups such that there were a variety of different L1s in each group, and speakers were balanced on intelligibility and accentedness. The accents present in Mixed group 1 were Albanian, Dutch, Japanese, Romanian, Bengali, and Hindi. The accents present in Mixed group 2 were French, German, Somali, Russian, Chinese, and Turkish. Since French and Spanish are both Romance languages, they may share many of the same properties that could be realized in accented English produced by native French and Spanish speakers. As a consequence, the exposure to French-accented English, however minimal, in the one training group may have influenced the apparent effects of generalization seen for listeners tested with Spanish.

To investigate this question, a one-way ANOVA was performed to compare listeners trained with either Mixed group 1, Mixed group 2, or not trained at all on their performance on the Spanish-accented test. A significant main effect of training group was found, $F\ (2, 37) = 4.71$, $p = .015$, *partial* $\eta^2 = .203$. Planned comparisons revealed that listeners trained with Mixed group 2 performed significantly better than untrained controls, $t\ (37) = 3.07$, $p = .002$, as well as better than listeners trained with Mixed group 1, $t\ (37) = 1.87$, $p = .035$. There was no significant difference between listeners trained with Mixed group 1 and untrained controls, $t\ (76) = 1.05$, $p = .151$. A one-way ANOVA was also used to investigate if Mixed training group had any effect on Korean-accented test performance. There were no significant differences between listeners trained with

Mixed group 1, Mixed group 2, or untrained controls when tested with Korean-accented speech, F (2, 37) = .088, p = .916, *partial* η^2 = .005. These findings tentatively suggest that exposure to the accented speech from Mixed group 2 may have facilitated processing of Spanish-accented speech. However, since six different accents were presented during training, it is impossible to pinpoint exactly which accent caused better test performance. However, the inclusion of a French speaker into Mixed group 2, since French is a historically similar language to Spanish, could potentially be the cause of transfer from learning to test in this case.

Intelligibility Analyses

Another reason that patterns of generalization may have differed for the Spanish- and Korean-accented tests is that since the Spanish- and Korean-accented speech differed in terms of baseline intelligibility and accentedness, differences in intelligibility may have contributed to the differences in how learning transferred to the generalization test. In order to examine effects of baseline intelligibility on the performance of listeners at test, performance with talkers of comparable baseline intelligibility across accent groups were compared. Recall that the groups of talkers in the test groups were balanced for baseline intelligibility, meaning that each listener encountered both high and low intelligibility talkers at test. The possible effects of intelligibility can be examined by comparing the performance at test on words produced by the three high intelligibility Spanish speakers (mean intelligibility = .580) and the three low intelligibility Korean speakers (mean intelligibility = .550). Based on these groupings, no significant difference in the baseline intelligibility of the two groups was found, t (10) = .817, p = .433. An ANOVA with training condition (Same-accent, Different-accent, Mixed-accent, and No

Training) and test accent (Korean- and Spanish-accented tests) was performed on these subsets of talkers at test. There was a main effect of test accent, *F* (1, 152) = 13,12, $p <$ .001, *partial* η^2 = .079, indicating that Korean-accented test items were transcribed more accurately that Spanish-accented test items, even though baseline intelligibility was similar. There was also a main effect of training condition, *F* (3, 152) = 3.03, p = .031, *partial* η^2 = .056. Planned comparisons revealed that there were significant differences between the Same accent group and the Different accent group, *t* (156) = 2.31, p = .011, the Same accent group and the Mixed accent group, *t* (156) = 1.67, p = .049, and the Same accent group and the No Training controls, *t* (156) = 2.69, p = .004. There were no significant differences among any other training groups or controls (Different and Mixed; or Mixed and No Training). There was no interaction between Test accent and Training condition, *F* (3, 152) = .781, p = .506, *partial* η^2 = .015.

To examine the effects of learning for the high intelligibility Korean-accented talkers and low intelligibility-Spanish accented talkers, two one-way ANOVAs were performed with Training Condition as the between-subjects variable. For the high intelligibility Korean-accented test, there was no significant main effect of training condition, *F* (3, 79) = 1.66, p = .182, *partial* η^2 = .056, but there was a trend toward specificity of learning. Planned comparisons showed that the group trained and test with Korean-accented speech transcribed the test items better than both the Spanish-accented trained listeners (p = .026) and the No Training controls (p = .082). For the low intelligibility Spanish-accented test, there was also no significant main effect, *F* (3, 79) = 1.11, p = .351, *partial* η^2 = .056, showing no benefit for training of any type.

When only the talkers in the middle to high range of intelligibility for both accent groups are considered, only training with the same accent as test facilitates transcription. Previous research using the same Spanish-accented stimuli in Experiment 2 has found evidence that the high intelligibility Spanish-accented talkers appeared to produce more systematic regularities in their accented speech than low intelligibility speakers, at least for vowel production (Nygaard, et al., 2006). Similarly, research on Spanish- and Korean-accented speech in general has shown that late bilinguals, who tend to be low intelligibility talkers, produce more variable L2 utterances (Lee, et al., 2006; Nygaard, et al., 2006; Schmidt & Flege, 1996). If there are fewer systematic regularities in low intelligibility accented speech, then listeners may have a more difficult time parsing the relevant levels of variation and learning the relevant reliable cues.

General Discussion

The experiments presented here examined how listeners perceptually adapt to accented speech. The purpose of these studies was to investigate to what degree the perceptual adaptation to accented speech was accent-specific and what that implies for underlying changes in processes of representation and categorization of spoken language. The current experiments built on evidence for perceptual learning for Spanish-accented speech in Sidaras et al. (2009). Experiment 1 replicated and extended previous findings by showing that listeners can perceptually adapt to Korean-accented speech in a similar training paradigm. Although there was little evidence of learning within the training blocks, trained listeners performed better on a transcription task at test than did untrained controls, indicating that learning did occur. Trained listeners generalized both to novel words and novel speakers of Korean-accented speech, indicating that learning was neither item- nor talker-specific, which provides evidence that listeners learned the systematic regularities that facilitated processing of Korean-accented speech. The results of Experiment 1 replicate those of Sidaras et al., providing evidence that listeners can perceptually adapt to Korean-accented speech in much the same way as to Spanish-accented speech.

Experiment 2 extended the perceptual learning paradigm to investigate specificity and generalization in the perceptual learning process by examining to what degree learning transfers between L1 accent groups. Listeners were trained with Spanish-, Korean-, or Mixed-accent talkers and were tested with novel words and unfamiliar talkers with either Spanish- or Korean-accents. As in Experiment 1, there was little evidence for learning within the training blocks, and listeners only showed evidence of learning when compared with untrained controls. The learning paradigm focuses on giving listeners extremely variable input, with six talkers producing items that vary greatly in phonological content and lexical properties. That listeners showed little improvement during training reflects the difficultly of processing highly variable stimuli from multiple talkers in the brief high variability training paradigm used in the current study. Regardless of performance during training, listeners trained and tested with the same accent performed better on test transcription than untrained controls. Listeners trained and tested with the same accent also performed better than listeners trained and tested with different accents. Listeners who were trained with talkers from many different L1 accent groups did not perform better than untrained controls, indicating that training with mixed accent talkers did not facilitate processing of a previously untrained accent. These findings provide evidence for specificity of learning, such that listeners showed a benefit of training when they were trained and tested on the same accent group. There was little evidence of overall generalization, since listeners showed learning for neither training with a different accent nor training with many different accents.

The current study showed no evidence of talker-specific learning for the accented talkers and little evidence that lexical properties affect the perceptual learning process for accented speech. Experiment 1 compared the transcription performance of listeners trained and tested with familiar or unfamiliar talkers and found that although trained listeners more accurately transcribed both familiar and unfamiliar accented speech than did untrained controls, there was no difference between trained listeners' performance for familiar or unfamiliar accented talkers. Listeners may be attending more to the properties that are diagnostic for the group of accented speakers as a whole rather than for the idiolect produced by any one talker. Listeners also gained relatively more exposure for the accent as a whole than for any one talker, further biasing them to learning properties of the accent rather than the talker-specific properties. There is also little evidence that lexical properties of the items included in the training and test phases affected the learning process. There was a consistent and clear main effect of *easy* and *hard* words at training and test for both experiments, with *easy* words being transcribed more accurately than *hard* words. This main effect was expected due to the low frequency and high neighborhood density of the *hard* words, which are easily confused with high frequency neighbors. However, there was no clear evidence that lexical properties affected other variables at training or test. If there had been effects of lexical properties on training conditions, it would indicate that listeners were forming representations that included lexical properties rather than only acoustic-phonetic information. It may be that with increased experience with accented speech, listeners would begin to show effects of representing lexical properties as well.

