Brain Plasticity in Speech Training in Native English Speakers Learning Mandarin Tones

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Christina Carolyn Heinzen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Dr. Yang Zhang

May 2014

Acknowledgements

Funding for this project provided by the Bryng Bryngelson Communication Disorders Research Fund and a CLA Brain Imaging Research Award to Dr. Yang Zhang. Additional thanks to Dr. Yue Wang, who provided the training stimuli and consulting service about training methods.

Many thanks to my advisor Dr. Zhang for his hard work and mentorship, and to my undergraduate assistants, Jiawei Li and Chloe Laabs, for their help with data collection. Additional thanks to my long-suffering and ever-supportive friends and family, who provided much-needed encouragement.

Dedication

This thesis is dedicated to all those hard-working individuals who, like me, struggled to learn a second language in adulthood.

Abstract

The current study employed behavioral and event-related potential (ERP) measures to investigate brain plasticity associated with second-language (L2) phonetic learning based on an adaptive computer training program. The program utilized the acoustic characteristics of Infant-Directed Speech (IDS) to train monolingual American English-speaking listeners to perceive Mandarin lexical tones. Behavioral identification and discrimination tasks were conducted using naturally recorded speech, carefully controlled synthetic speech, and non-speech control stimuli. The ERP experiments were conducted with selected synthetic speech stimuli in a passive listening oddball paradigm. Identical pre- and post- tests were administered on nine adult listeners, who completed two-to-three hours of perceptual training. The perceptual training sessions used pair-wise lexical tone identification, and progressed through seven levels of difficulty for each tone pair. The levels of difficulty included progression in speaker variability from one to four speakers and progression through four levels of acoustic exaggeration of duration, pitch range, and pitch contour.

Behavioral results for the natural speech stimuli revealed significant training-induced improvement in identification of Tones 1, 3, and 4. Improvements in identification of Tone 4 generalized to novel stimuli as well. Additionally, comparison between discrimination of across-category and within-category stimulus pairs taken from a synthetic continuum revealed a training-induced shift toward more native-like categorical perception of the Mandarin lexical tones. Analysis of the Mismatch Negativity (MMN) responses in the ERP data revealed increased amplitude and decreased latency for pre-attentive processing of across-category discrimination as a result of training. There were also laterality changes in the MMN responses to the non-speech control stimuli, which could reflect reallocation of brain resources in processing pitch patterns for the across-category lexical tone contrast. Overall, the results support the use of IDS characteristics in training non-native speech contrasts and provide impetus for further research.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

As technology continues to advance, we find ourselves in a world that is continually shrinking. In this increasingly globalized world, where it is possible to communicate with people on the other side of the globe in a mere fraction of a second, learning a second language (L2) is becoming more valuable than ever. However, learning an L2 as an adult can present a significant challenge. Many adults have watched enviously as children acquire languages seemingly without effort. Meanwhile, the adults arduously pour over mountains of flash cards and contract armies of tutors, yet still struggle to even perceive the unfamiliar speech sounds, or phonemes. Why is learning an L2 as an adult so challenging, and what can be done to ameliorate the difficulty?

This chapter provides the literature review that motivated the current thesis project on the perceptual training of Mandarin lexical tones in nonnative adult listeners. The first section introduces categorical perception and its importance from a cross-language perspective. The second section offers a brief discussion of theoretical accounts for the difficulties for L2 learning in adulthood, including effects of early language exposure, the critical period hypothesis, and the Native-Language-Neural-Commitment theory. The third section introduces lexical tones in Mandarin Chinese and the theoretical as well as practical issues on perceptual training studies. The fourth section describes how neurophysiological measures can add to the study of brain mechanisms underlying speech perception. The fifth section outlines the research questions for the thesis project, which combines behavioral and electrophysiological measures to assess training-induced changes in lexical tone perception.

1.1 Categorical perception

Learning L2 speech sounds requires individuals to identify which acoustic and phonetic features in the acoustic stream are relevant. The listener must then form representations of the new speech categories in their memory stores (Mattock, Molnar, Polka, & Burnham, 2008). This is accomplished through the combined utilization of computational cognitive skills, social skills (Kuhl, 2007), and the auditory system's natural tendencies. These new L2 speech categories form the basis of what is known as categorical perception.

Categorical perception is the ability to perceive continuous acoustic signals as discrete linguistic categories such that there is better discrimination between stimuli that lie in different categories than for equally separated stimuli within the same category (Liberman, Harris, Hoffman, & Griffith, 1957). This phenomenon is believed to be linguistically universal; that is, it is present in native speakers of any language (Zhang, 2013). Categorical perception is characterized by specific patterns in identification and discrimination. When asked to label synthesized stimuli that lie along a continuum, listeners will exhibit a sudden, sharp change at the boundary between categories. For equally separated stimuli pairs, discrimination is better across categories than within categories. Categorical perception aids the listener in decoding the infinite variations in the acoustic input by mapping them onto a finite set of phonemic categories (Liberman et al., 1957). Factors such as language experience, coarticulation, and stimulus length influence this process (Diehl, Lotto, & Holt, 2004).

Because each speaker is different, the listener must have representations of each phonemic category that are normalized across speakers so that they can accurately perceive the speech sounds regardless of the speakers' gender, age, and dialect (Johnson, 2005). This can be especially challenging for adult L2 learners to achieve. Listeners tend to exhibit a perceptual 'bias' toward prior category knowledge (e.g. their native language phonemic categories), so they may have trouble achieving native-like categorical perception in an L2 (Callan et al., 2003; Zhang et al., 2009). A failure to form new categories for the L2 speech sounds leads to the presence of residual characteristics of the first language (L1) phonology (speech sound system) that are present in the learner's L2: a foreign accent. Thus examining categorical perception in a cross-linguistic perspective is one way to reveal phonetic learning. There are many factors that contribute to the failure to achieve native-like perception and pronunciation in the L2; they are described in the following section.

1.2 Acquisition of L2 phonology in adulthood

Early language experience has a considerable impact on later language perception (Iverson et al., 2003; Kuhl & Williams, 1992; Kuhl et al., 2008). In fact, though young infants are able to perceive speech sound contrasts in any language, by six months of age they already exhibit a perceptual bias toward native language contrasts. As children approach their second birthday, their perception of contrasts in their native language improves, while their perception of contrasts in foreign languages decreases (Kuhl & Williams, 1992).

Iverson et al. (2003) postulate that early linguistic experience alters low-level auditory perception, impeding adaptations for higher-level processing of L2 sounds. These early-level changes shape categorical perception. For example, while native English speakers have distinct categories for /r/ and /l/, native Japanese speakers combine both into the same category. This categorization scheme based on early language experience can be difficult to adjust to accommodate the sound system of an L2, even with training (Zhang et al., 2009). While native speakers of a language perceive the speech sounds categorically, for non-native speakers, acoustic contrasts that occur within a speech sound category may be equally salient as contrasts that occur across categories (Hallé, Chang, & Best, 2004).

As frustrating as the influence of early language experience may seem for adults attempting to learn an L2, this process is designed to assist with learning the native language. Experience with a language enhances perception of acoustic features that are relevant to perceiving speech contrasts in that language (Krishnan, Xu, Gandour, & Carianib, 2005), while minimizing those features that are irrelevant (Iverson et al., 2003). These early perceptual skills predict later linguistic skills, with better perception of speech contrasts in the native language leading to faster language advancement in that language (Kuhl et al., 2008). The many aspects of language (e.g. phonetics, syntax, etc.) follow unique developmental trajectories. L2 learning in adulthood presents even more differences in acquisition. The remainder of this chapter will focus on phonetic and phonological acquisition in adulthood, as that is the primary interest of this research project.

Unfortunately, the same process that facilitates acquisition of native language contrasts can interfere with perception of speech contrasts in an L2 (Chandrasekaran, Gandour, & Krishnan, 2007; Iverson et al., 2003; Krishnan et al., 2005; Zhang et al., 2009; Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005). Research using brain-imaging techniques suggests that experience with the native language leads to a 'neural commitment' to the native language, which makes the processing of non-native speech sounds less efficient; this is referred to as the Native-Language-Neural-Commitment Theory (Chandrasekaran et al., 2007; Zhang et al., 2009, 2005).

There are two main theories of L2 acquisition at the segmental level: Best's Perceptual Assimilation Model (PAM) (Best & McRoberts, 2003) and Flege's Speech Learning Model (SLM) (Flege, 1987, 1995, 1997). The PAM focuses on dynamic articulatory information and states that the difficulty in acquiring L2 speech sound contrasts depends upon how those contrasts are mapped onto L1 contrasts. Mature listeners will perceptually assimilate non-native speech sounds to phonemes in their L1 that are the most similar. When perceiving non-native speech sound contrasts, if both sounds fall within the same L1 category, the learner will exhibit poor discrimination. If the sounds fall in different categories, the learner will exhibit accurate discrimination, and so on along the continuum (Best & McRoberts, 2003).

Flege's SLM states that learners are more likely to form new phonetic categories for L2 sounds that are clearly different from L1 sounds. L2 contrasts that are similar to

L1 categories are more difficult to learn (Hao, 2012; Nenonen, Shestakova, Huotilainen, & Näätänen, 2005; Y. Wang, Jongman, & Sereno, 2003; Y. Wang, Spence, Jongman, & Sereno, 1999). Because of this, listeners will have better categorical perceptions for speech sounds that are different from their L1 speech sounds (Piske, MacKay, & Flege, 2001).

The most important predictor of L2 speech sound acquisition is age, with learners who begin at a young age achieving more native-like mastery of the L2. Simply stated, the more fully developed the L1, the more interference it will have in L2 acquisition, leading to a foreign accent (Piske et al., 2001). The clear discrepancy in difficulty between adults and children acquiring an L2 has led some to conclude that there is a 'critical period' for language learning. This Critical Period Hypothesis, first postulated by Lenneberg (1957), states that maturational changes in brain structures used to learn and process language impede learning an L2 due to lost neural plasticity and failure to develop functional neural reorganization (Flege, Yeni-Komshian, & Liu, 1999; Lenneberg, 1967; Piske et al., 2001).

Though this may paint a bleak picture for adults struggling to learn an L2, there is still hope. Recent research has begun to question whether this critical period for language learning is as absolute as its name implies (Birdsong, 1999; Flege et al., 1999; Gullberg, Roberts, Dimroth, Veroude, & Indefrey, 2010; Perani, 1998; Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000; Werker & Tees, 2005). While some young learners never achieve native-like proficiency, some older learners do (Piske et al., 2001). Bialystock and Hakuta (1999) argue that age affects cognitive and linguistic factors that are related to language acquisition, not language acquisition itself. Additionally, Werker and Tees (2005) suggest that the term ‘optimal period’ more accurately reflects L2 learning because at least some level of language-learning abilities are retained into adulthood. Wayland and Guinon conclude that “even after years of research, it remains unclear whether the critical period is conditioned by biological or environmental variables” (2004, p. 683). Further research into adult L2 learning is required to fully understand what variables are critical for native-like acquisition.

In addition to previous linguistic experience and age of acquisition, there are several other factors that contribute to L2 speech acquisition. One might ask whether formal instruction in the L2 is beneficial. Unfortunately, most formal instruction is targeted to overall communication or reading/writing skills rather than speech sound acquisition and therefore has little effect beyond increasing L2 speech sound input (Wayland & Guion, 2004). The L2 learning process is known to be affected by length of residence in the L2-speaking country, motivation, continued use of L1, and shifts in language use (Flege et al., 1999; Sancier & Fowler, 1997). However, while linguistic experience and age of acquisition are known to play a substantial role in L2 acquisition, the relative influence of each of these other factors is uncertain (Piske et al., 2001).

1.3 Lexical tones

In addition to consonant and vowel speech sound contrasts, many languages have lexical tones. Lexical tones are changes in pitch used to differentiate between word meanings (Yip, 2002). For example, the Mandarin word /ma/ can have one of four meanings depending upon the tone associated with it: ‘mother,’ ‘scold,’ ‘hemp,’ and ‘horse’ (Chin, 2006). Tones can be further broken down into level tones, those that maintain a constant relative pitch, and contour tones, which rise, fall, or both. Some estimates state that up to 60-70% of the world’s languages include lexical tones, including Mandarin, which has some 885,000,000 speakers (Yip, 2002). In the United States, a rapid growth in native and non-native populations of tone-language speakers has made knowledge of lexical tone acquisition increasingly relevant (US Census Bureau Public Information, 2010).

Mandarin has four main tones: High (T1), Rising (T2), Falling-Rising or Low-Dipping (T3), and Falling (T4). As each of these tones is relative to each speaker’s pitch range, each tone must be normalized across multiple speakers for categorical perception (Chao, 1947). Native speaking infants are already sensitive to differences in syllabic pitch contours (Nazzi, Floccia, & Bertoncini, 1998). These infants typically acquire T1 and T4 first, and these tones are also typically the easiest for L2 learners to acquire (Wan, 2007). Of all four tones, T3 has the most allophones and is often the most difficult for non-native speakers to perceive (J. Lee, Perrachione, Dees, & Wong, 2007).

Auditory processing of tone information is accessed at a similar point in time as information provided by vowels and consonants, and it contributes to word processing in the same manner (Schirmer, Tang, Penney, Gunter, & Chen, 2005). While pitch is the primary characteristic of lexical tones, other cues, such as duration, amplitude, voice quality (e.g. creaky, breathy) also play a role in lexical tone perception (C.-Y. Lee & Hung, 2008; C.-Y. Lee & Lee, 2010; C.-Y. Lee, Tao, & Bond, 2008).

Perception of lexical tones begins before the listener even attends to the stimuli. In fact, in native speakers, “pre-attentive encoding of abstract auditory rules of lexical tones can predict perception during a later attentive stage” (X.-D. Wang, Gu, He, Chen, & Chen, 2012, p. 3). This pre-attentive abstraction of rules in speech by the auditory sensory intelligence is what enables listeners to perceive speech in noisy environments without drawing on conscious resources (X.-D. Wang et al., 2012). The interaction between memory and language in pre-attentive phonological processing causes difficulty for L2 learners (Sittiprapaporn, Chindaduangratn, Ter Vaniemi, & Khotchabhakdi, 2003).

When considering higher-level cortical processing, pitch information is typically processed in the right hemisphere; however, lexical tones are processed primarily in the left hemisphere in native speakers of tone languages (Y. Wang, Jongman, & Sereno, 2001; Y. Wang, Zhang, Cooper, & Dovan, 2011). In non-native speakers, on the other hand, lexical tones are processed bilaterally (Y. Wang et al., 2001). Xi, Zhang, Shu, Zhang and Li (2010) suggest that the acoustic and phonological information in lexical tones are processed in parallel. Because of this parallel processing, linguistic experience with lexical tones has been shown to influence perception of non-speech sounds (Xi, Zhang, Shu, Zhang, & Li, 2010).

The phonological representations of each tone are stored in the memory as independent units from segments. Like any other aspect of speech and language, lexical tones are subject to errors (Wan, 2007). Just as phonemic awareness training is known to help children to learn phonemes, some have suggested that tone awareness may assist in acquiring tones (Wong, Perrachione, & Parrish, 2007). Lexical tone perception, like perception of consonants and vowels, is influenced by language experience (Peng et al., 2010). Despite the similarities, however, lexical tones form a distinct linguistic entity with a unique developmental trajectory (Mattock & Burnham, 2006; Mattock et al., 2008; Werker & Tees, 1984).

Compared with vowels and consonants, there is a dearth of knowledge regarding lexical tones (Wong, Perrachione, Gunasekera, & Chandrasekaran, 2009). Studying lexical tones can further clarify the discussion of age-based constraints on L2 learning (Piske et al., 2001). Lexical tones can also be used to determine what aspects of speech processing are pre-attentively versus attentively (X.-D. Wang et al., 2012). Furthermore, the study of lexical tones assists in understanding the nature of brain plasticity in L2 acquisition, the underlying interactions between acoustics and abstract categories, and the contribution of foreign phonetic features (see Wong et al., 2009 for review; Wong, Perrachione, et al., 2007). Finally, lexical tones are unique in their relationship to pitch. While pitch is typically processed in the right hemisphere, lexical tones, along with other linguistic stimuli, are processed in the left hemisphere. This presents an interesting opportunity to study the relationship between pitch processing and linguistic processing (Y. Wang et al., 2011).

In addition to more theoretical applications of lexical tone research, there are also practical applications in disorder remediation and L2 learning. Like any other aspect of speech and language, speakers of tone languages are susceptible to disorders, such as Parkinson’s, dysarthria, cerebral palsy, dyslexia, and hearing loss, which can impact their production of lexical tones. Clinicians need to understand lexical tones in order to treat these disorders. Despite the obvious need, there are few assessments available for assessing lexical tone deficits and no normative data on typical lexical tone perception, production, and developmental trajectory (see Wong et al., 2009 for review). Though phonemic awareness is well-established as correlating with language learning, there is little information on tone awareness, and there are no tone awareness sections on standard phonemic and morphological awareness assessments (Wong & Perrachione, 2007).

A second practical application for lexical tone research is L2 acquisition. Behavioral methods have demonstrated that there are individual differences in L2 learning (e.g. Wong, Perrachione, et al., 2007), but few researchers have studied the neurological mechanisms associated with these differences. Furthermore, the most effective training procedures for each learner profile have yet to be established. Further research in this area could lead to more effective pedagogical techniques and, as a result, greater proficiency for L2 learners.