That learning facilitated processing of novel words and voices only for talkers with the same accent as presented during training is consistent with the evidence for specific learning found by Maye et al. (2008). They found that that learning extended to items that varied in ways that were predictable according to the properties of the accent, but not to items that varied in ways that were not consistent with the speech presented during training. In the current study, listeners showed the greatest transfer of learning when the accent was the same at training and test and little evidence for generalization from training with Korean- to Spanish-accented speech or with Spanish- to Korean-accented speech. These findings parallel those of Liss et al. (2002) showing that listeners can generalize to novel dysarthric speakers, but only when the speakers share the same deficit type. The findings of Maye et al. (2008) and Liss et al. (2002) along with those of the current study indicate that perceptual learning of spoken language characteristics generalizes only to other instances of spoken language that share properties with the speech that was learned.

Both Spanish- and Korean-accented speech differ in specific ways from one another and from native English speech. These specific differences in accent properties produce unique challenges for listeners as they encounter either Korean- or Spanish-accented speech, resulting in different patterns of perceptual confusions for the two accents. For example, Sidaras et al. (2009) found that untrained listeners were most likely to confuse Spanish-accented vowel /ʌ/ with /æ/, /a/, /o/, and /u/, and to confuse vowel /i/ with /I/. However, Alexander et al. (2008) found that untrained listeners were most likely to confuse Korean-accented vowel /I/ with /i/ and /e/ and /ʌ/ with only /a/. The differences in vowel confusion patterns for untrained listeners reflect the ways in which native Spanish and Korean speakers produce accented English. Neither the native Korean nor Spanish vowel inventories contain the vowels /I/ and /æ/, explaining why those vowels would initially be highly confusable with similar vowels. Accented speakers would have to approximate the English vowel using temporal and spectral cues without relying on a similar vowel from their L1, which results in the use of systematic variation in different dimensions in Spanish- and Korean-accented speech. For example, Alexander et al. (2008) found that Korean-accented speakers made distinctions between /I/ and /i/ in spectral qualities and distinctions between /æ/ and /a/ in temporal qualities. However, Sidaras et al. (2009) found that Spanish-accented speakers made distinctions between /I/ and /i/ mainly in temporal qualities and distinctions among /ʌ/, /æ/, and /a/ in spectral qualities. Similar patterns of differences in how the speakers realize Spanish- versus Korean-accented English should be present for other vowels as well as consonant contrasts. To adapt to the set of properties for any accent would involve learning which cues are diagnostic for the speech sounds in that accent. Only accents that share many of the same acoustic-phonetic properties would show cross-accent transfer.

One way that specificity of learning for accented speech might emerge at the acoustic-phonetic level is that listeners may differentially attend to or weight cues that are diagnostic for the particular accent and ignore cues that are not reliable for retrieving linguistic structure. For example, in line with the findings of Sidaras et al. (2009) and Alexander et al. (2008) mentioned previously, listeners may learn to attend to spectral distinctions for some contrasts and temporal distinctions for others. Attentional weighting could be one important and flexible mechanism that allows listeners to shift their focus from one type of cue to phoneme identity to a novel cue as a function of learning. In the current study, listeners who received accented training learned to attend to systematic cues such that they more accurately transcribed words than untrained controls. This learning and its specificity indicates both that accented speech must contain systematic regularities that are consistent across talkers and that listeners are sensitive to those regularities and can readily shift strategies of speech perception to encompass new information.

Additionally, listeners must map the accented utterances onto existing acoustic-phonetic categories in order to include those items as members of particular categories or accented subcategories (Dupoux & Green, 1997; Francis, et al., 2000; Greenspan, et al., 1988; Liss, et al., 2002). The process of mapping the items may allow for greater attention to shared properties within categories and to contrastive properties between categories, leading to the increased perceived similarity for items within a category and increased perceived difference for items across categories found in previous research (Goldstone, 1994, 1995). Mapping from novel utterances to existing categories can occur through the use of lexical context during unsupervised learning, as in the rapid adaptation paradigm used by Clarke and Garrett (2004), or mapping may occur in supervised learning contexts in which the category mapping is provided through feedback or presenting the written items during learning (Liss, et al., 2002).

Once novel utterances are mapped to existing categories, listeners may engage in attentional weighting for those cues that are most relevant for retrieving linguistic structure for multiple accented talkers. However, they may also engage in processes of stimulus imprinting at the acoustic-phonetic level, retaining in memory many different cues that may or may not be diagnostic for other talkers with the same accent or even other tokens from the same talker. Listeners might employ mechanisms of stimulus imprinting, forming representations of the utterance that capture both systematic and nonsystematic variability, when the learning context is extremely variable and has little systematicity. Since there would be fewer relevant cues to attentionally weight, listeners might form detailed representations of properties that vary nonsystematically. This type of learning might result in a general broadening of acoustic-phonetic categories as listeners gained experience with many, varied utterances that shared few common properties. As category boundaries stretch to incorporate the highly variable information present in the utterances, listeners may become more likely to accept more variable items into broadened categories, allowing listeners to perhaps generalize learning to previously untrained accented speech.

Although there was strong evidence in the current study for specificity of learning, there was also some evidence for generalization of learning for those listeners trained with the group of speakers from many different L1 accent groups and tested with Spanish-accented speech. Listeners trained with the mixed accent group of speakers showed better performance on the transcription task than untrained controls, indicating that perceptual learning transferred to the Spanish-accented words heard during test. On the surface, this seems to indicate that listeners gained a benefit for Spanish from learning many other accents. However, there are a number of factors that may have contributed to the presence of generalization effects for the Spanish-accented test. The first, and most likely, explanation is that listeners perceptually adapted to the speech produced by the French-accented talker in the one training group that showed the generalization effect.

Although generalization from a single talker to other accented talkers seems more difficult than having multiple talkers with similar accents to use in learning for generalization, Weil (2001) found that training with a single accented talker facilitated language processing for a similarly accented talker. It is possible that listeners learned information about French-accented speech that provided a benefit for the processing of Spanish-accented speech. It is less clear whether listeners employed mechanisms of attentional weighting, learning idiosyncratic characteristics of the talkers presented in the Mixed accent group, or processes of stimulus imprinting, learning many different cues, including cues from the French-accented talker that generalized to Spanish-accented speech. That experience with the second Mixed accent group, containing no native speakers of Romance languages did not produce benefits for listeners tested with Spanish-accented speech and that no effects of generalization to Korean-accented speech were found provides evidence that generalization does not occur without the presence of acoustic-phonetic similarity between accents. Future research should examine the extent to which accents overlap in their acoustic-phonetic properties and to what degree similar accents cross-facilitate one another in perceptual learning. Additionally, it may be possible to disentangle mechanisms of attentional weighting and stimulus imprinting through training with highly constrained stimuli, giving listeners some systematic variability that is diagnostic for the recovery of linguistic structure and other, less systematic variability that would generalize to a set of test stimuli. If listeners show a benefit at test, it might point toward the use of stimulus imprinting as well as attentional mechanisms.

Another explanation for the apparent evidence for generalization for the Spanish-accented talkers is that the Spanish- and Korean-accented speakers who provided the stimuli for these experiments differed greatly in baseline intelligibility. The Korean-accented speakers were overall more intelligible than the Spanish-accented speakers. Not only did that result in Korean-accented speech being transcribed more accurately overall, but it may have interacted with the training and test processes. When only speakers from both accent groups with comparable baseline intelligibility were considered, the effect of generalization disappeared. Previous research has shown that accented speech produced by low proficiency and low intelligibility speakers is more variable (Lee, et al., 2006; Nygaard, et al., 2006; Schmidt & Flege, 1996). Low intelligibility talkers have greater non-systematic variation than high intelligibility talkers, which might result in either greater impetus for stimulus imprinting or for greater probability of shared acoustic-phonetic characteristics with other highly variable talkers. If listeners were trained with extremely variable speech, such as speech from different accents, then it might be likely that there would be more overlap with highly variable, low intelligibility talkers (such as those in the Spanish-accented test group) than more systematic high intelligibility talkers (such as those in the Korean-accented test group). The intelligibility differences of the Korean- and Spanish-accented talkers may have played a role in how the learning context transferred to the items presented at test.

If specificity of learning accounts for the results of the present study, then what implications does that have for issues of perceptual learning and category or representational change in spoken language? For listeners to learn something specific about a particular accent, they must be attending to properties of the accent as well as retaining those accent properties in memory. If listeners attend to the diagnostic cues of an accent while receiving feedback on which acoustic-phonetic category that item should map onto, then listeners would be amassing information that tells them what properties the accented speech has in common as well as to which category each sound belongs. With this information, listeners can begin to build accent-specific filters, or subcategories that allow for facilitation of spoken language belonging to that accented in the future. This kind of learning mechanism is consistent with recent work investigating statistical learning of spoken language, in which listeners learn probabilistic regularities within language input (Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996; Tyler & Cutler, 2009). The probabilistic information allows listeners to learn which properties are most likely to co-occur or relate to one another, aiding in acquisition of word boundaries (Saffran, et al., 1996), as well as potentially many other properties of language.

There are a number of ways in which categories might change to accommodate the new influx of information about accented speech. Listeners may form new accented categories or subcategories that, through the use of lexical context and explicit feedback during learning, map onto existing categories (Dupoux & Green, 1997; Francis, et al., 2000; Greenspan, et al., 1988; Liss, et al., 2002). Before training, the accented items would be more confusable than native speech, but through experience, listeners would develop an accent-specific filter that would identify accented speech sounds as accented members of preexisting native speech categories (Liss, et al., 2002). Another hypothesis is that listeners amass exemplars of accented speech sounds that are similar, in some way, to current acoustic-phonetic categories. As listeners accumulate experience with accented items, the boundaries of the categories shift to include the new items as category members (See Figure 9). These kinds of shifts can be localized to particular phonetic contexts, as with the stimuli used by Dahan et al. (2008), in which vowels were only raised or lowered when used with particular consonants, or they can be chain shifts, where many categories shift in a systematic way, as in the "accent" constructed by Maye et al. (2008) in which all front vowels were lowered. These systematic shifts in categories are referred to as chain shifts and have been used to explain historical changes in spoken English, such as the Great Vowel Shift and the Northern Cities Shift (which is currently underway) (Ettlinger, 2007; Yule, 2006).

If listeners change their category boundaries through perceptual learning of accented speech, it is possible that their experience with accented speech sounds may alter their language processing more generally. The processes and representations underlying spoken language are extremely flexible and seem to adapt to even small amounts of novel input, such as the adaptation found by Clarke and Garrett (2004) in which listeners became faster at processing accented speech after exposure to only 4 accented sentences. Ettlinger (2007) modeled how experience with one set of altered vowel utterances can produce a set of chain shifts that alter all other vowel categories. Ettlinger's model can be used to explain how vowel categories shift across generations of speakers, as in the Great Vowel Shift, or it may also be able to explain how listeners generalized the properties of accented speech presented by Maye et al. (2008) to set of vowels they had yet to encounter in the constructed "accent." Maye et al. hypothesize that in such a flexible and ever-changing system, there must be a mechanism in place to maintain native acoustic-phonetic speech sounds as well as multiple accents or dialects. The mechanism behind the maintenance of multiple unique sets of acoustic-phonetic characteristics may be that accented speech is represented in subcategories, specific to the particular accent, or it could be simply that the bulk of a listener's experience comes from other native speakers of the language. However, experience with accented speech might actually perturb normal speech processing, changing not just the category boundaries, but also the category midpoints (Floccia, et al., 2006), ultimately changing how listeners process spoken language more generally (Clarke-Davidson, et al., 2008; Kraljic & Samuel, 2005, 2006). However, such an overarching change is unlikely without extreme amounts of experience with a particular accent or dialect, since the total amount of experience with native speech is likely to outweigh any brief exposure to accented speech such as the adaptation found with exposure to only four accented sentences shown by Clarke and Garrett (2004). If extended (or even brief) exposure to accented speech does have an effect on speech perception in general, it would indicate that listeners are not forming separate-but-related accented categories or subcategories but were including the accented tokens into their overall acoustic-phonetic categories for spoken language.

Another question raised by the current study is the extent to which accent similarity can drive perceptual adaptation. Do languages that are similar in acoustic-phonetic characteristics have similar L2 accented realizations? Would perceptual learning of one accent facilitate another similar accent? If the group of listeners trained with Mixed group 2 and tested with Spanish-accented speech showed better transcription performance due to experience with the historically similar French L1 of the accented English produced by the talker, then it would seem that many L1 realizations of accented speech could facilitate accented speech with shared properties. Future research should examine what degree of etymological and acoustic-phonetic similarity is necessary for one accent to facilitate linguistic processing for another accent. For example, would similar languages such as Dutch and English produce more similar accents in an L2 than less similar languages such as Russian and English? If accented speech perception can be facilitated by similar languages, it may be the case that individuals who encounter a variety of different accents might have an increased ability to understand many accented speakers, which might explain why Bent and Bradlow (2003) found an interlanguage intelligibility effect using Asian and European accents, and Stibbard and Lee (2006) found no benefit for Korean and Saudi Arabian accents.

It should be noted that even brief training with accented speech produces perceptual adaptation to novel speakers of that same accent. Although experience with one accent may not facilitate processing of other accents, experience may facilitate processing of similar accents. These findings have implications for theories of spoken language processing. Specifically, the current study provides evidence that listeners attend to surface characteristics of accented speech and retain them in memory, which supports previous research for spoken language more generally (Bradlow, et al., 1999; McLennan & Luce, 2005; Nygaard, et al., 2000; Palmeri, et al., 1993). The perceptual learning of accented speech facilitates spoken language processing (Bradlow & Bent, 2008; Gass & Varonis, 1984; Sidaras, et al., 2009) and provides evidence that listeners can learn information not only about particular talkers but also about groups of talkers who share systematic differences in speech properties. That listeners learn information that spans groups of talkers indicates that they can correctly apportion variability to the correct sources, resulting in greater speech perception performance for novel talkers who share the same properties (Sidaras, et al., 2009). Additionally, listeners perceptually learn the systematic regularities specific to particular accents, which has implications for the kinds of changes taking place in listeners' categorization processes and underlying representations during perceptual learning. The specificity of perceptual learning of accented speech constrains the type of mechanisms that may be employed during perceptual learning of accented speech and spoken language more generally.

One of the remarkable features of the speech perception system is that it can both readily adapt to new types of speech not previously encountered, altering representational structures and categorization processes, and it can maintain stable linguistic perception, providing perceptual constancy in the face of a great deal of variation. The experiments presented here show how listeners adapt to accented speech by unfamiliar talkers in a very brief training paradigm, showing accent-specific learning that is congruent with accounts of speech perception which posit perceptually-detailed representations for spoken language. The extent to which language users engage in perceptual adaptation and learning suggests that the perceptual system may be able to produce the appearance of stable linguistic perception specifically because it can adapt so readily to the current speech input.

References

Alexander, J. D., Nygaard, L. C., & Sidaras, S. K. (2008). The contribution of vowel production to the intelligibility and accentedness of nonnative speech. Paper presented at the 155th annual meeting of the Acoustical Society of America. Paris, France.

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 115* (6), 3171-3183.

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

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

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61* (2), 206-219.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech .1. Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20* (3-4), 255-272.

Burda, A. N., & Hageman, C. F. (2005). Perception of accented speech by residents in assisted-living facilities. *Journal of Medical Speech-Language Pathology, 13* (1), 7-14.

Burda, A. N., Scherz, J. A., Hageman, C. F., & Edwards, H. T. (2003). Age and understanding speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills, 97* (1), 11-20.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 521-533.

Clarke, C. M. (2000). Perceptual adjustment to foreign-accented English. *Journal of the Acoustical Society of America, 107*, 2856(A).

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116* (6), 3647-3658.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics, 70* (4), 604-618.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108* (3), 710-718.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., 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), 222-241.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning, 54* (4), 655-679.

Duke, J. E., & Sidaras, S. K. (2006). The effects of training context on perceptual learning of accented speech. Paper presented at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan. Honolulu, HI.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology-Human Perception and Performance, 23* (3), 914-927.

Ettlinger, M. (2007). Shifting Categories: An Exemplar-based Computational Model. Paper presented at the 29th Annual Conference of the Cognitive Science Society.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425* (6958), 614-616.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica, 43* (4), 155-171.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *Journal of the Acoustical Society of America, 92* (1), 128-143.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40* (4), 467-491.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41* (1), 78-104.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32* (5), 1276-1293.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62* (8), 1668-1680.