1.3.1 Lexical tone acquisition in non-tone language speakers

Tone languages are notoriously difficult for non-tone language speakers to acquire. To begin, tones are not typically marked orthographically, so there is no reinforcement as the learner reads and writes (Hao, 2012). Furthermore, speakers of Germanic languages, such as English, process pitch at the suprasegmental level, using it more holistically for stress and intonation; tone languages, on the other hand, use pitch for phonemic contrast at the segmental level (Francis, Ciocca, Ma, & Fenn, 2008; Y. Wang et al., 1999). Tones are not as critical a factor in syntagmatic planning in tone languages as stress is in Germanic languages. Tone is also more closely linked to lexical meaning, while stress is more closely linked to phrasal prosody (see Wan, 2007 for review).

Because lexical tones are not salient in English, English-speaking learners must develop novel processes capable of integrating tones and phonetic contrasts (Y. Wang, Jongman, et al., 2003). To perceive lexical tones, the listener must be able to combine the following perceptual cues: height of pitch onset and offset, pitch slope, turning point timing, duration of syllable, amplitude, and voice quality. Because English does not use pitch phonemically, English speakers are more sensitive to pitch height, as opposed to pitch direction and contour as native speakers of tone-languages (Chandrasekaran et al., 2007; Chandrasekaran, Sampath, & Wong, 2010; Kaan, Wayland, & Keil, 2013; X. Wang, 2013; Wayland & Guion, 2004). These differences in perceptual cue weighting shape categorical perception, which further increases the difficulty of L2 acquisition (Callan et al., 2003; Peng et al., 2010; Zhang et al., 2009).

It is commonly thought that native speakers of tone languages are better able to acquire another tone language as an L2 because of their previous experience with tracking pitch variations for lexical tones. However, this is not always the case (X. Wang, 2013). The relationship between L1 and L2 tone languages is far more complex. Even certain similar L1 features, such as pitch-accent, may not transfer to L2 tone acquisition (X. Wang, 2013). Furthermore, the supposed ‘advantage’ of tone L1 speakers may present itself only at certain levels of processing. While experience with L1 tones may make listeners more sensitive to pitch differences between tone pairs, they may actually struggle more when asked to overtly categorize L2 pitch contours (Hao, 2012).

While the overall accuracy of perception by individuals with a tone L1 may be similar to individuals with a non-tone L1, the specific tone confusions are dependent upon interactions with the individual L1s. Speaking a tone L1 may even be detrimental to L2 learning because the listeners must suppress pre-learned L1 tone categories. Lexical tones differ across languages, so speakers of tone L1s may still be placing more importance on different dimensional cues than those required for distinguishing L2 contrasts (Hao, 2012).

Regardless of language background, however, listeners can benefit from perceptual training. Wang (2013) found that, although the perception of Mandarin tones was initially more difficult for Hmong L1 speakers than for non-tone (English and Japanese) L1 speakers, the Hmong learners were not disadvantaged in that they improved significantly with training, just as the others did. While these are encouraging results, the best training procedure for specific learner profiles is yet to be firmly established.

Just as early language experience can influence perception, musical training is a major factor in the acquisition of L2 tone languages. Individuals with a musical background tend to be more successful learners of tone languages as L2s (Sleve & Miyake, 2006; Wong & Perrachione, 2007). Researchers have suggested that musical training may facilitate lexical tone processing because it engenders a more robust and faithful neural encoding of pitch in the subcortical auditory system and further develops the higher cortical structures involved in pitch processing (Wong et al., 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007). Wong and colleagues (2007) suggest that musical training assists in lexical tone processing because the perceptual training with feedback changes the weighting of auditory perceptual cues.

Musical training has been shown to facilitate linguistic pitch processing (Bidelman, Gandour, & Krishnan, 2011; Chandrasekaran, Krishnan, & Gandour, 2009; Chandrasekaran et al., 2010), but the best type of training for learning lexical tones remains to be discovered (Wayland, Herrera, & Kaan, 2010). Training in one domain may not necessarily transfer to the other domain, and the advantage of musical training in lexical tone perception varies as a function of the amount of auditory input (C.-Y. Lee & Hung, 2008; Wayland et al., 2010). Chandrasedaran et al. (2009) suggested that pre-attentive auditory perception is domain general; thus there is an advantage with musical training. Attentive perception, on the other hand, is sensitive to categorical representations of speech sounds (Chandrasekaran et al., 2009).

1.3.2 Perceptual training studies

The complexity of the perception of speech sounds has many implications for adults who are attempting to learn an L2. As discussed earlier, there are multiple factors that influence a person’s ability to perceive speech sound contrasts in a foreign language. Despite this, however, Flege (1999) states that the difficulty adults face is not that they have lost the ability to learn new speech sounds, but that they have already learned their L1 sounds so well. If this is true, then it should be possible to train adults to be able to overcome the neural commitment to their L1 and acquire their L2 speech sounds. Indeed, laboratory training procedures for adults have been shown to improve perception of both trained and untrained stimuli, demonstrating that the auditory perceptual system remains at least somewhat malleable over the lifespan (Wayland & Guion, 2004).

Training can improve perception of lexical tones by shifting perceptual cue weighting and neural activation (see Kaan et al., 2013 for review). Perceptual training can lead to changes in neurological responses of the auditory system to speech contrasts (Y. Wang, Sereno, Jongman, & Hirsch, 2003; Zhang et al., 2009). Learning-induced plasticity also leads to changes in higher cortical structures; for example, the recruitment of additional areas specialized for functions similar to new language functions (Y. Wang et al., 2011). Additionally, perceptual training has been shown to lead to increased neural sensitivity and increased neural efficiency, which allows for the allocation of resources to more abstract linguistic functions (Y. Wang et al., 2011; Zhang et al., 2009). It has also been shown to improve production of both trained and untrained speech contrasts (Y. Wang, Jongman, et al., 2003).

There have been many training studies on acquisition of L2 speech sound contrasts using a variety of training methods to assist various learner populations (See Appendix A for table) (Chandrasekaran et al., 2010; Francis et al., 2008; Kaan, Wayland, Bao, & Barkley, 2007; Kraus et al., 1995; Y. Wang, Jongman, et al., 2003; Y. Wang, Sereno, et al., 2003; Zhang et al., 2009). Training can be brief or intensive, address perception and production individually or in combination, target the auditory domain exclusively or include visual information, be conducted at the phonemic or word level, and involve a variety of inter-stimulus intervals (X. Wang, 2013; Wong & Perrachione, 2007). Computer training programs have been used to provide more, high-quality exposure in less time, meaning that fewer hours of training are required before observing results (Thomson, 2012; X. Wang, 2013; Wu, Yang, Lin, & Fu, 2007).

Studies also vary in their criteria for ending training. Many studies limit their training to a certain number of sessions, but some researchers have suggested that this practice means learners may not be reaching their full potential. They suggest using proficiency level instead (Wong & Perrachione, 2007). Given the wide variety of training methods and training objectives, it is no surprise that researchers have yet to encounter the ideal training procedure.

One decision that must be made when developing perceptual training programs is whether or not to include any sort of visual cues. Visual cues range from video of the speaker to simple arrows representing the pitch contours. Chen and Massaro (2008) found that there is some useful information about Mandarin lexical tones present in visible the head, neck, and mouth movements and that visual lexical tone identification could be improved with training. That being said, there is a certain exchange between additional information provided by visual cues and the increased repetitions possible without the visual cues (X. Wang, 2013).

There is a large degree of individual variability in the benefits of visual information (Massaro, 2001). Some researchers have suggested that the ability to use visual information for speech perception is dependent upon prior experience using it (Ronnberg et al., 1999). Burham, Lauw, Lau, and Stokes (2000) found that there was no enhancement in lexical tone perception in an auditory-visual condition when compared to an auditory-only condition. As a full video of the speakers' heads and necks is not feasible in the current study's design, a static picture of each speaker will be the only visual cue.

Additionally, research has demonstrated that using the characteristics of Infant-Direct Speech (IDS) may facilitate L2 perceptual learning (Iverson et al., 2003; Zhang et al., 2009). IDS, also known as 'motherese,' is the exaggerated speech caregivers use when addressing infants; it has been shown to attract the infants' attention and to facilitate positive affect throughout the interaction (Werker & McLeod, 1989). This speaking style, with its simplified language, slower rate, high degree of repetition, and exaggerated intonation, may assist infants as they form speech perceptual categories from the distributional properties of the acoustic input (Maye, Werker, & Gerken, 2002; Werker et al., 2007). For native speakers of Mandarin, IDS includes exaggerated tone pitch and duration, while still maintaining the critical cues for distinguishing between the tones (Liu, Tsao, & Kuhl, 2007). By incorporating these characteristics into a perceptual training program, it may be possible to facilitate the acquisition of Mandarin lexical tones by non-tone L1 speakers.

Along with the variables discussed above, it is important to consider the degree of stimulus variability in training procedures. High-variability training generally results in the best outcomes, though there is still a great degree of individual variability (Chandrasekaran et al., 2007; Iverson et al., 2003; Y. Wang et al., 1999; Wong & Ettlinger, 2011). The individual variability demonstrates that, like any other area of speech, language, and hearing sciences, perceptual tone training cannot be formatted into a ‘one-size-fits-all’ design. While great progress has been made in developing training procedures, there is still much research needed to further refine them. For example, while some studies found that the effects of training generalized to untrained stimuli, others did not (Kaan et al., 2007; Zhang et al., 2009). Furthermore, while training can lead to improved behavioral identification, it does not always lead to a native-like categorical perception of speech sounds (Zhang et al., 2009). These caveats highlight the need for more research in perceptual training methods.

1.4 Behavioral and neurophysiologic measures

As with any method, there are advantages and disadvantages to using behavioral and brain measurements. Behavioral measures have the distinct advantage of being readily applicable to ‘real-world’ outcomes. That is to say, measuring an individual’s accuracy in producing lexical tones readily translates to that person’s ability to produce those tones in conversations outside the laboratory setting. However, behavioral methods can be less precise, especially when utilizing listener ratings (Piske et al., 2001). Furthermore, behavioral measures of timing are often confounded by other processes, such as attention, working memory, motor response capabilities, motivation, and decision making (Foster et al., 2013).

In contrast, brain measures are not necessarily hindered by such confounds. Brain measures can be used to determine accuracy separately for each dimension (e.g. frequency, intensity) (Näätänen et al., 2012). Brain measurements can also measure pre-attentive processing, which would be impossible using behavioral measurements

(Sittiprapaporn et al., 2003). Many brain measurements are non-invasive, and they correspond well to behavioral measures (Näätänen et al., 2012). Together these factors illustrate how brain measurements can provide additional information not revealed by behavioral measures.

Despite the many advantages to using brain measures, there are disadvantages. In order to utilize brain measurements, researchers must have access to what is often expensive equipment. Brain measure techniques can be time-consuming and require a significant level of expertise (Näätänen et al., 2012). Researchers must not only know how to use a variety of complex measuring techniques, they must know when each is appropriate. Finally, though brain measurements are precise, they may not always correlate to real-world results. A significant improvement in a brain measurement may not lead to a practically significant improvement for the learner.

1.4.1 MMN

Despite the disadvantages, brain measures remain a valuable tool for studying language. Studies focusing on the time course of auditory processing often utilize an electroencephalogram (EEG) because of its refined temporal resolution (Kuhl, 2010; Wong et al., 2009). The Mismatch Negativity (MMN) component of the Event-Related Potential (ERP) waveform is a commonly- used method for assessing pre-attentive auditory processing (Näätänen et al., 2012; Näätänen, Jiang, Lavikainen, Reinikainen, & Paavilainen, 1993). It is elicited when an anomalous stimulus is presented amidst a series of presentations of some standard stimulus. This is referred to as the Oddball Paradigm.

The anomalous, or ‘deviant,’ stimulus can vary along various parameters, including duration, pitch, and inter-stimulus interval.

The MMN is derived by comparing neural responses to the ‘deviant’ stimulus to that of the standard, and usually involves a negative peak around 100-250 milliseconds following the onset of the deviant stimulus (Näätänen et al., 1993). These early neural processes involved in detecting a change in the stimulus occur before frontal lobe activation directs attention to the change. This suggests that the change is detected in a forward-looking process where the actual input is compared with the input that was predicted based upon regularities in the previous stimuli. Therefore, MMN responses reflect the pre-attentive capabilities for encoding physical features of repetitive stimulus and abstract attributes, such as direction and frequency change (Näätänen, 1990; Näätänen et al., 2012, 1993).

Because it is designed to examine pre-attentive processing, the MMN is a passive measure. The participant is engaged in a distracter task, such as watching a silent movie, while the stimuli are presented and the responses measured through scalp electrodes. For this reason MMN analysis can be used in developmental studies involving very young infants, or treatment studies involving severely impaired clients for whom behavioral measures are difficult to obtain (Foster et al., 2013; Näätänen et al., 2012). In studies of healthy adults, a pre-attentive measure like MMN is useful for avoiding the limitations presented by a wandering attention.

Studies of MMN responses indicate that detection of auditory deviance is organized hierarchically, with peak responses to deviants to simple and complex regularities occurring within different time frames (Cornella, Leung, Grimm, Escera, & Mansvelder, 2012). Another relevant component of the ERP is the P3a, a positive ERP with a fronto-central distribution peaking around 270 milliseconds. Like the MMN, the P3a is measured by comparing neural responses to the deviant and standard stimuli. However, while the MMN measures pre-attentive processing, the P3a indicates an involuntary attention shift, which can be useful to researchers analyzing MMN.

Like many brain responses, MMN responses are influenced by acquired knowledge. There is a more robust response to words and to phoneme changes that are relevant to the listeners' native language than to non-words and non-native phoneme changes. This suggests that the MMN in part involves cortical memory traces from previous language experience (Salisbury, 2012; Sittiprapaporn et al., 2003) Linguistic tones engender a long latency in the MMN because of the integration of pitch variation and linguistic information over time. This interaction with linguistic information may be why native speakers exhibit longer latencies than non-native speakers, for whom lexical tones have no linguistic element (Y. Wang et al., 2011). For this same reason, native tone language speakers are more accurate in discrimination tasks with a longer inter-stimulus interval, while non-native listeners are more accurate with shorter inter-stimulus intervals (Wayland & Guion, 2004). These characteristics make MMN an effective method of studying the L2 speech sound acquisition and the influence of early language experience.

Regardless of etiology, MMN responses can indicate decreased discrimination accuracy, decreased sensory-memory duration, abnormal perception and attention, or cognitive decline (Näätänen et al., 2012). The MMN tends to shorten in peak latency and increase in amplitude with improved behavioral discrimination (Kraus et al., 1995; Tamminen, Peltola, Toivonen, Kujala, & Näätänen, 2013). This makes MMN an ideal measure for training studies, as it can be utilized to demonstrate changes in perception that occur as a result of training (e.g. Kaan et al., 2007).

1.5 Research questions

The research design for the current project follows Zhang et al. (2009), which demonstrated that substantial neural plasticity for L2 phonetic learning can be induced in adulthood by intensive training in a laboratory setting. This study aims to answer the following questions with regard to native English speakers with no previous experience with tone languages and no formal musical training:

1. Will a computer-based perceptual training program utilizing the characteristics of IDS lead to improved behavioral perception of Mandarin lexical tones?
2. Will the same training program lead to a shift towards categorical perception of the Mandarin lexical tones?
3. Will the training program lead to pre-attentive neural responses to Mandarin tones that reflect perceptual improvements?

Previous research on the infant-directed speech and child-directed speech in Mandarin has shown age-related changes in terms of acoustic exaggeration in the language input that a child receives (Liu et al., 2007; Liu, Tsao, & Kuhl, 2009). In comparison with adult-directed speech, the Mandarin lexical tones in IDS are longer in duration with higher overall pitch as well as wider pitch range and more exaggerated pitch contour. The training program will utilize an exaggeration method focusing on four levels of stimulus duration, pitch range, and pitch contour. The training software was developed by my advisor, Dr. Zhang. Modified speech input using durational manipulations was previously employed in the Fast ForWord software program for treating children with learning disabilities (Scientific Learning Corporation, 1999); however, it remains to be tested whether durational manipulation in combination with spectral manipulations is applicable to L2 phonetic training for Mandarin lexical tones.

We will use both behavioral and MMN measures to quantify the learning process. With regard to the first question, Zhang et al. (2009) utilized this type of training procedure to effectively train Japanese participants in the perception of non-native consonant (English /l-r/) contrasts (Zhang et al., 2009). However, no previous studies have utilized the exaggeration method to train perception of non-native lexical tone contrasts. Given the previous results, the participants in this study are expected to improve their behavioral accuracy and decrease their reaction time following training.

Regarding the second question, participants are expected to improve their discrimination accuracy for across-category stimulus pairs, but not for within-category stimulus pairs. This would reflect a shift towards categorical perception. However, participants may not necessarily exhibit exact native-like categorical perception boundaries following training, as previous studies have not produced this result (Zhang et al., 2009).

Regarding the brain measures, studies have shown that shorter MMN peak latencies and increased peak amplitudes accompany increased behavioral discrimination (Kraus et al., 1995). These results are expected for the across-category speech stimuli but not for the within-category speech stimuli in the present study. Changes in responses to the non-speech stimuli are expected to be insignificant (Zhang et al., 2009), or to reflect the changes in the speech stimuli due to a transfer of learning from speech to non-speech stimuli (Xi et al., 2010).

Chapter 2: Methods

2.1 Participants

The experimental procedure, including subject recruitment, followed protocols approved by the Institutional Review Board at the home institution. Participants in the current study were college students at the University of Minnesota Twin Cities campus (See recruitment flyer in Appendix B). Eleven right-handed, monolingual speakers of English (six female, five male) ages 18-26 with no prior experience with Mandarin Chinese participated in this study. According to self-reports, all participants had hearing acuity that was within normal limits, none had any history of speech and language disorders, and none had greater than two years of professional music training.