Francis, A. L., Nusbaum, H. C., & Fenn, K. (2007). Effects of training on the acoustic-phonetic representation of synthetic speech. *Journal of Speech Language and Hearing Research, 50* (6), 1445-1465.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*, 412-429.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34* (1), 65-89.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 22* (5), 1166-1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105* (2), 251-279.

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), 152-162.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology-General, 123* (2), 178-200.

Goldstone, R. L. (1995). Effects of categorization on color-perception. *Psychological Science, 6* (5), 298-304.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality - female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50* (6), 524-536.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual-learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14* (3), 421-433.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21* (2), 205-228.

Hall, G. (1991). *Perceptual and Associative Learning*. Oxford: Clarendon.

Halle, M. (1985). Speculation about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic Linguistics* (pp. 101-114). New York: Academic Press.

Helson, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review, 55*, 297-313.

Hervais-Adelman, A., 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), 460-474.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English |r|-|l| to Japanese adults. *Journal of the Acoustical Society of America, 118* (5), 3267-3278.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. J. J. W. Mullennix (Ed.), *Talker Variability in Speech Processing* (pp. 145-165). New York: Academic Press.

Joos, M. A. (1948). Acoustic phonetics. *Language, 24*, Supplement 2, 1-136.

Jusczyk, P. W. (1993). From general to language-specific capacities - the WRAPSA model of how speech-perception develops. *Journal of Phonetics, 21* (1-2), 3-28.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51* (2), 141-178.

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), 1-15.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21* (3), 451-468.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55* (4), 306-353.

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition, 28* (3), 487-513.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 112* (6), 3022-3030.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological*

Review, 95 (4), 492-527.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /1/ - A 1st report. *Journal of the Acoustical Society of America, 89* (2), 874-886.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39*, 155–158.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 1-24.

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), 543-562.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82* (3), B101-B111.

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, & Cognition, 31*, 306-321.

Monsen, R. B., & Engebretson, A. M. (1977). Study of variations in male and female glottal wave. *Journal of the Acoustical Society of America, 62* (4), 981-993.

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

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85* (1), 365-378.

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

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 48* (2), 159-182.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285-310.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech - A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93* (2), 1097-1108.

Nash, R. (1977). *Comparing English and Spanish: Patterns in phonology and orthography*. New York: Regents Publishing Co.

Nissen, S. L., Dromey, C., & Wheeler, C. (2004, Nov 18-20). First and second language tongue movements in Spanish and Korean bilingual speakers. Paper presented at the Annual Convention of the American-Speech-Language-Hearing-Association. Philadelphia, PA.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47* (2), 204-238.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception: Progress Report, 10*, 357-372.

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26* (5), 1228-1244.

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

Nygaard, L. C., Sidaras, S. K., Duke, J. E., & Rasmussen, S. T. (2006). Acoustic correlates of accentedness and intelligibility of Spanish-accented English vowels. Paper presented at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan. Honolulu, HI.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech-perception as a talker-contingent process. *Psychological Science, 5* (1), 42-46.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences, 13* (1), 14-19.

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

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19* (2), 309-328.

Pisoni, D. B. (1993). Long-term-memory in speech-perception - Some new findings on talker variability, speaking rate and perceptual-learning. *Speech Communication, 13* (1-2), 109-125.

Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In K. J. a. J. W. Mullennix (Ed.), *Talker Variability in Speech Processing* (pp. 9-32). San

Diego: Academic Press.

Rice, K. (Ed.). (2002). *Vowel Place Contrasts*. Westport, CT: Praeger.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274* (5294), 1926-1928.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes - Codification and repetition effects in word identification. *Journal of Experimental Psychology-General, 114* (1), 50-77.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15* (2), 123-148.

Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech Language and Hearing Research, 42* (1), 56-64.

Schmidt, A. M., & Flege, J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish English bilinguals. *Phonetica, 53* (3), 162-179.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395-408.

Sebastian-Galles, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics, 62* (4), 834-842.

Sidaras, S. K., Alexander, J. D., & Nygaard, L. C. (2009). Perceptual Learning of

Systematic Variation in Spanish-accented Speech. *Journal of the Acoustical Society of America, 125* (5), 3306–3316.

Spitzer, S. M., Liss, J. M., Caviness, J. N., & Adler, C. (2000). An exploration of familiarization effects in the perception of hypokinetic and ataxic dysarthric speech. *Journal of Medical Speech-Language Pathology, 8* (4), 285-293.

Stibbard, R. M., & Lee, J. I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *Journal of the Acoustical Society of America, 120* (1), 433-442.

Summerfield, Q., & Haggard, M. P. (1973). Vocal tract normalization as demonstrated by reaction times. *Report of Speech Research in Progress, 2* (2), 12-23. Queens University of Belfast.

Tamminen, J., & Gaskell, M. G. (2006, Jul 26-29). Newly learned spoken words show long-term lexical competition effects. Paper presented at the 28th Annual Conference of the Cognitive-Science-Society. Vancouver, Canada.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2* (3), 339-363.

Tsukada, K., Birdsong, D., Mack, M., Sung, H. Y., Bialystok, E., & Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica, 61* (2-3), 67-83.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126* (1), 367-376.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition - Patterns and parameters 1. Recognition of backward voices. *Journal of Phonetics,* *13* (1), 19-38.

Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition - Patterns and parameters 2. Recognition of rate-altered voices. *Journal of Phonetics, 13* (1), 39-52.

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35* (1-2), 103-113.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9*, 325-329.

Weil, S. A. (2001). Foreign accented speech: Encoding and generalization. *Journal of the Acoustical Society of America, 109*, 2473(A).

Yang, B. G. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24* (2), 245-261.

Yule, G. (2006). *The Study of Language* (3rd ed.). Cambridge: Cambridge University Press.

Appendix

Easy Words		*Hard Words*	
was	reach	ban	wad
down	mouth	bead	bud
work	teeth	bean	dame
long	gas	bug	lace
both	jack	bum	lame
thought	check	chat	pad
does	king	cheer	chore
put	shape	comb	cod
give	learn	caught	hack
young	ship	den	kin
thing	neck	dune	kit
peace	watch	fade	pawn
god	judge	fin	bun
five	dog	goat	gut
gave	vote	knob	lice
death	league	lad	mid
shall	thick	mall	wick
real	page	mat	hurl
south	hung	mitt	mote
job	join	mole	teat
love	shop	pat	hash
full	roof	pet	hid
wife	leg	pup	hoot
voice	lose	rat	mace
girl	theme	rhyme	main
live	soil	rum	moan
move	pull	sane	mum
food	chain	soak	rim
size	curve	suck	route
cause	path	ten	wail
chief	dirt	wed	hum
faith	vice	white	sill
pool	rough	con	beak
deep	fool	doom	hag
firm	noise	rut	wave
serve	wash	toot	weed

Table 1.

Accentedness and Intelligibility for Korean-, Spanish-, and Mixed-accented Talkers

Speaker Group	Language	Gender	Mean Accentedness Ratings	Mean Intelligibility (Sentences)	Mean Intelligibility (Words)
Korean Group 1	Korean	Female	4.34	88.4%	53.3%
		Female	5.57	84.8%	58.3%
		Female	3.31	93.6%	64.3%
		Male	5.12	84.8%	59.2%
		Male	3.17	93.2%	67.5%
		Male	3.14	94.9%	81.0%
Korean Group 2	Korean	Female	4.39	93.5%	68.5%
		Female	3.28	91.3%	60.1%
		Female	5.35	83.0%	56.8%
		Male	2.39	98.3%	88.8%
		Male	5.06	81.9%	42.8%
		Male	3.45	95.5%	77.0%
Spanish Group 1	Spanish	Female	5.59	75.6%	32.9%
		Female	4.43	83.0%	39.7%
		Female	3.10	89.8%	68.8%
		Male	4.77	65.9%	54.5%
		Male	2.83	90.5%	58.3%
		Male	4.01	82.9%	49.2%
Spanish Group 2	Spanish	Female	4.31	85.5%	48.9%
		Female	6.17	74.6%	35.2%
		Female	4.54	82.6%	53.5%
		Male	3.55	81.8%	42.9%
		Male	2.68	90.7%	60.3%
		Male	4.75	89.0%	52.8%
Mixed Group 1	Albanian	Female	4.46	93.4%	57.7%
	Dutch	Female	4.59	93.4%	64.7%
	Japanese	Female	4.97	95.0%	79.5%
	Romanian	Male	4.35	97.5%	76.6%
	Bengali	Male	1.73	94.8%	88.9%
	Hindi	Male	5.00	93.8%	70.2%
Mixed Group 2	French	Female	4.34	92.3%	73.7%
	German	Female	2.24	96.6%	72.5%
	Somali	Female	1.23	98.4%	92.7%
	Russian	Male	3.17	98.4%	90.9%
	Mandarin	Male	5.39	74.2%	53.7%
	Turkish	Male	4.59	93.8%	72.2%

Table 2.