Each participant read and signed an informed consent form prior to the experiments. Participants were compensated at a rate of ten dollars per hour of participation. One participant withdrew from the study due to subsequent difficulty in scheduling the training and post-test sessions. Participants whose pre-test data already closely mirrored the accuracy of native speakers in the identification and discrimination data of the synthetic speech stimuli (Xi et al., 2010) were eliminated from analysis (See Appendix E). Participants who did not have sufficient number of ERP trials due to excessive blinking and other muscle movement artifacts were also excluded from further analysis. In total, there were seven participants (4 females and 3 males) whose data were included in the statistical analysis.

2.2 The training program

The experimental protocol and training sessions closely followed the approach in Zhang et al. (2009), which incorporated the characteristics of IDS into a computer training program to enhance perception of non-native speech sound contrasts. Some characteristics of IDS that have been suggested for inclusion in perceptual training include: 1) signal enhancement to exaggerate spectral and temporal cues for pitch and phonetic categories, 2) adaptive speech modification based on learning, 3) visible articulation cues to induce cross-modality learning, 4) a large stimulus set with a high degree of variability, and 5) self-initiated selection without the overt requirement of identification (Zhang et al., 2009).

Zhang and Cheng (2011) pointed out a number of limitations in the previous software version for speech training as used in Zhang et al. (2009). One limitation was portability as the previous program was implemented using Macromedia Authorware on an Apple PowerPC. Another was data management in that data collection saved separate output files for pre-test, post-test, and each training session for each trainee without a systematic database structure. A third limitation was scalability as the program could not be easily adapted for training speech stimuli other than the English /r-l/ contrasts and could not be implemented for remote access/learning over the internet.

The current implementation was based on an improved Speech Assessment and Training (SAT) version to overcome the limitations (Cheng & Zhang, 2013; Zhang & Cheng, 2011). The SAT program used the JAVA script language that could be used across computer platforms and accessed either online or offline via internet. A database structure was implemented using Microsoft ACCESS, which allowed efficient data management, access, retrieval, and analysis. The program also implemented a user interface that included login ID and password integrated with the database program, which could allow secure online sessions for future large scale speech training. These new features in the SAT program would allow more research/teaching options and practical applications for L2 learners.

Speech stimuli in the training sessions for the current project were real monosyllabic Mandarin words including all four Mandarin tones: High Level (T1), Rising (T2), Low-Dipping (T3), and Falling (T4), in Consonant-Vowel (CV) and CVNasal contexts. The naturally produced speech stimuli in the training program were previously used in a Mandarin tone training study (Wang et al., 1999), which included productions of four native speakers of Mandarin (two male, two female).

The natural productions were then analyzed and resynthesized in Praat (www.fon.hum.uva.nl/praat/) to follow an adaptive training procedure in which training stimuli were acoustically modified to four levels of exaggeration (Zhang et al., 2009). For Level 1, duration and pitch range were exaggerated by 220%, and pitch height was exaggerated by 124%. For Level 2, duration and pitch range were exaggerated by 180%, and pitch height was exaggerated by 116%. For Level 3, duration and pitch range were exaggerated by 140%, and pitch height was exaggerated by 108%. Level 4 consisted of the natural speech productions with no exaggeration. These exaggeration levels were chosen based on a consensus from three native Mandarin speakers about the quality of modified speech such that the stimuli could be exaggerated, but not distorted so much that they no longer resembled speech.

In the training sessions, static pictures of each speaker were used as visual cues instead of full videos of the speakers' heads and necks (See Appendix D). We utilized exaggeration of three parameters that are important for the perception of lexical tones: syllable duration, pitch range, and pitch height. As mentioned in the previous chapter, modified speech input using durational manipulations was previously employed in the Fast ForWord software program for treating children with learning disabilities (Scientific Learning Corporation, 1999). The software received mixed reviews in this capacity (Strong, Torgerson, Torgerson, & Hulme, 2011). It remains to be tested whether durational manipulation is applicable to L2 phonetic training in adulthood and how the perception of lexical tones may be facilitated by durational exaggeration in combination with spectral exaggeration of the pitch in the speech input that the listener receives.

2.3 Stimuli in Pre- and Post- Tests

The speech stimuli in the pre- and post- tests included natural speech and synthetic speech. To assess generalization of training effects to untrained stimuli, the natural speech stimuli in the pre- and post-testing not only included one speaker that was not used in the training sessions, but also stimuli in novel phonetic contexts there were not used in the training (Wang et al., 1999).

The synthesized speech stimuli were an 11-step continuum between /pa/ with Tone 2 and /pa/ with Tone 4, which were previously used in an ERP study to assess categorical perception of lexical tones (Xi et al., 2010; Figure 1). The original stimuli were recorded from a female native speaker at a sampling rate of 44.1 kHz. They were then digitally edited to have a duration of 200 milliseconds each using SoundForge (SoundForge9, Sony Corporation, Japan). Next, pitch tier transfer was performed using the Praat software (www.fon.hum.uva.nl/praat/). This produced two stimuli /pa2/ and /pa4/ which were identical apart from lexical tone. These identical stimuli were then used as endpoints in a 10-interval lexical tone continuum using a morphing technique in STRAIGHT Matlab (Kawahara, Masuda-Katsuse, & de Cheveigne, 1999; Mathworks Corporation, USA). All 11 stimuli were normalized in root mean square intensity.

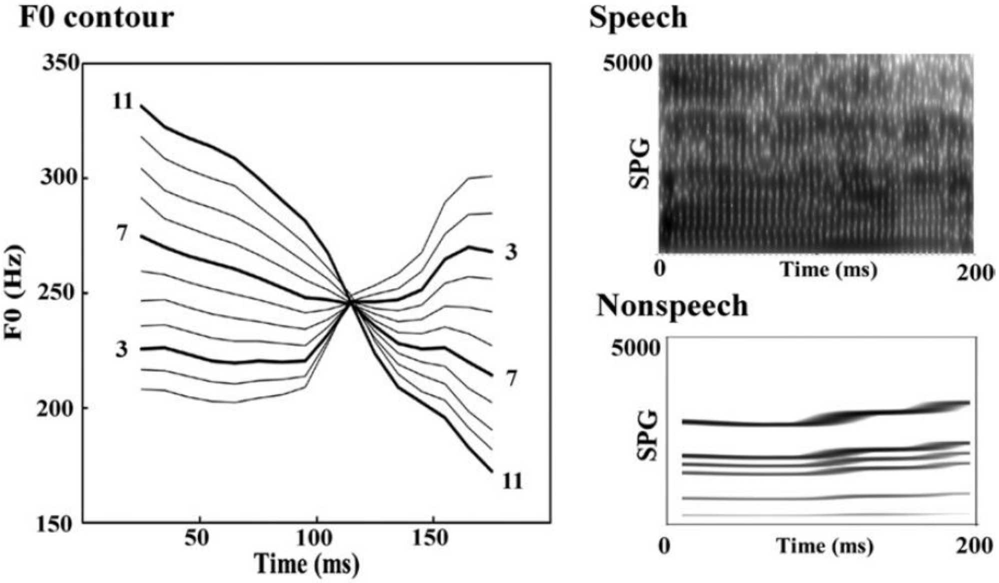


Figure 1. Schematic representation of the synthetic speech and nonspeech stimuli. The left panel shows the F0 variations along the lexical tone continuum, and Stimuli 3, 7 and 11 are highlighted as 3 and 7 form the across-category pair and 7 and 11 form the within-category pair based on identification data of native Mandarin Chinese speakers (reproduced from Xi et al., 2010).

Non-speech stimuli that tracked the fundamental frequency contours in the synthetic speech were also created as control stimuli to properly assess linguistic vs. acoustic processing. The synthesis method for creating acoustic control stimuli of lexical tones followed Xu, Gandour, and Francis (2006). The non-speech stimuli were harmonic tones with the same pitch, amplitude, and duration as the speech stimuli; they differed only in their spectral components. Harmonics 2, 4, 5, 9, 10, and 11 were omitted to decrease the perceptual similarity with the speech stimuli. This left six equal-amplitude harmonics of f0: 1, 3, 6, 7, 8, and 12. These harmonics were selected to reduce the perceptual resemblance of "speechness" (Xu et al., 2006). All the stimuli were normalized to have the same duration and average RMS intensity as the speech stimuli. The same set of nonspeech stimuli were used in a previous ERP study (Xi et al., 2010).

2.4 Pre-test

The pre-test was conducted approximately one week prior to training. The pre-test consisted of an EEG recording session and behavioral perceptual tests.

2.4.1 EEG recording

During the EEG recording session, the participants sat in a comfortable chair in an acoustically and electrically treated booth (ETS-Lindgren Acoustic Systems). A stretchable cap with electrodes sewn into it was fitted on the participants' head. Continuous EEG data were recorded (bandwidth = 0.016–200 Hz; sampling rate = 512 Hz) using the ASA-Lab system with REFA-72 amplifier (TMS International BV) and a 64-channel WaveGuard cap (ANT Inc., The Netherlands). The EEG cap used shielded wires for 65 sintered Ag⁄AgCl electrodes (including the ground channel) in the international 10–20 montage system and the intermediate locations. The ground was positioned at AFz. Adjustments on individual electrodes were made to keep impedances at or below 5 k Ohm. Active shielding technology on the WaveGuard cap allows high-quality EEG recording in conditions where skin impedances are relatively higher than conventional standards (Rao, Zhang, & Miller, 2010).

Participants were engaged in a distracter task of watching a silent movie of their choice while the test stimuli were presented binaurally via Etymotic Research ER3A headphone system with 3M E-A-RLINK ear tips. They were instructed to sit as still as possible, to ignore the stimuli, and to attend to the film and the subtitles on the movie screen. The 20-inch LCD TV monitor was placed approximately two meters away from the chair. Prior to the session, participants who wore contacts were instructed to wear glasses if possible to minimize excessive blinking.

Test stimuli included 3, 7, and 11 from the 11-step synthesized speech continuum between Tone 2 and Tone 4. The test stimuli were presented at 60 dB sensation level following the Double Oddball Paradigm in Xi et al. (2010), in which an MMN response was elicited when a series of some standard stimulus (step 7) is interrupted by deviant stimuli (steps 3 and 11). The standard stimuli were presented at 80% of the total trials and the two deviants were each presented at 10% of trials. Deviants were never presented consecutively. The inter-stimulus interval ranged from 600-700 milliseconds. The same procedure was used for the non-speech stimuli. The presentation order for the two blocks of the speech and non-speech stimuli for the EEG experiment were counter balanced among the participants.

Following the EEG test, the participants were directed to a hair washing station with shampoo, towels, and a hair dryer to wash the conductive gel out of their hair. The entire procedure for the EEG experiment without counting the behavioral test lasted approximately 1.5 hours.

2.4.2 Behavioral pre-test

Behavioral pre-tests included identification and discrimination sessions for the synthetic speech stimuli, a discrimination session for the synthetic non-speech stimuli, and an identification task (in two sessions) for the natural speech stimuli. Each pre-test session lasted approximately 15-20 minutes. During all the behavioral pre-tests, participants sat at a Dell computer with a customized Direct IN-PCB keyboard (Empirisoft Corporation) which recorded button responses with sub-millisecond accuracy. The test stimuli were presented auditorily in random order through Etymotic Research ER3A headphone system with 3M E-A-RLINK eartips at 60 dB sensation level. No visual articulation cues and no corrective feedback were provided during any of the behavioral pre-tests. For all of the pre-tests, behavioral responses were transmitted to a Dell computer and saved in a coded file.

The identification task for the synthetic speech stimuli required the participants to identify whether a rising or falling tone had been presented. Each sound from the speech continuum was presented 20 times in random order. The inter-stimulus interval for the stimulus pairs ranged from 1500-1700 milliseconds. The discrimination task for the speech stimuli was designed to determine whether discrimination abilities were enhanced when sounds resided in different tone category boundaries. Participants were presented with pairs of sounds with a silent interval of 250 ms in between from the speech continuum and asked to indicate whether the two sounds were the same or different. Seven sound pairs were presented 10 times each in random order. Pairs 11-7 and 7-11 represented pairs for which both stimuli resided in the same tone category, a within-category difference, based upon data from native speakers (Xi et al., 2010). Pairs 7-3 and 3-7 represented pairs where the stimuli resided in different tone categories, an across-category difference. Pairs 11-11, 3-3, and 7-7 were used as foil trials. The inter-stimulus interval ranged from 1500-1700 milliseconds. The discrimination task for the nonspeech control stimuli followed the same design as the discrimination task for the synthetic speech stimuli. Each of the speech stimulus pairs were replaced with the corresponding nonspeech stimulus pairs from the synthetic continuum.

No discrimination task was used for the natural speech stimuli. There were two sessions of pre-test identification. In the first session, the natural speech stimuli consisted of 120 natural productions of monosyllabic CV and CVNasal Chinese words. The stimuli contained 30 exemplars of each of the 4 tones, produced in different phonetic contexts than the stimuli used in training. The stimuli were produced by two native speakers (one male and one female). One speaker's productions were excluded from training entirely, while the other speaker's productions were used in training, but in different phonetic contexts than for the productions used to assess generalization. The second session used 200 natural productions of monosyllabic CV and CVNasal Chinese words produced in different phonetic contexts. There were 50 exemplars for each tone. In both identification sessions, participants were asked to identify which of the four tones had been presented. The test sessions were semi-self-paced in that the next stimulus would be presented when the participant had responded to the previous stimulus. Participants were allotted a maximum of 5 seconds to respond to each stimulus.

Throughout the experiments, an experimenter observed the participants from a video monitor in the control room via an intercom system in order to ensure proper data collection and timely correction of problematic issues. The monitoring video was not recorded.

2.5 Training sessions

Each participant underwent approximately two to three hours of perceptual training distributed in hour-long sessions over the course of one-to-two weeks. The training consisted of a software program modified from Zhang et al. (2009) to train lexical tone contrasts (See Appendix D for sample screenshots). The software was run on a Lenovo U430 laptop computer using Sennheiser headphones (HD 212 Pro) to ensure high-quality audio. Stimuli were presented binaurally at approximately 70 dB SPL. Each participant completed the training at his or her own pace in a sound-treated booth. The program incorporated the following characteristics:

1. Self-directed listening: The participants selected the stimuli by clicking on iconic buttons depicting the tones. In this way, the presentation was self-paced by each participant. If the participant selected the same icon 30 times, an automatic message would instruct the participant to select a different icon.
2. Visual Cues: a static photograph of the speaker was situated in the bottom left portion of the screen.
3. Large Stimulus Sets with High Variability: A total of 480 different items spoken by 4 speakers (2 male, 2 female) were used for training. The items consisted of 30 exemplars of each tone at each of the four levels of exaggeration.
4. Adaptive Scaffolding: The training followed a pair-wise paradigm. There were six total contrasts trained in the following order: T1-T2, T1-4, T2-T4, T1-T3, T3-T4, T2-T3. This order was selected such that the training progressed from the least difficult contrasts to the most difficult contrasts according to (Y. Wang et al.,

1999). Training for each contrast progressed through seven levels of difficulty. The first level contained the most exaggerated stimuli spoken by only one speaker. The second, third, and fourth level maintained the exaggeration level, but added one additional speaker with each level. The fifth, sixth, and seventh level maintained all four speakers but decreased the exaggeration, finally using natural speech in the seventh level. For each level, there were 30 exemplars of each tone. Participants had to pass a short identification quiz at the end of each level with 90% accuracy to move on to the next level. Each quiz presented five trials of each tone in the pair, randomly drawn. If the participant failed the quiz, the repeated the level once more before moving on.

2.6 Post-test

A post-test identical to the pre-test was conducted approximately one week following training. Like the pre-test, the post-test consisted of both EEG and behavioral measures.

2.7 Outcome measures

1. Behavioral effects on trained and untrained natural and synthetic speech stimuli.
    a. Accuracy: Discrimination and identification accuracy was measured in terms of percent correct, taking into consideration correct identification, false positives, correct rejection, and misses.

   b. Reaction Time: Reaction time was measured in milliseconds. The pre- and post-test reaction times were compared to determine if there was a significant reduction.
   c. Identification function slope and boundary location for the synthetic continuum: Participants' perceptual boundaries between Tones 2 and 4 were compared to those of native speakers to determine whether the training induced native-like categorical perception.
2. MMN Responses
   a. Peak latency: Peak latency was measured in milliseconds post-stimulus onset.
   b. Amplitude: Mean amplitude was calculated for a window of 40 milliseconds around the MMN peak.

2.8 Data analysis

2.8.1 Behavioral data analysis

Behavioral identification and discrimination scores were evaluated in terms of percent correct accuracy, taking into account all possible response categories (hits, correct rejections, misses, and false alarms). Percent correct identification for each tone for the pre- and post-tests were compared using paired one-tailed t-tests to determine whether there was a statistically significant improvement. Reaction time for the natural speech condition pre- and post-training was also compared using t-tests.

The identification data from the synthetic speech continuum were analyzed to determine the location of the phonemic boundary, the slope of the boundary, and the width of the boundary. These values were calculated using the probit analysis of individual identification curve (Finney, 1971). The boundary location was defined as point with 50% crossover, meaning for half of the trials the stimulus was identified rising, and for the other half of the trials, it was identified as falling. The boundary width was defined as the linear distance between the 25th and 75th percentiles as determined by the mean and standard deviation obtained from probit analysis (Hallé et al., 2004). In order to better fit the asymptotic property of the probit function, we replaced the 0% with 0.1% and the 100% with 99.9% for both respective ends of individual identification curves. The slope, location, and width of the phonemic boundary were compared between pre- and post-tests using a parametric t-test and nonparametric Wilcoxon signed-rank test to determine whether training engendered more native-like categorical perception.