Experiment 1: Means and Standard Deviations for Training and Test

	Easy Words		Hard Words		All Words	
Training Block	*M*	*SD*	*M*	*SD*	*M*	*SD*
Block 1	0.809	0.069	0.635	0.080	0.722	0.052
Block 2	0.814	0.058	0.639	0.080	0.727	0.052
Block 3	0.809	0.066	0.646	0.089	0.727	0.059
	Easy Words		**Hard Words**		**All Words**	
Test Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Same Talkers	0.821	0.092	0.634	0.102	0.727	0.074
Different Talkers	0.851	0.088	0.609	0.086	0.730	0.050
No Training	0.799	0.090	0.563	0.094	0.681	0.066

Table 3.

Experiment 2: Means and Standard Deviations for Training

		Easy Words		Hard Words		All Words	
Training Accent	Training Block	*M*	*SD*	*M*	*SD*	*M*	*SD*
Korean	Block 1	0.798	0.059	0.642	0.082	0.720	0.058
	Block 2	0.828	0.075	0.667	0.077	0.747	0.059
	Block 3	0.805	0.071	0.635	0.082	0.720	0.058
Spanish	Block 1	0.772	0.102	0.579	0.112	0.676	0.090
	Block 2	0.741	0.091	0.564	0.124	0.652	0.088
	Block 3	0.751	0.090	0.532	0.058	0.642	0.100
Mixed	Block 1	0.847	0.066	0.701	0.077	0.774	0.058
	Block 2	0.871	0.081	0.694	0.068	0.782	0.056
	Block 3	0.831	0.066	0.724	0.074	0.778	0.055

Table 4.

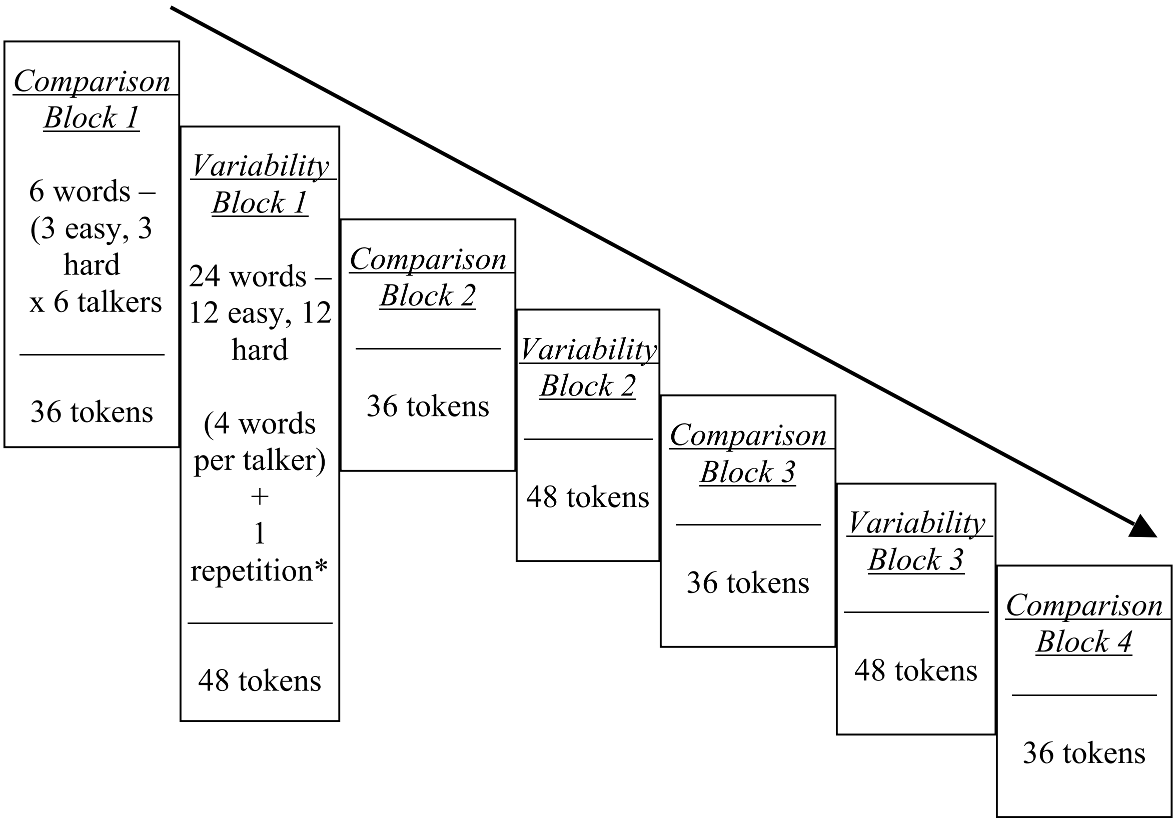
Experiment 2: Means and Standard Deviations for Test

		Easy Words		Hard Words		All Words	
Test Accent	Training Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Korean	Same Accent	0.800	0.055	0.621	0.109	0.710	0.067
	Different Accent	0.752	0.075	0.550	0.129	0.651	0.083
	Mixed Training	0.758	0.066	0.575	0.105	0.667	0.079
	No Training	0.781	0.066	0.540	0.102	0.660	0.045
Spanish	Same Accent	0.602	0.063	0.390	0.105	0.496	0.037
	Different Accent	0.577	0.088	0.379	0.099	0.478	0.045
	Mixed Training	0.613	0.087	0.377	0.107	0.495	0.057
	No Training	0.592	0.063	0.319	0.103	0.455	0.049
Spanish and Korean (all data)	Same Accent	0.701	0.116	0.505	0.158	0.603	0.121
	Different Accent	0.665	0.120	0.465	0.143	0.565	0.110
	Mixed Training	0.685	0.106	0.476	0.145	0.581	0.110
	No Training	0.687	0.115	0.429	0.151	0.558	0.114

Figure 1.

Experiment 1 Design. Listeners heard Comparison and Variability Blocks in alternation followed by the generalization test.

Training phase



* repetition used different talker/voice pairings to provide variability in experience.

Generalization test – either Same or Different Talkers

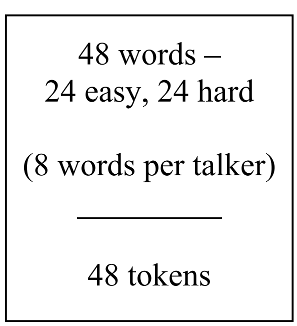


Figure 2.

Experiment 1: Transcription performance across training blocks for easy and hard words for listeners trained and tested with Korean-accented speech.

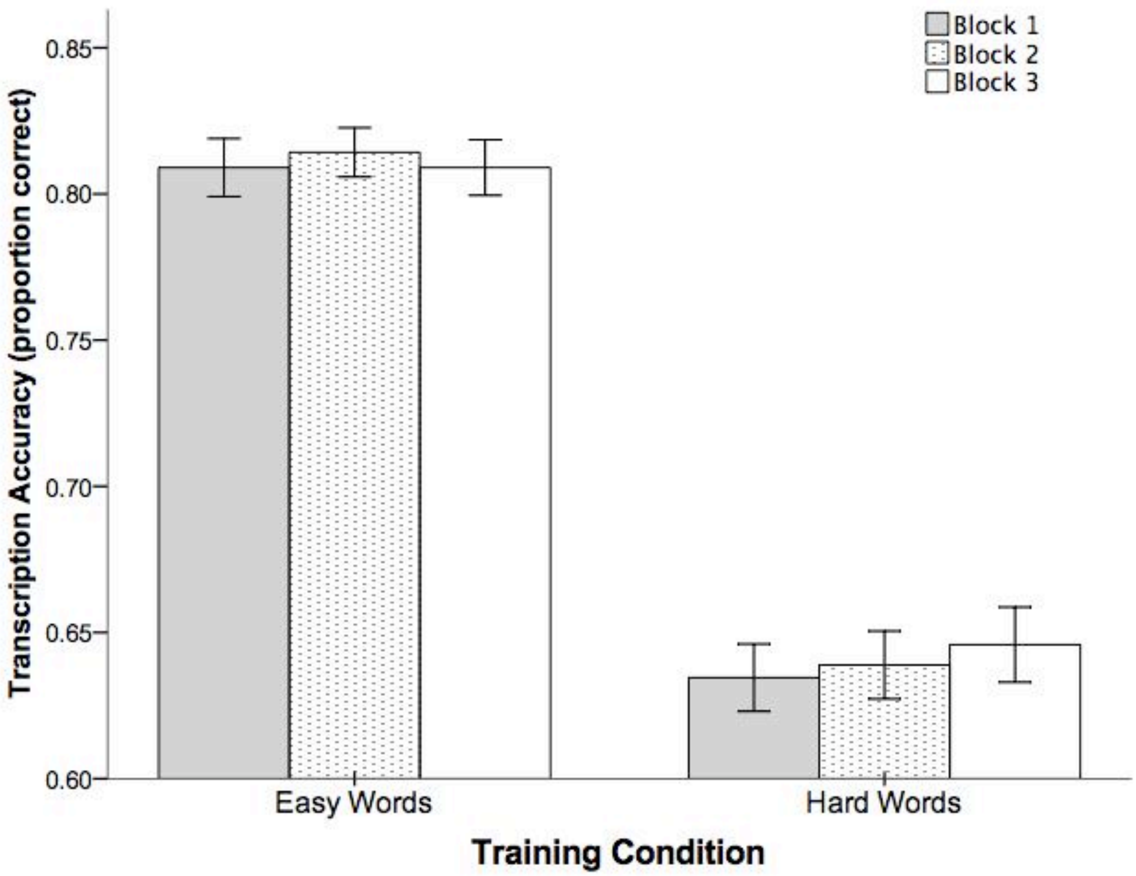


Figure 3.

Experiment 1: Transcription performance at test for listeners trained and tested with Korean-accented speech.

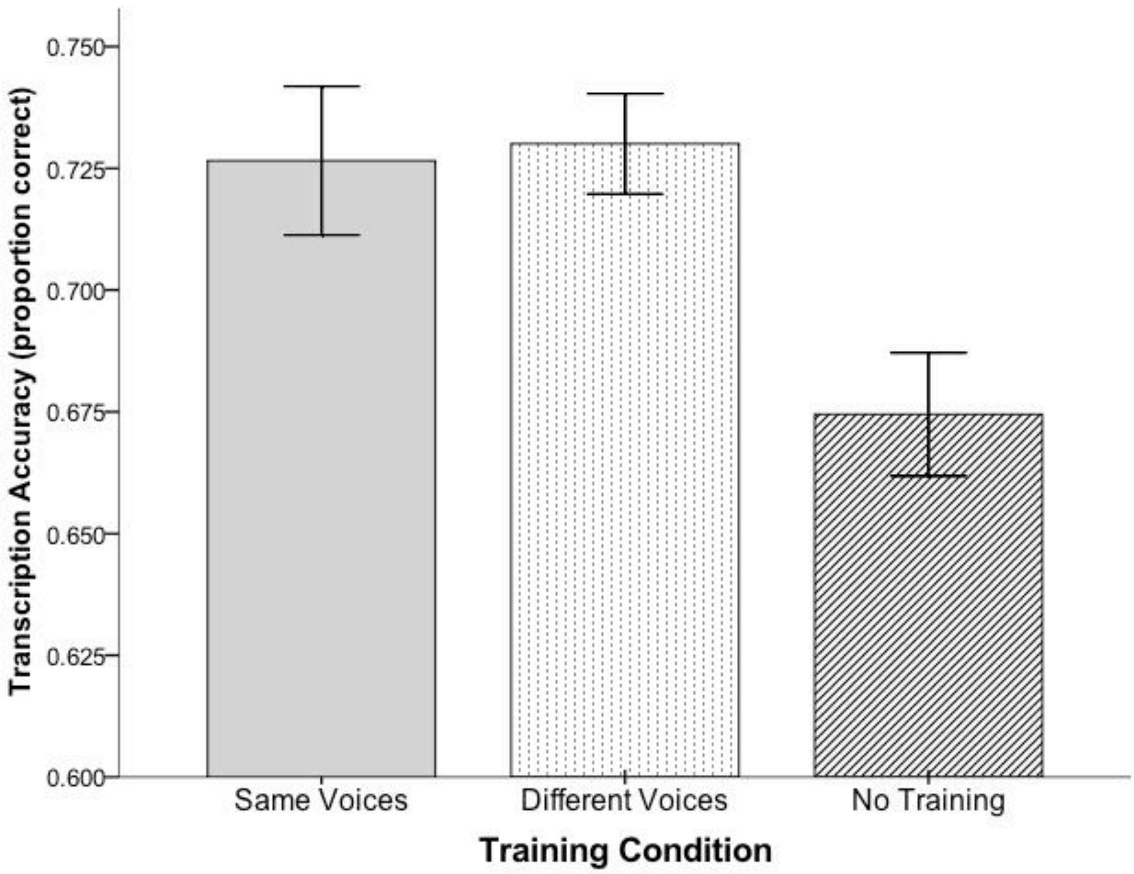


Figure 4.

Experiment 2 Design. Listeners received training with Korean-, Spanish-, or Mixed accented speech and we tested with either Korean- or Spanish-accented speech.

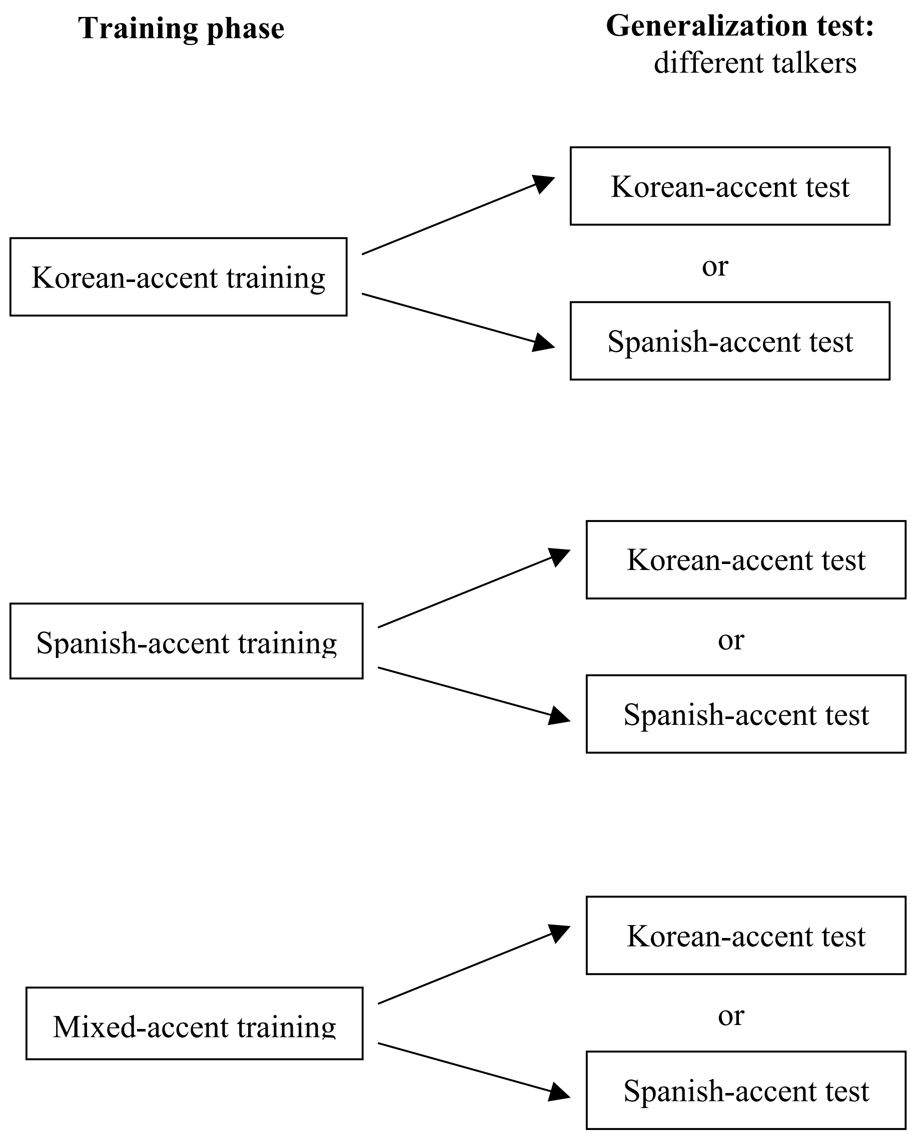


Figure 5.