Finally, discrimination of sound pairs from the synthetic speech and non-speech stimuli was analyzed in terms of percent correct discrimination for both within- and across-category sound pairs using the formula described by Xu et al.(2006). All four types of pairwise comparison (AB, BA, AA, and BB, where stimuli A and B were separated by two steps on the continuum) were used. Stimuli pairs AB and BA represented ‘different’ pairs, and pairs AA and BB represented ‘same’ pairs. Adjacent comparison types contained overlapping AA or BB trials. Accuracy (P) for each comparison type was defined by P = P(“S|S) x P(“D”|D) x P(D), where P(“S”|S) represented the percentage of correct ‘same’ responses to ‘same’ pairs. Similarly, P(“D”|D) represented the percentage of correct ‘different’ responses to ‘different’ pairs. P(S) and P(D) represented the overall percentages of ‘same’ and ‘different’ pairs in each type, respectively. Pre- and post-test values were compared using paired one-tailed t-tests to determine whether the training resulted in increased sensitivity to across-category differences and decreasing or no change in sensitivity to within-category differences (Xi et al., 2010).

As all of the participants reached at least 90% accuracy during each training session, the session-by-session accuracy data from the quizzes in between training sessions were not reported here.

2.8.2 MMN analysis

Analysis of the averaged ERP data from the individual subjects was focused on the MMN response. ERP averaging was performed offline in BESA (Version 6.0, MEGIS Software GmbH, Germany). Artifact correction was first applied to the raw EEG data to minimize influences due to horizontal and vertical eye movements (eye drift and blink, respectively). The correction parameters were set at 100.0 µV for horizontal electrooculogram (HEOG) and 150.0µV for the vertical electrooculogram (VEOG). This allowed for the preservation of trials that may otherwise have been obscured by eye movements. The EEG data were then bandpassed at 0.5–40 Hz. The ERP epoch length was 800 ms, including a pre-stimulus baseline of 100 milliseconds. The automatic artifact scanning tool in BESA was applied with the artifact rejection criterion set at plus or minus 50µV (Ille & Berg, 2002). If the absolute difference between two adjacent sample points exceeded 75 µV, the trial would also be excluded from analysis. The averaging of standard stimuli used only the standards presented before deviants (Zhang et al. 2009).

After ERP averaging for each subject, their MMN responses were extracted by subtracting the averaged standard response from the averaged deviant response. To improve signal to noise ratio of the data, three electrodes sites were defined (Figure 2). The left site included channels F3, FC3, and C3, the mid site included channels (Fz, FCz, Cz), and the right site included channels (F4, FC4, and C4). This channel grouping selection was based on visual inspection of the scalp topography of the individual MMN data. Similar channel grouping was used in previous studies (Zhang et al., 2011). Mean MMN amplitude for each electrode site (left, mid and right) was calculated by averaging the samples in a window of 40 ms around the MMN peak. Statistical analysis only included participants with at least 80 accepted deviant trials in each condition (Zhang et al., 2009). MMN peak latency and amplitude for within- and across-category pairs before and after training were compared using a repeated measure ANOVA. Further post-hoc tests were also conducted for a better understanding of interaction effects.

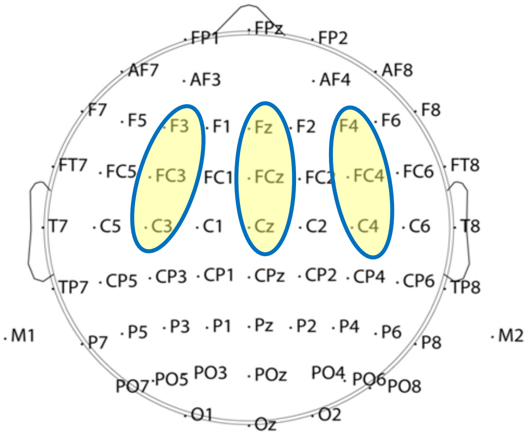


Figure 2: Electrode Grouping. To improve the signal-to-noise ratio, electrode channels were grouped as follows in statistical analysis: Left (F3, FC3, C3), Mid (Fz, FCz, Cz), and Right (F4, FC4, C4) (Zhang et al., 2011).

Chapter 3: Results

3.1 Behavioral results

While we would expect reaction time to decrease with improved behavioral perception, there were no significant pre-test versus post-test changes in reaction time for any of the behavioral tasks due to large inter-subject variability. Some subjects showed reductions in reaction time whereas others showed increases.

As we had hoped, the identification data of the synthetic speech continuum showed significant training effects (Figure 3). The overall trend suggested that participants exhibited a shift toward more native-like categorical perception. Data from two exceptional participants were excluded from analysis because their pre-test identification functions exhibited near-perfect categorical perception (see Appendix E); thus there was little change from pre- to posttest. For the seven subjects who were kept in statistical analysis, paired t-test results revealed significant training-induced changes in the identification slope (t(6) = 3.64, p = 0.0108). A non-parametric test, the Wilcoxon signed-rank test, revealed a significant reduction of phonetic boundary width (p = 0.0156).

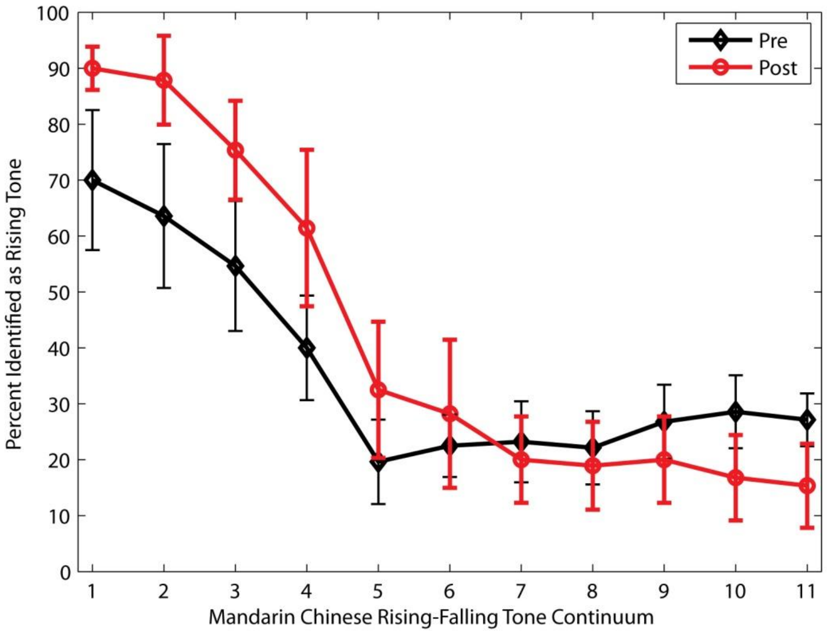


Figure 3. Identification function of the synthetic continuum. This figure depicts the group identification function for the speech tone continuum both pre- (black diamonds) and post- (red circles) training.

The discrimination data for the synthetic speech stimuli showed corresponding training-induced changes (Figure 4). As with the identification function, data from two exceptional participants were excluded from analysis because they exhibited near-perfect discrimination in their pre-tests (see Appendix E). As could be expected with a shift toward more native-like categorical perception, paired t-test results for the pre-post data revealed significant improvement in across-category discrimination (Pair 3-7) (t(6) =

3.87, p = 0.008). There was no significant change for within-category discrimination (Pair 7-11).

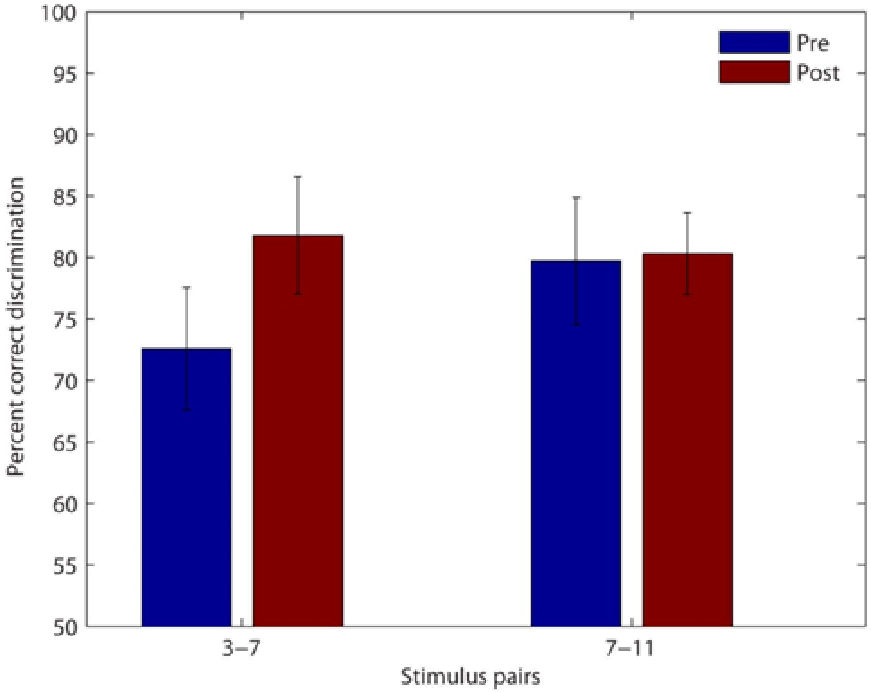


Figure 4: Within- versus across-category discrimination for speech stimuli. This figure depicts percent correct discrimination between across- and within-category pairs (3-7 and 7-11 respectively) from the speech continuum.

On the other hand, the discrimination scores for the non-speech control stimuli (Figure 5) showed a high level accuracy (above 95% on average) in both pre- and post-tests. As expected, there were no significant pre-post changes or significant differences for the across- vs. within-category comparison.

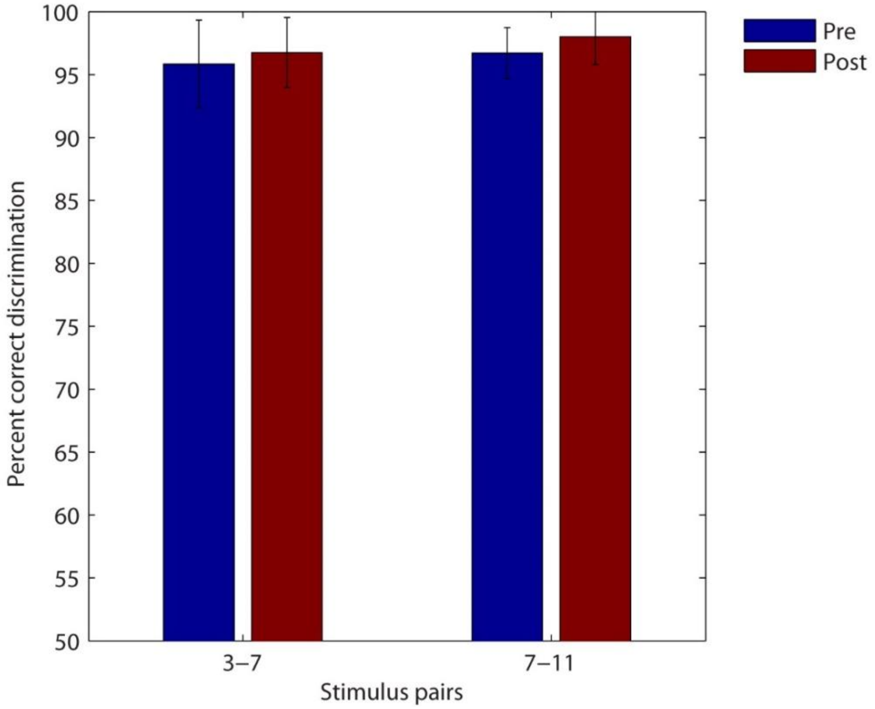


Figure 5: Within- versus across-category discrimination for nonspeech stimuli. This figure depicts percent correct discrimination between across- and within-category pairs (3-7 and 7-11 respectively) from the non-speech continuum.

To analyze training-induced changes in identification of each individual tone (see Figure 6), data from all nine participants were analyzed, as the two exceptional participants for the synthetic continuum tasks were not statistical outliers in this task. Statistical outliers were defined as data falling outside +/- 3 z scores. Repeated measures ANOVA results reveal a significant main effect of training (F(1,8) = 16.60, p = 0.004), and an interaction effect of training and tone category (F(3,24) = 3.76, p = 0.024). Excluding the data from the two participants who showed native-like categorical perception did not change the statistical significance of the main training effect and the interaction effect.

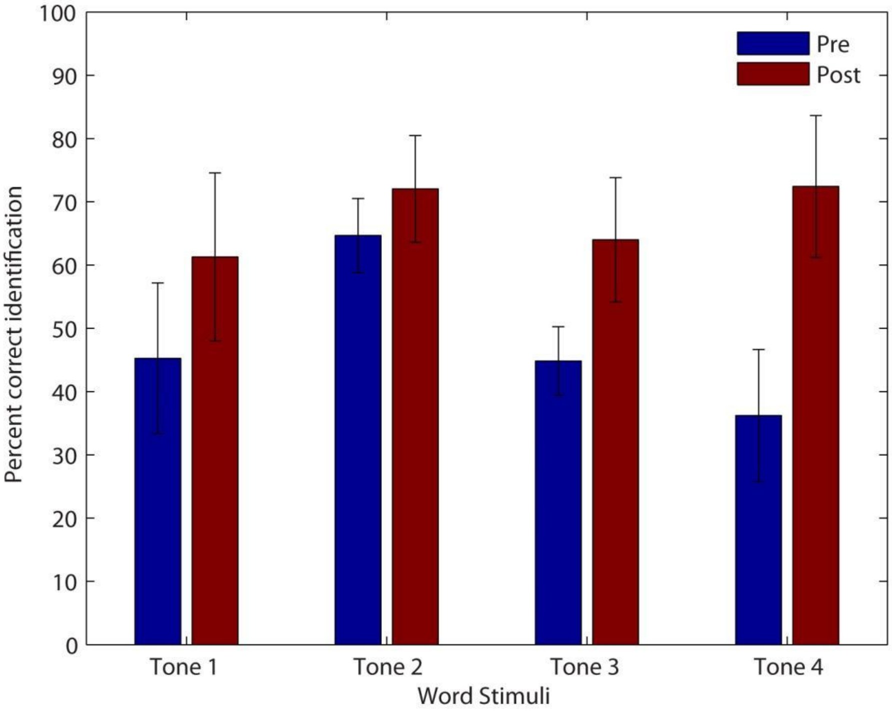


Figure 6. Identification using natural words**.** This figure represents percent correct tone identification in natural words. Stimuli include the exact stimuli used in training, as well as words used in training, but spoken by novel talkers.

One-tailed paired t-tests revealed significant differences in pre- and post-scores for Tones 1, 3, and 4 (respectively, t = 2.19, p = 0.03; t = 2.53, p = 0.02; t = 4.08, p = 0.002). Improvement in identification of Tone 2 was not significant, but demonstrated a clear trend (t = 1.66, p = 0.07).

Table 1

Percent correct tone identification in natural words

Tone	Percent Correct Pre	Percent Correct Post	Training Gains
1	45.26	61.3	16.04*
2	64.66	72.03	7.37
3	44.83	63.98	19.15*
4	36.21	72.41	36.2*

* = *significant change*

Table 1 describes the percent correct identification in natural words for stimuli used in training, and words used in training, but spoken by novel talkers.

Regarding generalization of training effects (Figure 7), data from all nine participants were once again analyzed, as the two exceptional participants were not statistical outliers in this task. The main effect of training was significant (F(1,8) = 8.80, p = 0.018), and the interaction between training and tone categories was also significant (F(3,24) = 3.24, p = 0.040). One-tailed paired t-tests revealed significant differences in pre- and post-scores only for Tone 4 (t = 3.65, p = 0.003). No significant differences occurred for any of the other tones (Tone 1: t = 1.64, p = 0.11; Tone 2: t = 1.58, p = 0.14; Tone 3: t = 1.87, p = 0.090).

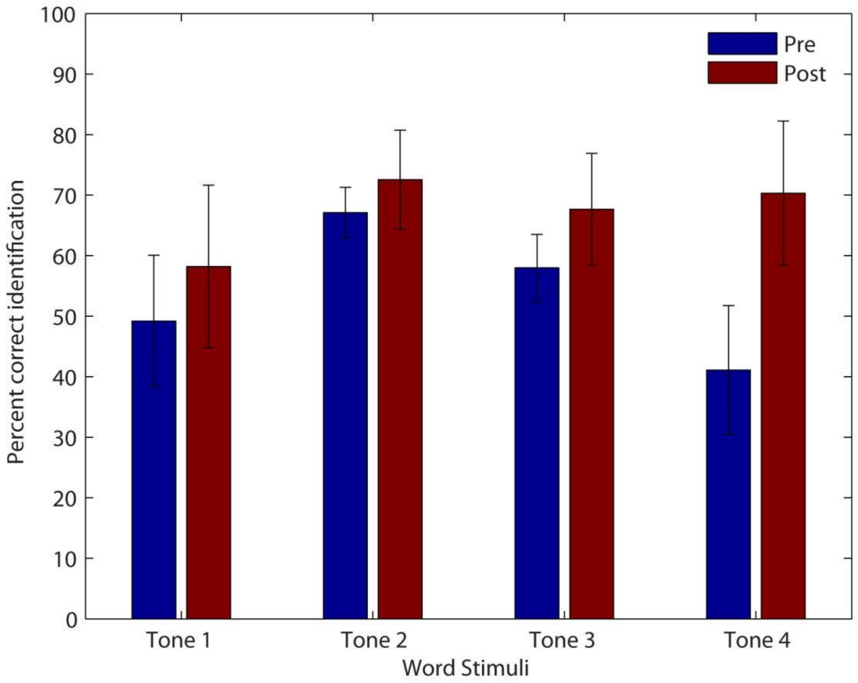


Figure 7. Identification generalization to novel stimuli. This figure demonstrates generalization of training effects to novel natural word stimuli that were not used in training sessions.

Table 2

Percent Correct Tone Identification in Untrained Natural Words

Tone	Percent Correct Pre	Percent Correct Post	Training Gains
1	49.18	58.21	9.03
2	67.11	72.57	5.46
3	57.97	67.66	9.69
4	41.1	70.32	29.22*

* = *significant change*

Table 2 describes the percent correct identification in natural words for novel stimuli.

3.2 MMN Results

We endeavored to discover whether increased perceptual skills lead to increased MMN amplitudes and decreased latencies for the speech stimuli, and no or comparable changes for the non-speech stimuli. Furthermore, we sought to determine whether training would result in a shift towards more native-like categorical perception, meaning higher amplitudes and shorter latencies for across-category stimulus pairs compared to within-category stimulus pairs. To this end, MMN amplitudes and latencies for both speech and non-speech stimuli were analyzed for each electrode site both before and after training for both within- and across-category stimuli pairs. The grand mean waveforms are depicted in Figure 8. Analysis of MMN amplitude and latency for speech and non-speech stimuli both before and after training for both within- and across-category stimuli pairs are described in the figures and tables to follow.

Table 3

Speech mean amplitudes and latencies

	Left		Mid		Right	
	Across	Within	Across	Within	Across	Within
Pre-Amplitude	-1.34	-0.81	-1.59	-0.79	-1.45	-0.92
Post-Amplitude	-2.15	-1.09	-2.63	-1.18	-2.08	-1.14
Pre-Latency	184.86	174	203.57	149.57	213	160.86
Post-Latency	164	170.43	171.43	178.71	170.86	165

Table 3 lists the mean amplitudes and latencies for the MMN responses to the across- and within-category speech stimuli before and after training for each electrode site.

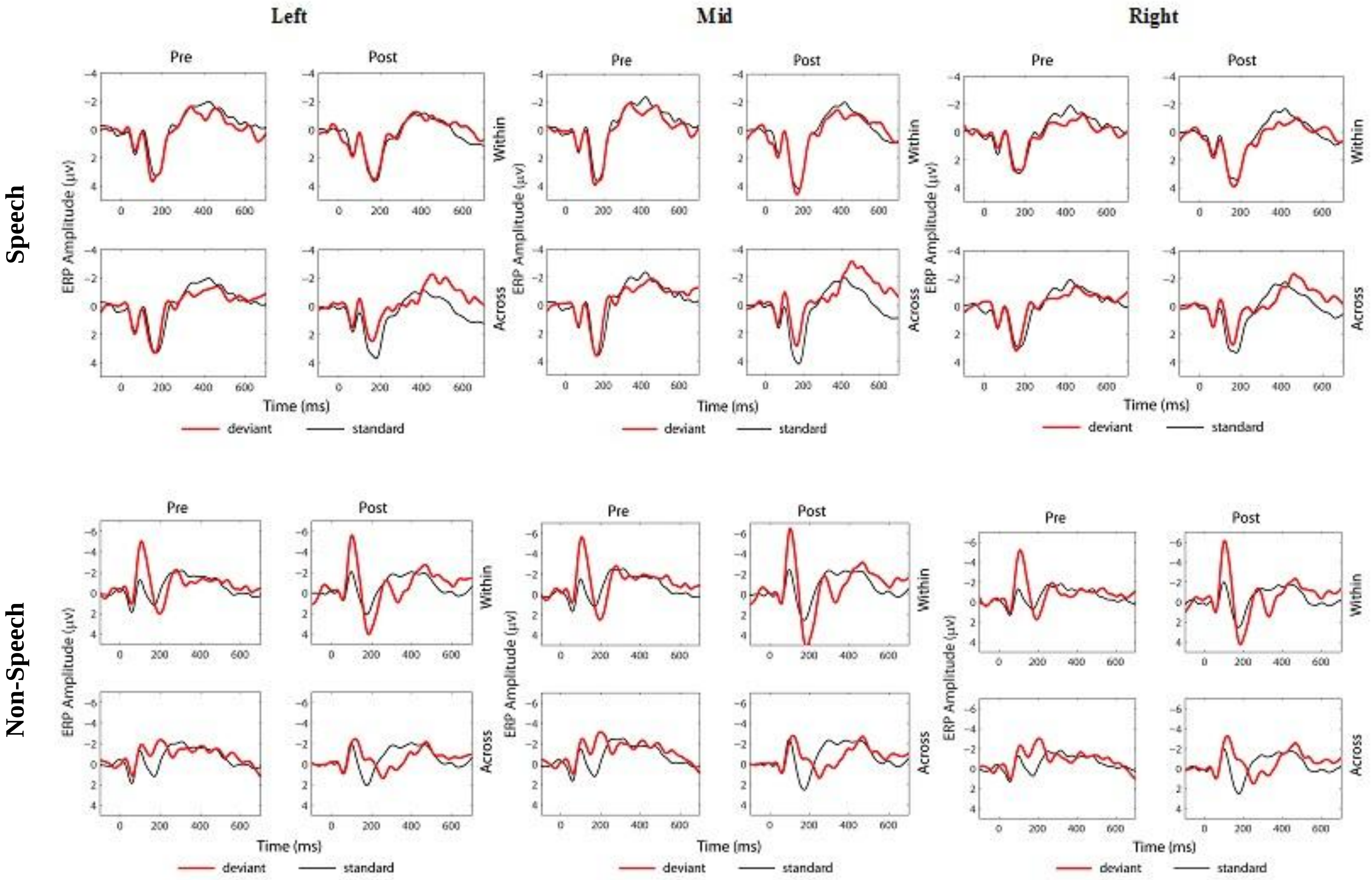


Figure 8. MMN Waveforms. This figure depicts grand-average waveforms from the left, mid, and right electrode sites elicited by across-category deviants, within-category deviants, and standards both pre- and post-training including data from seven participants. The top row presents the waveforms elicited by the speech stimuli; the bottom those elicited by the non-speech stimuli.

For MMN amplitude for the speech stimuli (Figure 9), repeated measures ANOVA revealed significant effects for training (pre- versus post-) (F(1,6) = 6.86, p = 0.040) and stimulus condition (within- versus cross-category) (F(1,6) = 7.07, p = 0.035). A marginally significant interaction effect was present for stimulus condition and electrode site (F(2,5) = 5.57, p = .053). There was no significant main effect for electrode site (F(2,12) = 3.37, p = 0.092). Further analysis using post-hoc one-tailed paired t-tests revealed significant training-induced MMN enhancement at left and mid electrode sites for the across-category condition, but not at the right electrode site (left: t(6) = 2.52, p = 0.045; mid: t(6) = 2.766, p = 0.03; right: t(6) = 1.54, p = 0.17). There were no significant changes in MMN amplitude for the within-category condition for any of the three electrode sites.

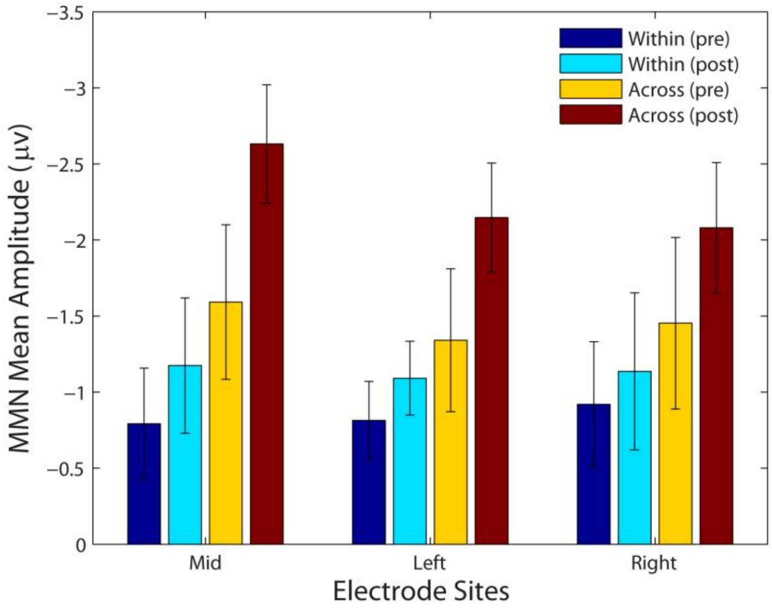


Figure 9. Speech MMN mean amplitude. This figure depicts mean amplitudes from the MMN responses elicited by the speech stimuli pre- and post-training for within- and across-category deviants using data from seven participants.

When considering MMN latency for the speech stimuli (Figure 10), repeated measures ANOVA revealed a non-significant trend toward an interaction effect for electrode site and stimulus condition (F(2,12) = 3.88, p = 0.084). There were no significant effects for any of the other main factors: training, stimulus condition, electrode site. Further analysis using one-tailed post-hoc paired t-tests revealed significant training-induced decreases of MMN latency at the mid and right electrode for the across-category condition (t(6) = 1.96, p = 0.049; t(6) = 2.72, p = 0.035 respectively). The left electrode site demonstrated a clear trend, but did not reach statistical significance (t(6) = 1.85, p = 0.056). There were no significant changes in MMN latency for the within-category condition for any of the three electrode sites of interest.

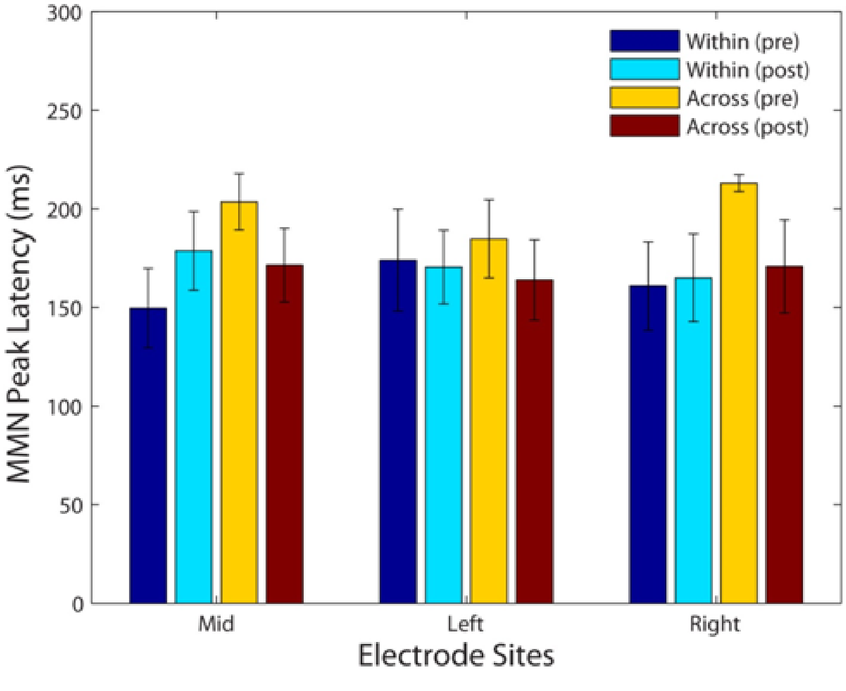


Figure 10. Speech MMN latency. This depicts mean peak latencies (in milliseconds) from the MMN responses elicited by the speech stimuli pre- and post-training for within- and across-category deviants using data from seven participants.

Data from the non-speech stimuli were analyzed as a control. We expected either no change, or a change comparable to those observed with the speech stimuli. The results reveal some interesting training effects, which are discussed further in the discussion section.

Table 4

Non-speech mean amplitudes and latencies

	Left		Mid		Right	
	Across	Within	Across	Within	Across	Within
Pre-Amplitude	-3.15	-3.87	-3.65	-4.17	-3.11	-3.94
Post-Amplitude	-2.86	-3.31	-3.59	-4.09	-3.72	-4.33
Pre-Latency	173.14	124.43	194.86	126.86	198.86	124.43
Post-Latency	173	127.29	163.71	126.29	158.43	123.86

Table 4 lists the mean amplitudes and latencies for the MMN responses to the across- and within-category non-speech stimuli before and after training for each electrode site.

For the MMN amplitude for the non-speech stimuli (Figure 11), repeated measure ANOVA revealed a significant effect for electrode site (F(2,12) = 19.52, p = 0.002), and a significant interaction effect for electrode site and training (F(2,12) = 11.19, p = 0.004). There was no main effect for training (F(1,6) = 6.65, p =0.04), and there were no significant training-induced changes in MMN amplitude at any of the three electrode sites for either stimulus condition.

Further repeated measures ANOVA revealed significantly larger MMN at the right site than at the left (F(1,6) = 59.67, p = 0.0002). There was also a significant interaction effect for hemisphere and stimulus condition (F(1,6) = 6.65, p = 0.04). Further post-hoc t-tests were completed to examine the hemisphere effect for each condition. Before training, MMN activities were bilateral for both across- and within-category conditions. After training, enhanced MMN activity was present in the right relative to the left hemisphere for the within-category condition (t(6) = 3.60, p = 0.01). A similar trend was observed for the across-category condition, but it did not reach significance (t(6) = 1.73, p = 0.067).

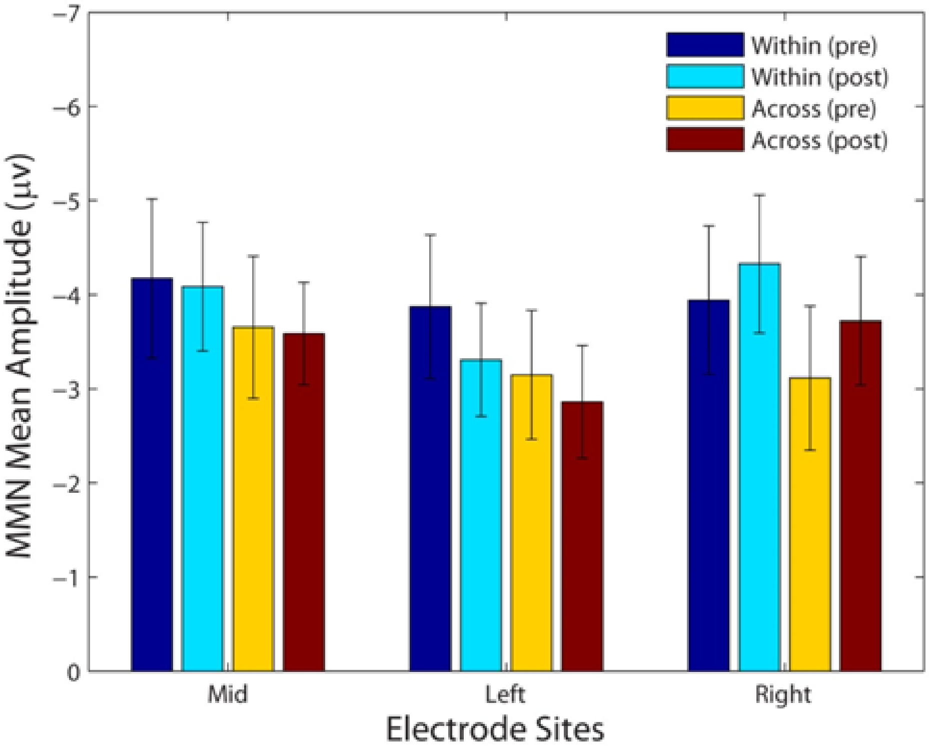


Figure 11. Non-speech MMN amplitude. This figure depicts mean amplitudes from the MMN responses elicited by the non-speech stimuli pre- and post-training for within- and across-category deviants using data from seven participants.

Regarding MMN latency for the non-speech data (Figure 12), repeated measures ANOVA revealed a highly significant main effect for stimulus condition (F(1,6) =

423.39, p = 0.000001) and a significant interaction effect for stimulus condition and training (F(1,6) = 12.26, p =0.013). Further analysis using post-hoc one-tailed t-tests revealed significant training-induced decreases in MMN latency at the mid and right electrode sites for the across-category condition (t(6) = 3.34, p = 0.008) and t(6) = 5.02, p = 0.001 respectively). There were no significant changes in MMN latency for the within-category condition at any of the three electrode sites of interest.