Experiment 2: Transcription performance across training for listeners trained with Korean-, Spanish-, or Mixed-accented speech.

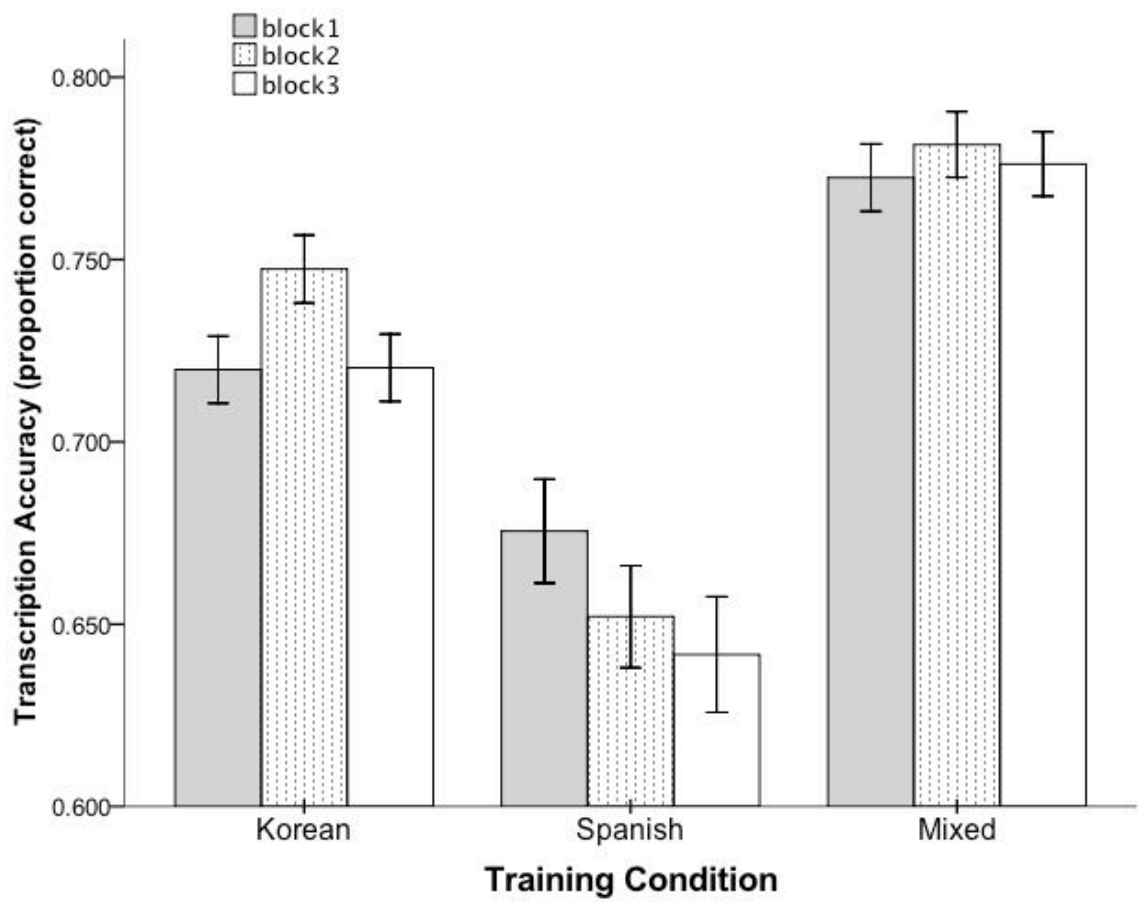


Figure 6.

Experiment 2: Transcription performance at test for listeners trained with Korean-, Spanish- or Mixed-accented speech for both Spanish- and Korean-accented tests.

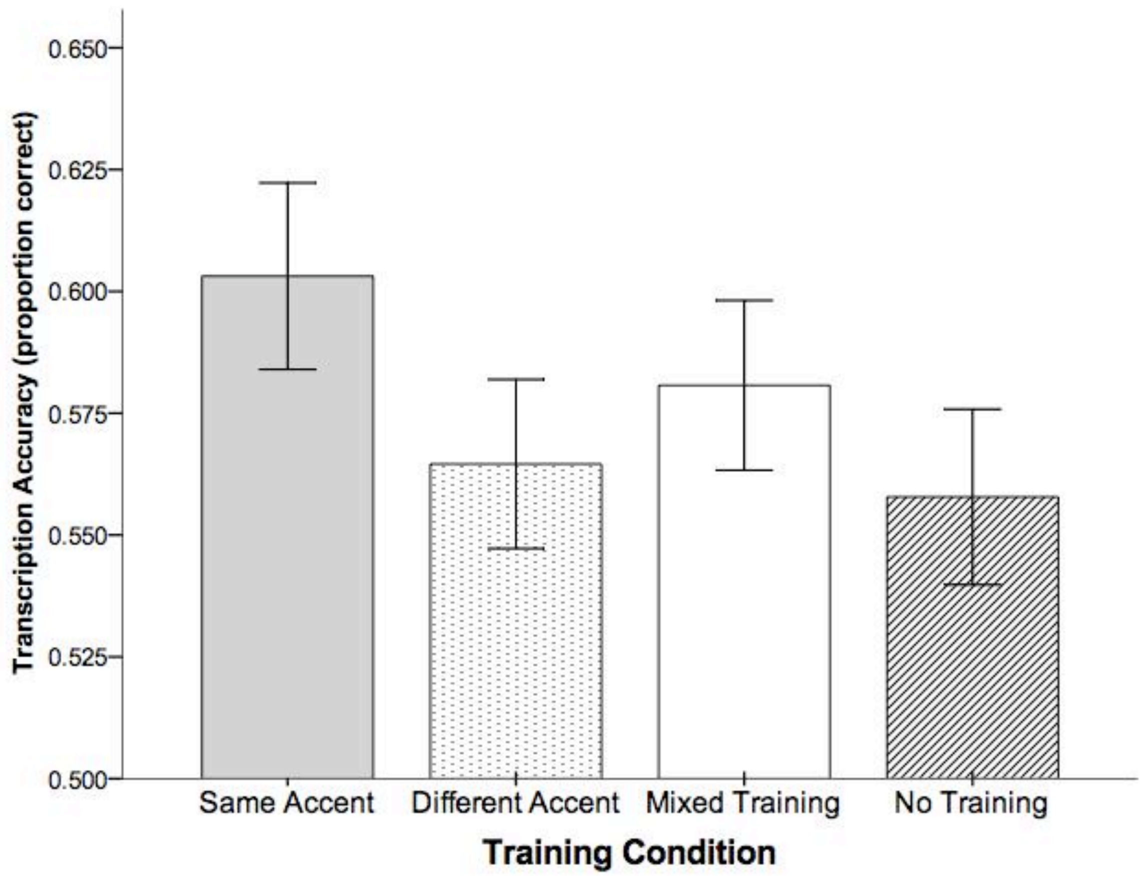


Figure 7.

Experiment 2: Transcription performance at test for listeners trained with Korean-, Spanish- or Mixed-accented speech and tested with Korean-accented speech.