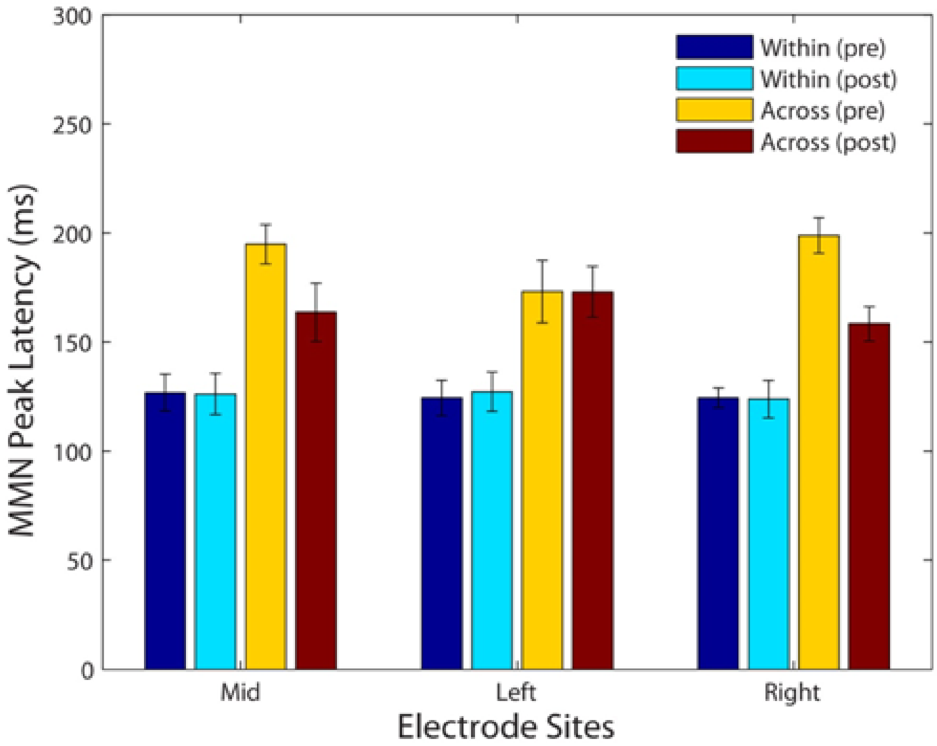
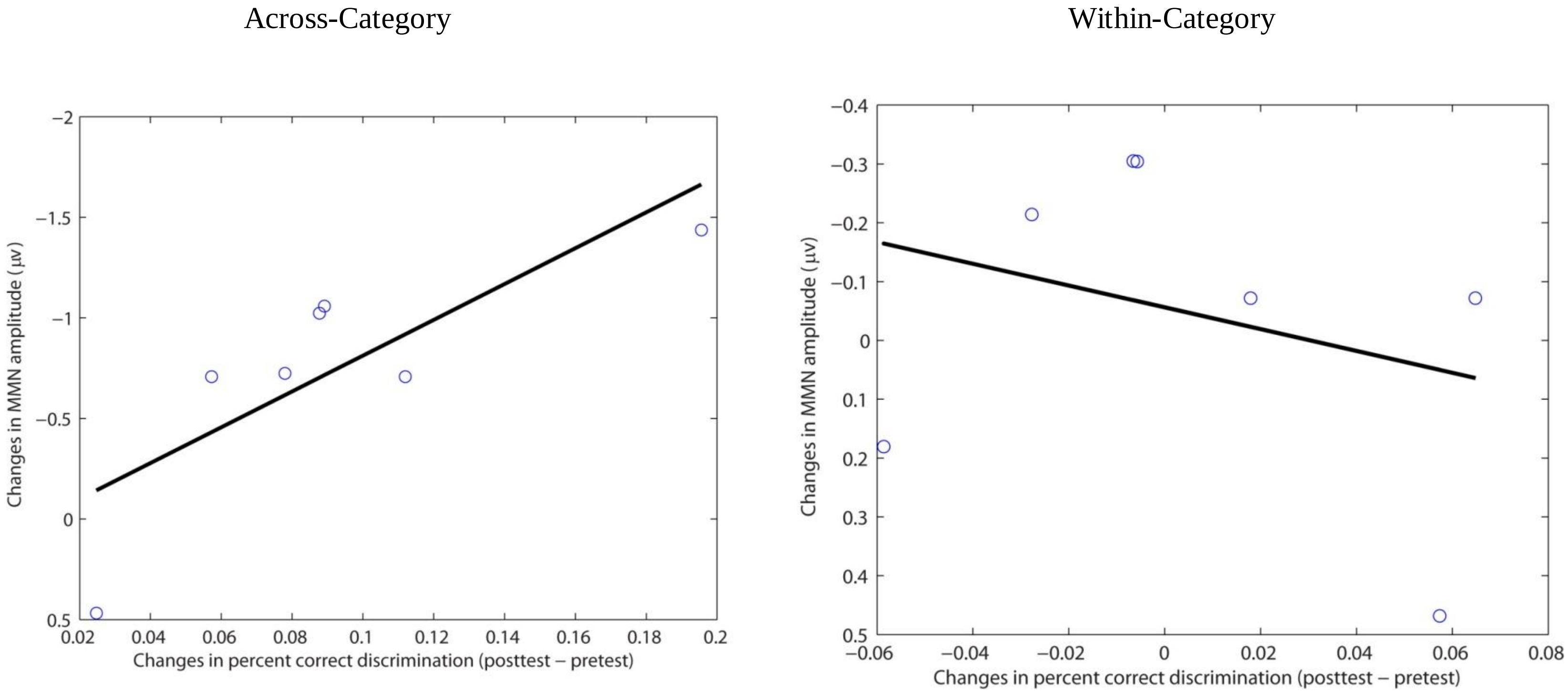


Figure 12. Non-speech MMN latency. This figure depicts mean latencies from the MMN responses elicited by the non-speech stimuli pre- and post-training for within- and across-category deviants using data from seven participants.

3.3 Brain-behavior correlation

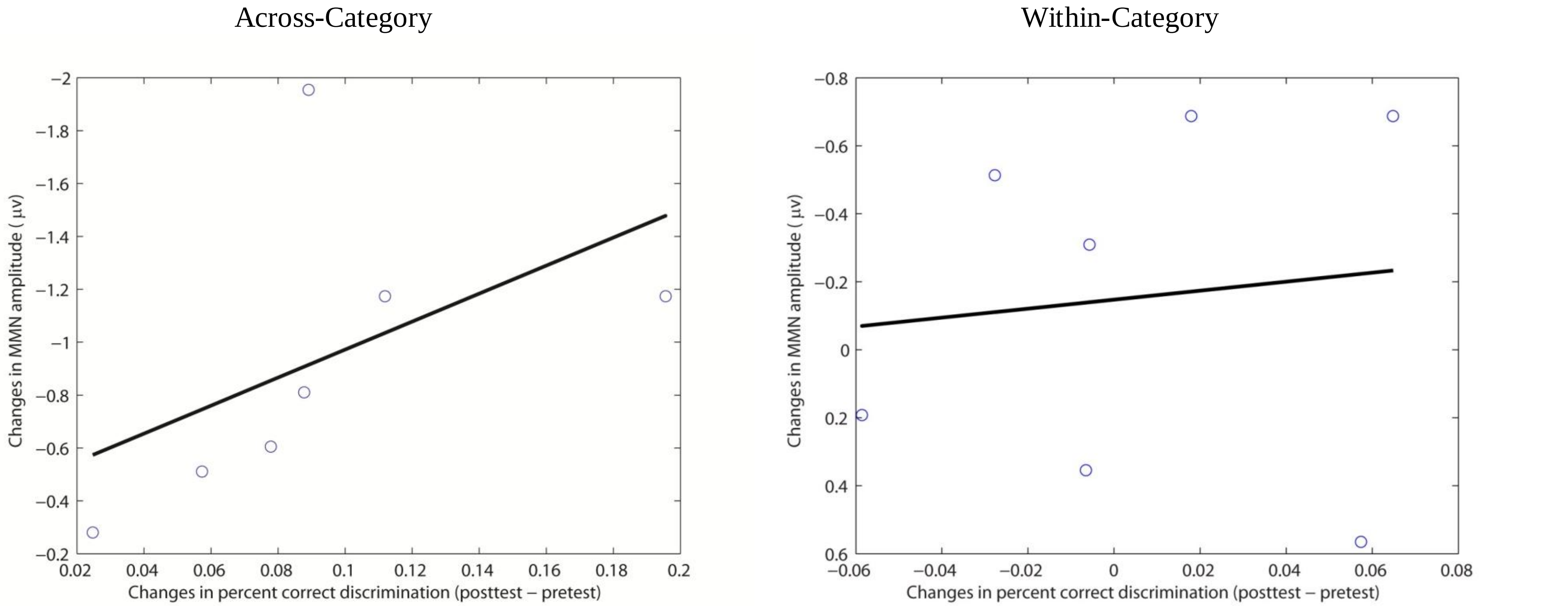
If behavioral changes in perception are reflected in the MMN measures, we would expect to see a strong correlation between the two. For the left site, correlation was marginally significant for the across-category (r= - 0.746, p = 0.054) condition but not for the within-category (r = 0.29, p = 0.53) condition. For the mid site, correlation was not significant for either condition (cross: r = - 0.50, p = 0.25; within: r = -0.11, p = 0.81). Finally, for the right site, correlation was not significant for either condition (cross: r = - 0.68, p = 0.09; within: r = -0.19, p = 0.69). See Figures 13-15 for summary.



Across-Category

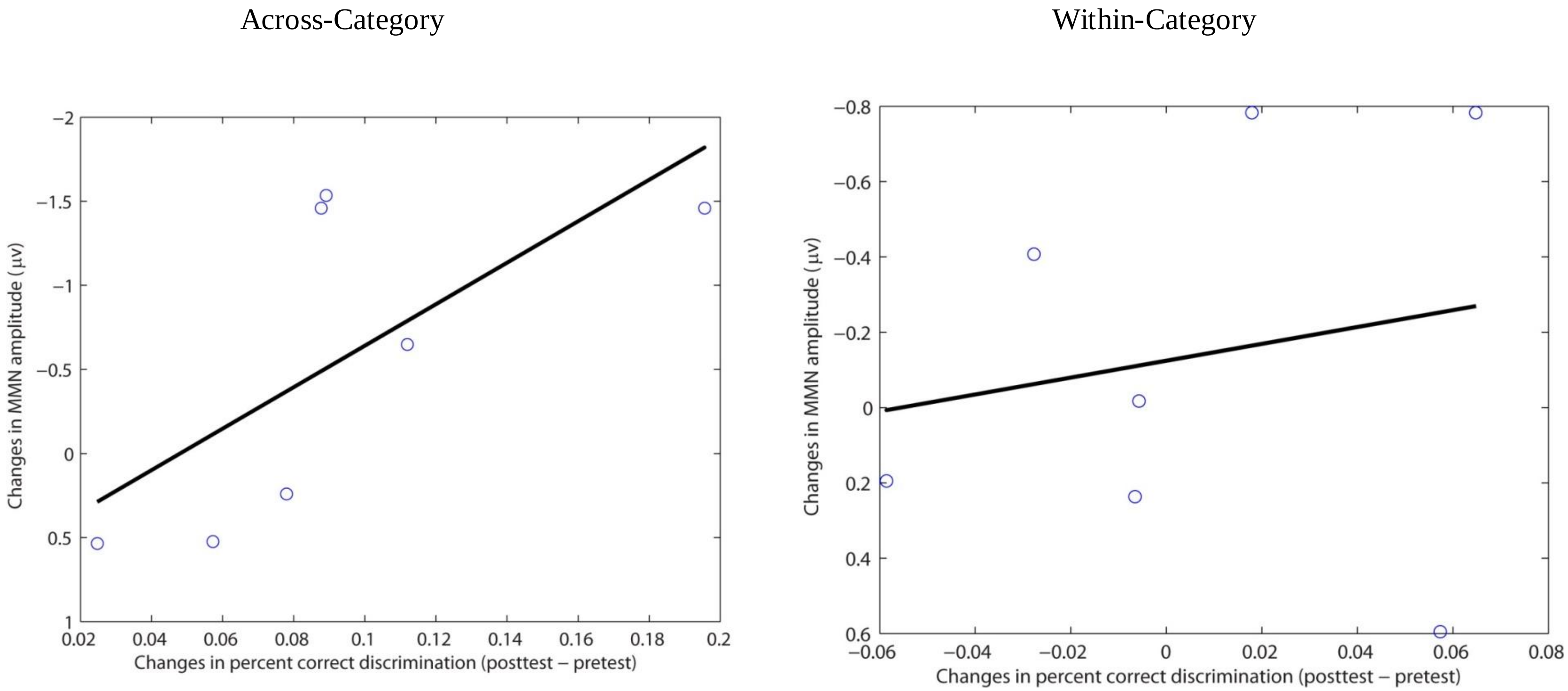
Within-Category

Figure 13. Brain-behavior correlation: Left electrode site. This figure depicts the correlation between the percent correct discrimination for across-category and within-category speech stimulus pairs (x-axis) and the changes in MMN amplitude (y-axis) for the left electrode site. Correlation was significant for both across-category (r= - 0.746, p = 0.054) and within-category (r = 0.29, p = 0.53) conditions.


Across-Category

Within-Category

Figure 14. Brain-behavior correlation: Mid electrode site**.** This figure depicts the correlation between the percent correct discrimination for across-category and within-category speech stimulus pairs (x-axis) and the changes in MMN amplitude (y-axis) for the mid electrode site. Correlation was not significant for either condition (Across: r = - 0.50, p = 0.25; Within: r = -0.11, p = 0.81).


Across-Category

Within-Category

Figure 15. Brain-behavior correlation: Right electrode site. This figure depicts the correlation between the percent correct discrimination for across-category and within-category speech stimulus pairs (x-axis) and the changes in MMN amplitude (y-axis) for the right electrode site. Correlation was not significant for either condition (Across: r = - 0.68, p = 0.09; Within: r = -0.19, p = 0.69).

Chapter 4: Discussion

4.1 Behavioral results

4.1.1 Does training improve behavioral identification?

Unlike previous studies which had shown a decrease in reaction time associated with improved perception (Zhang et al., 2009), participants in the current study exhibited no significant changes in reaction time. Despite the lack of change in reaction time, the overall results suggest that the training was successful in improving behavioral identification for trained words. Participants' perceptual accuracy differed for each tone, and training was more effective for some tones than for others. Tones 1 and 3 were of comparable difficulty pre-training (45.26% and 44.83% respectively), and improved by similar amounts (16.04% and 19.15% respectively). Tone 2 proved to be the easiest to perceive initially (64%), but did not demonstrate significant improvement (7.37%). Tone 4 proved to be the most difficult to identify pre-training (36.21%), but it also exhibited the greatest improvement, an impressive 36.2%.

Previous research has postulated that Tones 1 and 4 are the easiest to for both native and non-native speakers to acquire, and that the latter is the easiest to perceive initially (Howell, Jiang, Peng, & Lu, 2012; Wan, 2007). Because Tone 3 has the most allophones, some researchers consider it the most difficult for non-native speakers to perceive (J. Lee et al., 2007). The current study, as well as the other training studies discussed below, found differing results regarding perception of individual tones by non-native speakers.

Like the current study, Wang (2013) and Wang et al. (1999) demonstrated perceptual improvements as a result of training. However, results regarding individual tones were different. Wang (2013) found Tone 3 to be the easiest to perceive in the pre-training measures, and only perception of Tones 1 and 2 significantly improved with training. Wang et al. (1999), however, observed no significant differences in perception of the four tones before training. They found significant improvement in perception of all four tones after training, with greatest improvement in perception of Tone 4, followed by Tones 3, 2, and 1 respectively. These differences could have resulted from distinct testing methodologies. While Wang et al. (1999) utilized a pair-wise identification task, the current study utilized a more realistic identification task in which participants heard a natural word and were required to select which of the four tones they heard.

For the current study, as well as those described above, participants' differing perceptual accuracy before training and differential improvements for each of the four tones are not well-explained by either the SLM or PAM model (Best & McRoberts, 2003; Flege, 1987, 1995, 1997). Following the PAM model, we would expect that the rising and falling tones would be easiest to perceive initially because rising and falling intonation patterns are used contrastively in English to distinguish between declarative and interrogative sentences. Though the rising tone was the easiest to perceive before training in the current study, the falling tone proved to be the most difficult. Following the SLM model, we would expect to observe the most improvement in tones that are the most different from English. However, we actually observed the most improvement in perception of Tone 4, which is similar to the falling intonation pattern in English. At least for the current study, neither model can be used to specifically address lexical tone learning based upon acoustic or phonological similarities between L1 and L2, especially given that L1 is a non-tone language, and L2 is a tone language.

An important consideration for assessing the success of any given training method is whether the training generalizes to untrained items. While overall training effects generalized to untrained stimuli in the current study, most of the generalization occurred for Tone 4. This is in contrast to Wang et al. (1999), who found training effects generalized to untrained stimuli for all four tones. Differences between the current study's results and those of other Mandarin tone training studies could be the result of the individual characteristics of the participants in each study, the specific training method and duration of training sessions, or the small sample size of the current study. Further research is necessary to determine which tones are most difficult for L2 learners to acquire, which training method is most successful for individual learner profiles, and which testing methodology most accurately captures perceptual improvement.

4.1.2 Does training engender more native-like categorical perception?

Results demonstrate significantly steeper slope and narrower phonemic boundary width in the identification function utilizing the synthetic speech continuum. This indicates that training increased categorical perception in non-native listeners, resulting in more native-like identification functions. This is encouraging because Zhang et al. (2009) did not produce this result in training adult Japanese speakers to learn the English /r-l/ contrast. Furthermore, across-category discrimination improved, while within-category discrimination did not, suggesting that participants were increasing their sensitivity to the salient tone differences. While within-category discrimination did not decrease, the fact that it remained stable while across-category discrimination improved suggests that participants were in the process of forming of new speech sound categories to accommodate previously irrelevant non-native speech sound contrasts.

4.2 MMN Results

4.2.1 Do brain measures reflect behavioral changes in perception?

With improvements in behavioral accuracy, we expected to see increased MMN amplitude and decreased MMN latency for across-category speech deviants, and no change for the within-category speech deviants (Kraus et al., 1995; Tamminen et al., 2013). For the non-speech control stimuli, we expected either no change or a change reflective of the changes in responses to the speech stimuli (Xi et al., 2010). Overall, the results support these expectations.

For the speech stimuli, MMN amplitude increased for the left and mid locations for across-category deviants, but not for within-category deviants. Furthermore, MMN latency decreased for the mid and right locations – with the same trend for the left location -- for the across-category deviants, but not for the within-category deviants. Responses to non-speech stimuli are generally in agreement with expectations as well. There were no significant main effects for MMN amplitude for either stimulus condition. Changes in MMN latency for the non-speech stimuli reflect the changes observed for the speech stimuli in that latencies reduced for the across-category deviants at the mid and right locations, and there were no changes in latency for the within-category stimulus condition.

As expected for non-native listeners, the speech stimuli were processed bilaterally (Y. Wang et al., 2001). A deeper look into the non-speech data, however, reveals some intriguing results. Before training, MMN responses were largely bilateral for both stimulus conditions. After training, however, MMN amplitudes were larger in the right hemisphere than the left hemisphere for the within-category stimulus condition, with the same trend for the across-category condition. Furthermore, MMN latencies decreased for the mid and right locations for the across-category condition.

These results suggest a training-induced reallocation of neural resources in the processing of non-speech pitch stimuli from the left to the right hemisphere. Similar examples of brain plasticity were found by Wang et al. (2011) and Zhang et al. (2009), who discovered reallocation of resources to more efficiently process L2 stimuli. While their discoveries were regarding speech stimuli, the reallocation of resources observed in the current study are not unreasonable. Xi et al. (2010) suggest that acoustic and phonological information from lexical tones are processed in parallel, so increased linguistic experience with lexical tones provided by the training could have influenced the processing of the non-speech stimuli (Bent, Bradlow, & Wright, 2006; Xi et al., 2010). Future studies using technology with high spatial resolution, such as Functional Magnetic Resonance Imaging (fMRI), may be better able to capture this phenomenon.

4.3 Acoustic versus linguistic processing

A closer look at the speech discrimination data before training shows that participants were behaviorally more accurate for within-category discrimination than for across-category discrimination, while MMN amplitude was greater for across-category deviants than within-category deviants. These conflicting results could be explained in terms of acoustic versus linguistic processing. Acoustic processing focuses exclusively on acoustic characteristics of the stimuli. Linguistic processing, on the other hand, includes linguistic information present in the stimuli.

For the speech discrimination task, the absolute acoustic difference between each stimulus in each pair of stimuli was the same, but pairs differed in the way that acoustic difference was realized. Both stimuli in the across-category pair were relatively flat acoustically (see Figures 1 and 16 for illustration), while one of the within-category stimuli demonstrated a more salient acoustic change in frequency slope than all the others. This would make within-category stimuli easier to discriminate using acoustic processing. Because the across-category stimuli fell into different linguistic tone categories, they would be easier to discriminate using linguistic processing for native Mandarin speakers (Xi et al., 2010).

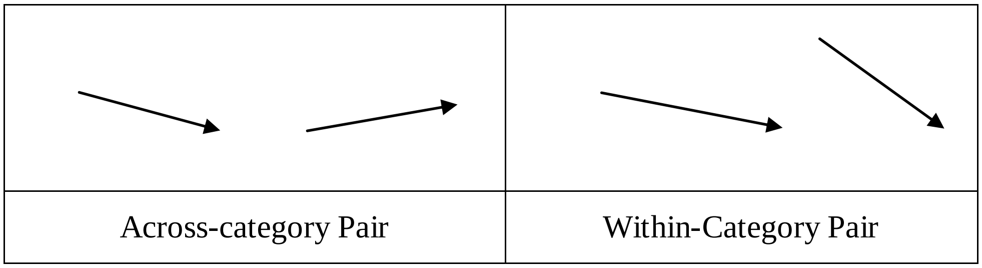


Figure 16. Schematic illustration of the across-category and within-category tone pairs designed to aid comprehension of the above paragraph.

Research has demonstrated that auditory processing of tone information is accessed at a similar point in time as information provided by vowels and consonants, and it contributes to word processing in the same manner (Lidji et al., 2010; Schirmer et al., 2005). Indeed, perception of lexical tones includes not only pitch perception, but also mapping all relevant acoustic information only linguistically relevant phonetic categories (C.-Y. Lee & Lee, 2010). Perception of lexical tones begins before the listener even attends to the stimuli. In fact, in native speakers, "pre-attentive encoding of abstract auditory rules of lexical tones can predict perception during a later attentive stage" (X.-D. Wang et al., 2012, p. 3). Because of this integration of linguistic and acoustic information, lexical tone stimuli would automatically be treated linguistically.

For the behavioral discrimination task, participants may have been consciously focusing their efforts on acoustic processing, resulting in better discrimination for within-category pairs. Because the MMN response is pre-attentive, however, participants would not have been able to focus their efforts on acoustic processing. This would explain the conflicting behavioral and MMN results from before training. Furthermore, training resulted in improved behavioral discrimination of across-category pairs, but not of within-category pairs. This suggests that participants switched to a more linguistic mode of processing.

Further information regarding acoustic versus linguistic processing can be obtained by examining portions of the ERP waveform that occur later than the MMN. In the current study, there is still a large mismatch between standard and deviant responses beyond 400 milliseconds after stimulus presentation, a phenomenon commonly referred to as sustained negativity. Sustained negativity is often observed in tasks requiring linguistic processing, such as detection of lexical-semantic or syntactic violations (Friederici, Pfeifer, & Hahne, 1993; Jiang, Tan, & Zhou, 2009; Li, Shu, Liu, & Li, 2006; Ye, Luo, Friederici, & Zhou, 2006). We suspect the presence of sustained negativity in the current study is the result of a training-induced shift from acoustic processing to linguistic processing of the lexical tone stimuli. Given the limitations of small sample size in the current study, further research on this topic is needed to confirm our suspicions.

In addition to MMN and sustained negativity, some researchers also analyze the P3a, a response in the ERP waveform occurring after the MMN that indicates an involuntary shift in attention toward the stimulus (Horváth, Winkler, & Bendixen, 2008; Polich, 2007). Though statistical analysis of the P3a response is beyond the scope of this study, some interesting observations were made. There appears to be an overall larger

P3a response after training than before in the grand average ERP waveforms for the speech stimuli. This would imply that the differences had become salient enough to the listeners to warrant a redirection of conscious attention.

4.4 IDS-based training program

Similar to Zhang et al. (2009), results from this study provide strong support for the use of IDS characteristics in perceptual training programs. Improvements in identification of tones in natural words testify to functionality of training gains. Not only did participants improve their identification skills for trained items, but training effects generalized to novel stimuli for Tone 4. These positive results suggest that this sort of multimodal adaptive training method is effective for L2 speech-sound contrast learning in adulthood. Not only was the training effective, it was also efficient. Participants were able to improve their perception of Mandarin tones after only 2-3 hours of training, compared to the 12 hours of training used by Zhang et al. (2009).

Furthermore, this training program was relatively user-friendly in that participants were not required to make any overt judgments regarding tones they heard. Participants self-selected the tones they heard, allowing them to complete the training at their own pace and in their own preferred manner. This is in contrast to other training studies that required participants to identify which tone had been presented (Francis et al., 2008; Kaan, Barkley, Bao, & Wayland, 2008; Kaan et al., 2007; J. Lee et al., 2007; X. Wang, 2013; Y. Wang, Sereno, et al., 2003; Y. Wang et al., 1999; Wayland & Guion, 2004; Wayland et al., 2010; Wong, Perrachione, et al., 2007).

4.5 Future applications

Though the current study was limited by the brevity of training and the small sample size, the results add support to the previous success of Zhang et al. (2009) for using computer programs incorporating the characteristics of IDS to train non-native speech sound contrasts. However, further research is required to determine exactly which characteristics are most conducive to L2 learning. Future studies using a larger sample size and longer training would allow for making stronger conclusions.

The differences between results from the current study and previous Mandarin tone training studies provide some important considerations for future research. Predictions about which tones will prove most difficult initially and which are most amenable to training should be made with caution. Different studies have produced differing results regarding perceptual accuracy when identifying individual tones. A consensus has not yet and may never be reached. As with training in any area, training speech sound contrasts must consider individual differences among trainees in terms of baseline perceptual abilities, motivation for training, and success with a given training method. In the current study, for instance, we found two exceptional listeners whose categorical perception of the Mandarin lexical tone continuum was native-like and yet their performance in identifying the four tones in natural word stimuli were not native-like. Further research is necessary to determine which tones are most difficult for L2 learners to acquire, which training method is most successful for individual learner profiles, and which testing methodology most accurately captures perceptual improvement.

Results from the current study challenge two widely-held assumptions regarding lexical tone learning. First is the assumption that listeners will exhibit poor discrimination and identification of non-native speech sound contrasts. As the two exceptional participants in this study demonstrate, some individuals are equipped with remarkable perceptual capabilities even before training. Though the percentage of Americans born with absolute pitch is low relative to other cultures (Dediu & Ladd, 2007; C.-Y. Lee & Lee, 2010), that does not guarantee that participants who have not had any significant musical training will not have absolute pitch. The second assumption is the common practice of reporting MMN responses from only the mid electrode site. As results from this study demonstrate, hemispheric differences may be present. Future research with a larger sample size may illuminate training effects in each hemisphere.

Though outside the scope of the current study, observations regarding the P3a and sustained negativity provide impetus for further research. While the current study's training focused exclusively on acoustic exaggeration, future studies could make more ecologically valid programs that incorporate more semantically-driven contexts. For example, pictures of the word meanings could accompany the audible presentation of the word. Other studies utilizing semantically-driven training have demonstrated success

(Chandrasekaran et al., 2010; J. Lee et al., 2007; Wong, Perrachione, et al., 2007; Wong & Perrachione, 2007; Wong, Skoe, et al., 2007), but they did not incorporate IDS characteristics and acoustic exaggeration as the current study did. Coupled with statistical analysis of the P2, P3a, and sustained negativity, this methodology would allow for more rigorous exploration of linguistic versus acoustic processing in Mandarin tone learning by non-native speakers.

The success of this training procedure is a testament to remaining brain plasticity in adults. While this method has been effective in training L2 learners, whether similar methods could be used in clinical settings for individuals with speech-sound disorders, dyslexia, auditory processing deficits, or other speech and language disorders remains to be seen.

Works Cited

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The Influence of Linguistic Experience on the Cognitive Processing of Pitch in Speech and Nonspeech Sounds. *Journal of Experimental Psychology. Human Perception & Performance*, *32*(1), 97–103. doi:10.1037/0096-1523.32.1.97

Best, C., & McRoberts, G. W. (2003). Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language & Speech*, *46*(2/3), 183–216.

Bialystok, E., & Hakuta, K. (1999). Confounded Age: Linguistic and Cognitive Factors in Age Differences for Second Language Acquisition. In *Second Language Acquisition and the Critical Period Hypothesis* (pp. 161–181). Psychology Press.

Bidelman, G., Gandour, J., & Krishnan, A. (2011). Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–434.

Birdsong, D. (1999). *Second Language Acquisition and the Critical Period Hypothesis*. Psychology Press.

Burnham, D., Lauw, C., Lau, S., & Stokes, S. (2000). Perception of Visual Information for Cantonese Tones (pp. 86–91). Presented at the Austrailian International Conference onSpeech Science and Technology, Canberra.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage*, *19*(1), 113–124. doi:10.1016/S1053-8119(03)00020-X

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, *25*(3), 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*(1), 1–9.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*(1), 456–465.

Chao, Y. R. (1947). *Cantonese primer*. Published for the Harvard-Yenching Institute [by] Harvard University Press.

Chen, T. H., & Massaro, D. W. (2008). Seeing pitch: Visual information for lexical tones of Mandarin-Chinese. *The Journal of the Acoustical Society of America*, *123*(4), 2356. doi:10.1121/1.2839004

Cheng, B., & Zhang, Y. (2013). Neural plasticity in phonetic training of /i-I/ contrast for adult Chinese speakers. *Journal of the Acoustical Society of America*, *134*, 4245.

Chin, T. (2006). *Sound systems of Mandarin Chinese and English: a comparison*. Lincom.

Cornella, M., Leung, S., Grimm, S., Escera, C., & Mansvelder, H. D. (2012). Detection of Simple and Pattern Regularity Violations Occurs at Different Levels of the Auditory Hierarchy. *PLoS ONE, 7*(8), 1–8. doi:10.1371/journal.pone.0043604

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, *104*(26), 10944–10949. doi:10.1073/pnas.0610848104

Diehl, R., Lotto, A., & Holt, L. (2004). Diehl 2004 SPercep.pdf, *55*, 149–179. doi:10.1146/annurev.psych.55.090902.142028

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge, UK: Cambridge University Press.

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47–65.

Flege, J. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange, *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. (1997). English vowel production by Dutch talkers: More evidence for the "similar" vs. "new" distinction. In A. James & J. Leather, *Second Language Speech: Structure and Process.* (pp. 11–52). Berlin/New York: Mouton de Gruyter.

Flege, J. (1999). Age of Learning and Second Langauge Speech. In *Second Language Acquisition and the Critical Period Hypothesis* (pp. 101–131). Psychology Press.

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

Foster, S. M., Kisley, M. A., Davis, H. P., Diede, N. T., Campbell, A. M., & Davalos, D. B. (2013). Cognitive function predicts neural activity associated with pre-attentive temporal processing. *Neuropsychologia*, *51*(2), 211–219. doi:10.1016/j.neuropsychologia.2012.09.017

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268–294.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192. doi:10.1016/0926-6410(93)90026-2

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, *60*, 5–24.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. doi:10.1016/S0095-4470(03)00016-0

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. doi:10.1016/j.wocn.2011.11.001

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, *79*(2), 139–147. doi:10.1016/j.biopsycho.2008.04.001

Howell, P., Jiang, J., Peng, D., & Lu, C. (2012). Neural control of fundamental frequency rise and fall in Mandarin tones. *Brain and Language*, *121*(1), 35–46. doi:10.1016/j.bandl.2012.01.004

Ille, N., & Berg, P. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*, 113–124.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57.

Jiang, X., Tan, Y., & Zhou, X. (2009). Processing the universal quantifier during sentence comprehension: ERP evidence. *Neuropsychologia*, *47*(8–9), 1799–1815. doi:10.1016/j.neuropsychologia.2009.02.020

Johnson, K. (2005). Speaker normalization in speech perception. In *The handbook of speech perception* (pp. 363–389). Malden, MA: Blackwell Pub.

Kaan, E., Barkley, C., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neuroscience*, *9*, 53–69.

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research*, *1148*, 113–122. doi:10.1016/j.brainres.2007.02.019

Kaan, E., Wayland, R., & Keil, A. (2013). Changes in Oscillatory Brain Networks after Lexical Tone Training. *Brain Sciences*, *3*(2), 757–780. doi:10.3390/brainsci3020757

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., & Trent, N. (1995). Central Auditory System Plasticity Associated with Speech Discrimination Training. *Journal of Cognitive Neuroscience*, *7*(1), 25–32.

Krishnan, A., Xu, Y., Gandour, J., & Carianib, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cortex*, *30*, 37.

Kuhl, P. K. (2007). Is speech learning “gated” by the social brain? *Developmental Science*, *10*(1), 110–120. doi:10.1111/j.1467-7687.2007.00572.x

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727. doi:10.1016/j.neuron.2010.08.038

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of* *the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. doi:10.1098/rstb.2007.2154

Kuhl, P. K., & Williams, K. a. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606.

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, *124*(5), 3235–3248.

Lee, C.-Y., & Lee, Y.-F. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *Journal of the Acoustical Society of America*, *127*(1), 481–490.

Lee, C.-Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics*, *36*(4), 537–563. doi:10.1016/j.wocn.2008.01.002

Lee, J., Perrachione, T. K., Dees, T. M., & Wong, P. C. (2007). Differential effects of stimulus variability and learners' pre-existing pitch perception ability in lexical tone learning by native English speakers. In *16th Meeting of the International Congress of Phonetic Sciences*. Retrieved from http://cns.northwestern.edu/pubs/pdfs/ICPhS_Jiyeon.pdf

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, Wiley.

Li, X., Shu, H., Liu, Y., & Li, P. (2006). Mental representation of verb meaning: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 18*(10), 1774–1787.

Liberman, A., Harris, K., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. doi:10.1037/h0044417

Lidji, P., Jolicœur, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology, 121*(4), 533–541. doi:10.1016/j.clinph.2009.12.018

Liu, H.-M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43*(4), 912.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language, 36*(04), 909–922. doi:10.1017/S030500090800929X

Massaro, D. W. (2001). Speech perception. In (P. Smelser, Baltes, & Kintsch, Eds.)*International encyclopedia of social and behavioral sciences*. Amsterdam: Elsevier. Retrieved from http://mambo.ucsc.edu/pdf/iesbs.pdf

Mattock, K., & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy, 10*(3), 241–265. doi:10.1207/s15327078in1003_3

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367–1381. doi:10.1016/j.cognition.2007.07.002

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

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(02), 201–233. doi:10.1017/S0140525X00078407

Näätänen, R., Jiang, D., Lavikainen, J., Reinikainen, K., & Paavilainen, P. (1993). Event-related potentials reveal a memory trace for temporal features. *NeuroReport*, *5*(3), 310–312. doi:http://dx.doi.org/10.1097/00001756-199312000-00033

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., & Ponton, C. (2012). The mismatch negativity (MMN) – A unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clinical Neurophysiology*, *123*(3), 424–458. doi:10.1016/j.clinph.2011.09.020

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*(4), 779–784. doi:10.1016/S0163-6383(98)90044-3

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, *92*(1), 26–32. doi:10.1016/j.bandl.2004.05.005

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. doi:10.1016/j.wocn.2010.09.003

Perani, D. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*(10), 1841–1852. doi:10.1093/brain/121.10.1841

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*(2), 191–215. doi:10.1006/jpho.2001.0134

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. doi:10.1016/j.clinph.2007.04.019

Rao, A., Zhang, Y., & Miller, S. (2010). Selective listening of concurrent auditory stimuli: An event-related potential study. *Hearing Research*, *268*(1–2), 123–132. doi:10.1016/j.heares.2010.05.013

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, *111*(1–2), 13–23. doi:10.1016/S0166-4328(00)00139-X

Ronnberg, J., Andersson, J., Samuelsson, S., Soderfeldt, B., Lyxell, B., & Risberg, J. (1999). A speechreading expert: The case of MM. *Journal of Speech, Language and Hearing Research*, *42*(1), 5.