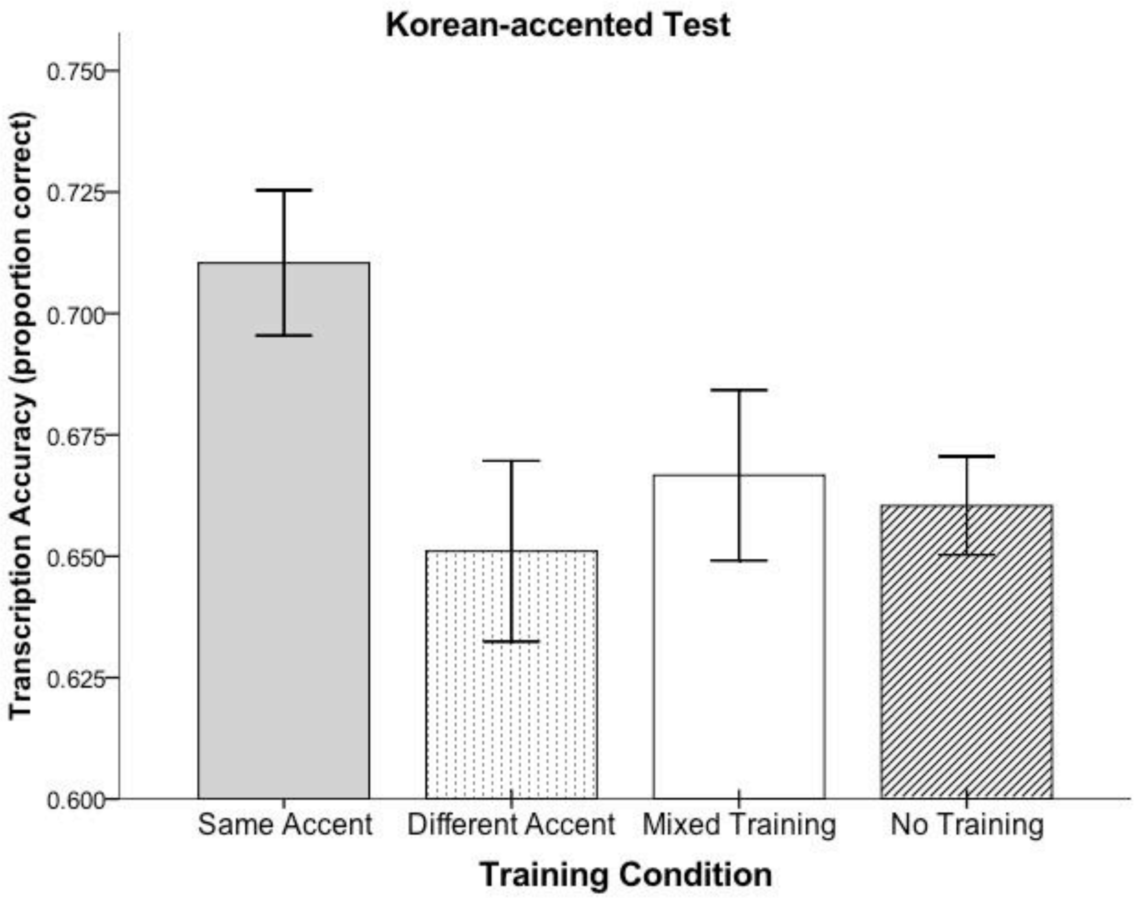


Figure 8.

Experiment 2: Transcription performance at test for listeners trained with Korean-, Spanish- or Mixed-accented speech and tested with Spanish-accented speech.

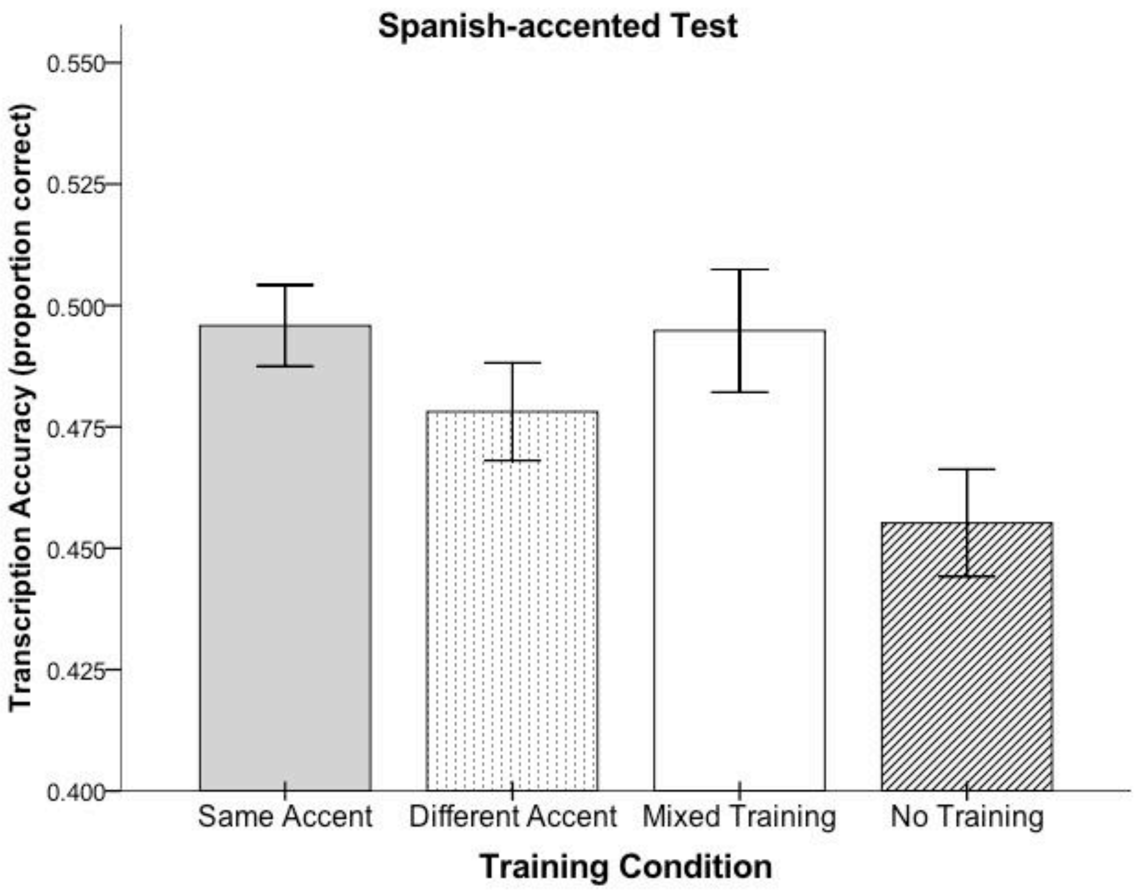
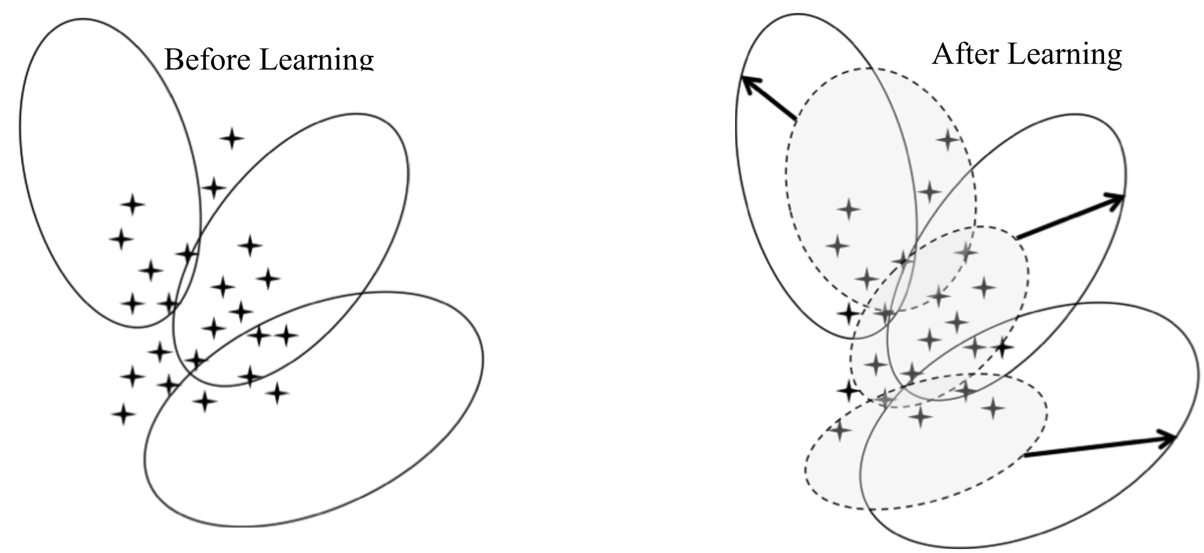


Figure 9.

Diagram of possible acoustic-phonetic category shifts. Listeners may be forming subcategories for the accented speech sounds and/or they may be shifting native category boundaries to include accented items.

Formation of subcategories



Category boundary shifts