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): Emitted MMN to violations of an auditory gestalt. *Psychophysiology*, *49*(4), 544–548. doi:10.1111/j.1469-8986.2011.01336.x

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436. doi:10.1006/jpho.1997.0051

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen, H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, *17*(1), 1–12.

Scientific Learning Corporation. (1999). *National field trial results: Results of Fast ForWord training for children with language and reading problems.* Berkeley, CA: Scientific Learning Corporation.

Sittiprapaporn, W., Chindaduangratn, C., Ter Vaniemi, M., & Khotchabhakdi, N. (2003). Preattentive Processing of Lexical Tone Perception by the Human Brain as Indexed by the Mismatch Negativity Paradigm. *Annals of the New York Academy of Science*, *999*(1), 199–203. doi:10.1996/annals.1284.029

Sleve, L. R., & Miyake, A. (2006). Individual Differences in Second-Language Proficiency: Does Musical Ability Matter? *Psychological Science*, *17*(8), 675–681. doi:10.2307/40064434

Tamminen, H., Peltola, M. S., Toivonen, H., Kujala, T., & Näätänen, R. (2013). Phonological processing differences in bilinguals and monolinguals. *International Journal of Psychophysiology*, *87*(1), 8–12. doi:10.1016/j.ijpsycho.2012.10.003

Thomson, R. I. (2012). Improving L2 Listeners' Perception of English Vowels: A Computer-Mediated Approach. *Language Learning*, *62*(4), 1231–1258.

US Census Bureau Public Information. (2010). New Census Bureau Report Analyzes Nation's Linguistic Diversity - American Community Survey (ACS) - Newsroom - U.S. Census Bureau. Retrieved July 18, 2013, from http://www.census.gov/newsroom/releases/archives/american_community_survey_acs/cb10-cn58.html

Wan, I.-P. (2007). On the phonological organization of Mandarin tones. *Lingua*, *117*(10), 1715–1738. doi:10.1016/j.lingua.2006.10.002

Wang, X. (2013). Perception of Mandarin Tones: The Effect of L1 Background and Training. *The Modern Language Journal*, *97*(1), 144–160. doi:10.1111/j.1540-4781.2013.01386.x

Wang, X.-D., Gu, F., He, K., Chen, L.-H., & Chen, L. (2012). Preattentive Extraction of Abstract Auditory Rules in Speech Sound Stream: A Mismatch Negativity Study

Using Lexical Tones. *PLoS ONE*, *7*(1), e30027. doi:10.1371/journal.pone.0030027

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic Perception of Mandarin Tones by Chinese and American Listeners. *Brain and Language*, *78*(3), 332–348. doi:10.1006/brln.2001.2474

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

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience*, *15*(7), 1019–1027.

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

Wang, Y., Zhang, Y., Cooper, A., & Dovan, M. (2011). EFFECTS OF TRAINING ON THE PROCESSING OF SPEECH AND NON-SPEECH TONE: AN EVENT-RELATED POTENTIAL STUDY. Presented at the Psycholinguistic Representation of Tone Conference, Hong Kong. Retrieved from http://www.sfu.ca/content/dam/sfu/lablab/pdfs/papers/plrt2011-wang-zhang-cooper-dovan_revision_submission.pdf

Wayland, R., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning*, *54*(4), 681–712.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*(4), 654–662. doi:10.1016/j.wocn.2010.10.001

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *43*(2), 230.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147–162. doi:10.1016/j.cognition.2006.03.006

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. doi:10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*(3), 233–251. doi:10.1002/dev.20060

Wong, P., & Ettlinger, M. (2011). Predictors of spoken language learning. *Journal of Communication Disorders*, *44*(5), 564–567. doi:10.1016/j.jcomdis.2011.04.003

Wong, P., Perrachione, T., Gunasekera, G., & Chandrasekaran, B. (2009). Communication Disorders in Speakers of Tone Languages: Etiological Bases and

Clinical Considerations. *Seminars in Speech and Language, 30*(03), 162–173. doi:10.1055/s-0029-1225953

Wong, P., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*(04). doi:10.1017/S0142716407070312

Wong, P., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, *28*(10), 995–1006. doi:10.1002/hbm.20330

Wong, P., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*. doi:10.1038/nn1872

Wu, J.-L., Yang, H.-M., Lin, Y.-H., & Fu, Q.-J. (2007). Effects of Computer-Assisted Speech Training on Mandarin-Speaking Hearing-Impaired Children. *Audiology & Neuro-Otology*, *12*(5), 307–312.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223–231. doi:10.1016/j.neuroscience.2010.06.077

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Acoustical Society of America*, *120*(2), 1063–1074.

Ye, Z., Luo, Y., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research*, *1071*(1), 186–196. doi:10.1016/j.brainres.2005.11.085

Yip, M. (2002). *Tone*. Cambridge University Press.

Zhang, Y. (2013). *Categorical Perception*. Unpublished encyclopedia article manuscript.

Zhang, Y., & Cheng, B. (2011). Brain Plasticity and Phonetic Training for English-as-a-Second-Language Learners. In D. J. Alonso, *English as a Second Language* (pp. 1–49). New York: Nova Science Publishers.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., … Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain: Neural coding of exaggerated speech. *Developmental Science*, *14*(3), 566–581. doi:10.1111/j.1467-7687.2010.01004.x

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., … Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage*, *46*(1), 226–240. doi:10.1016/j.neuroimage.2009.01.028

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703–720. doi:10.1016/j.neuroimage.2005.02.040

Appendix A

Table 5

Non-native speech Sound Contrast Training Studies

Study	# Participants: Native Language	Trained Language	Hours of Training	Variable of Interest	Use of Exaggeration	% Accuracy Improvements
(Kaan et al., 2013)	10: English 10: Mandarin 11:Thai	Thai	unstated, 2 days	Alpha and Gamma Waves	No	Not measured
(Y. Wang et al., 2011)	19: Canadian English	Mandarin	2 hours	Speech vs Non-Speech	No	36-71%
(X. Wang, 2013)	37: Hmong 22: Japanese 20: English	Mandarin	6 hours	Language Background	No	14-24%
(J. Lee et al., 2007)	47: English	English pseudowords mimicking Mandarin	4 hours	Word Learning	No	Average 56%
(Wayland & Guion, 2004)	6: Thai 6: English 5: Taiwanese-Mandarin 1: Mandarin	Thai	2.5 hours	Inter-Stimulus Interval (ISI)	No	Chinese listeners outperformed English. No ISI effect.
(Wong, Perrachione, et al., 2007; Wong & Perrachione, 2007)	17: English	Synthesized English pseudowords mimicking Mandarin	until criterion reached	Word Learning	No	Average 50.98%
(Wayland et al., 2010)	30: non-tone language speakers 15: non-musicians 15: musicians	Naturally produced /ba/, /bi/ and /bu/ re-synthesized to have either a rising or falling pitch	1.5 hours	Musical Experience	No	3.3-4.9%

(Chandrasekaran et al., 2010)	16: English	English pseudowords mimicking Mandarin	4.5 hours	Cue-Weighting	No	Increased identification and labeling of pitch direction
(Y. Wang et al., 1999)	8: English	Mandarin	5.3 hours	High-Variability Training	No	Average 21%
(Kaan et al., 2007)	10: English 10: Mandarin 10: Thai	Thai	1 hour	MMN	No	Decreased reaction time
(Y. Wang, Sereno, et al., 2003)	6: English	Mandarin	5.3 hours	fMRI	No	Average 24%
(Kaan et al., 2008)	12: Mandarin 12: English 11: Thai	Synthesized Thai	2 hours	Language Background	No	7.17-14.01%
(Francis et al., 2008)	10: Mandarin 10: English 12: Cantonese	Cantonese	10 hours	Language Background	No	8.3-16.7%
(Zhang et al., 2009)(Y. Wang, Jongman, et al., 2003)	9: Japanese	English /l/ and /r/	12 hours	Computer Training	Yes	Average 21.6%

Appendix B

Participants Needed

for a research study on bilingualism and brain plasticity in speech training

Who: American students ages 18-25

What: You will be participating in a speech-training study learning Mandarin tones and participating in a pre-test and post-test.

When: The study will be conducted over 4 weeks, about 15 hours total.

Reimbursement: $150 total

If interested in the bilingualism and brain plasticity study, contact: zhanglab@umn.edu	If interested in the bilingualism and brain plasticity study, contact: zhanglab@umn.edu	If interested in the bilingualism and brain plasticity study, contact: zhanglab@umn.edu	If interested in the bilingualism and brain plasticity study, contact: zhanglab@umn.edu	If interested in the bilingualism and brain plasticity study, contact: zhanglab@umn.edu	If interested in the bilingualism and brain plasticity study, contact: zhanglab@umn.edu

Appendix C
CONSENT FORM FOR ADULT PARTICIPANTS

Bilingualism and Brain Plasticity in Speech Training

You are invited to participate in a research study titled "**Bilingualism and Brain Plasticity in Speech Training**". This study is being conducted in the Department of Speech-Language-Hearing Sciences at the University of Minnesota. You were selected as a possible participant because you fit the profile we are interested in studying and have no history of brain damage. The target populations of the study are right-handed, monolingual speakers of English with normal hearing between the ages of 18-45 who have no prior formal or informal learning experience with Mandarin Chinese, no history of speech and language disorders, and no professional musical training.

This form may contain words or language that is unfamiliar to you. Please ask the researcher if you would like something explained to you. We ask that you read this form and ask any questions that you may have before agreeing to be in the study.

The researchers in this project include Yang Zhang (Ph.D.) of the Department of Speech-Language-Hearing Sciences, and Christina Heinzen (B.A.) of the Department of Speech-Language-Hearing Sciences.

Background Information

The purpose of the study is to examine whether perceptual training can affect perception of non-native speech contrasts. We will take behavioral and brain measures both before and after a computerized perceptual training program.

We hope that the results of this study will help us better understand how the adult brain is able to learn to distinguish between speech contrasts that are not a part of the listener's native language. The results of this experiment may lead to a better understanding of the neurological characteristics of second-language learning in adulthood.

Procedures

If you agree to be in this study, your participation will include the following: a Pre-test one week before training, 10 days of hour-long training sessions, and a Post-test one week following training. The Pre-test and Post-test will include an electroencephalogram (EEG) and a behavioral test of perception.

In the EEG recording sessions, you will sit in a comfortable chair in a sound-treated booth. A stretchable cap with electrodes sewn into it will be fit on your head much like a shower cap. The electrodes will touch the scalp on different spots to record electrical brain activities corresponding to those individual spots. The experimenter will put conductive gels on each electrode to automatically record your brain activities as you listen to sequences of sounds while reading a book of your choice. The set-up will take us about 15 minutes.

During the recording you will be asked to sit as still as possible and to relax your face and muscles as much as possible. If you normally wear contact lenses, we suggest you wear glasses to minimize excessive blinking, which can interfere with testing. After testing, your hair will be messy. A hair wash station with sanitized combs, shampoo, towels, and a hair dryer is available for you to use after the experiment is finished.

During the computer perceptual tests and training sessions, you will be directed to listen to sounds and select the one you heard. The program will provide you with corrective feedback and will automatically increase in difficulty as you improve. The training will last approximately ten hours, which will be distributed in hour-long sessions over the course of two weeks.

Throughout the experiments, we will be watching you from a video monitor in the control room via an intercom system. The monitoring system is necessary to ensure proper data collection and timely correction if we see any problematic data. The monitoring video is not recorded. You can stop the session at any time for any reason simply by telling the experimenter that you would like to stop.

Risks and Benefits

You may choose to end participation at anytime without negatively impacting your relationship with the University of Minnesota or the researchers.

This study follows the standard procedures in neurophysiological studies, and there is no known risk. However, you may be bored of listening to the stimuli. Taking a short break may prevent the occurrence of boredom. Application of gel for recording brainwaves on the scalp is a standard procedure. The gel is made of non-toxic and non-allergenic materials and is completely safe; it can be washed off easily using tap water. Accommodations will be made so that you can wash your hair and clean up after the experiment has been completed.

Findings in this study may help us better understand how the adult brain is able to learn to distinguish between speech contrasts that are not a part of the speaker's native language. The results of this experiment may lead to a better understanding of the neurological characteristics of second-language learning in adulthood.

Compensation

It will take a total of 15 hours over the course of 4 weeks to complete this study. You will receive compensation of $10.00 per hour of your participation in this study. If you are unable or unwilling to complete the study, you will be compensated at a pro-rated rate of $10.00 per hour of participation completed.

Confidentiality

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. Video recording of the session will be used only for analysis purposes. All such recording will be kept securely and will be destroyed after analysis of data.

Voluntary Nature of the Study

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions

The researchers conducting this study are: Dr. Yang Zhang and Dr. Zhang's research assistants.
You may ask any questions you have now. If you have questions later, you are encouraged to contact them at 115 Shevlin Hall, 164 Pillsbury Drive, Minneapolis, MN 55455, Phone: (612) 624-3322, or email: zhang470@umn.edu, heinz096@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Experiment Consent

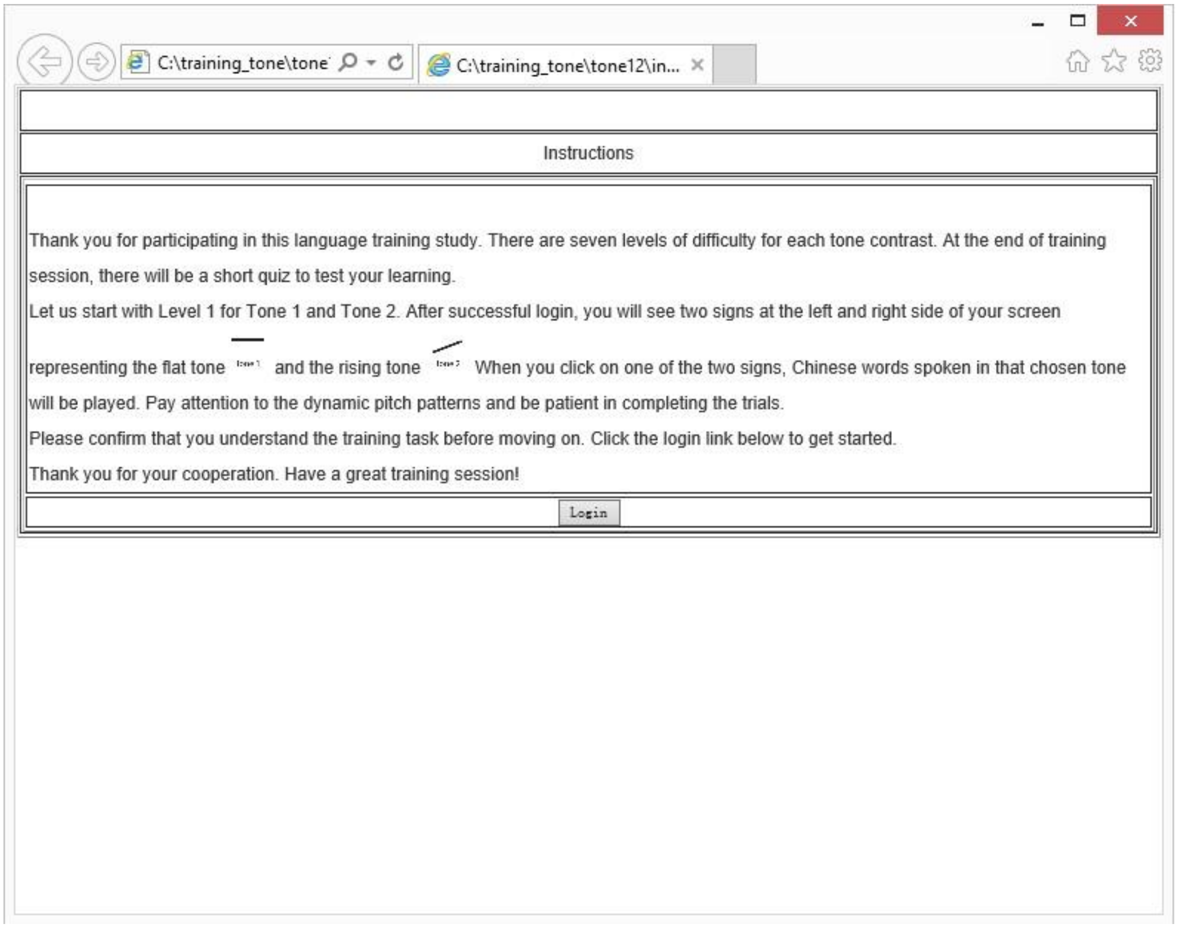
I have read the above information. I have asked questions and have received answers. I consent to participate in the adult study.

Signature: ____ Date: _____________

Signature of Investigator: ____________________________ Date: ______________

Appendix D

Sample screenshots of training program



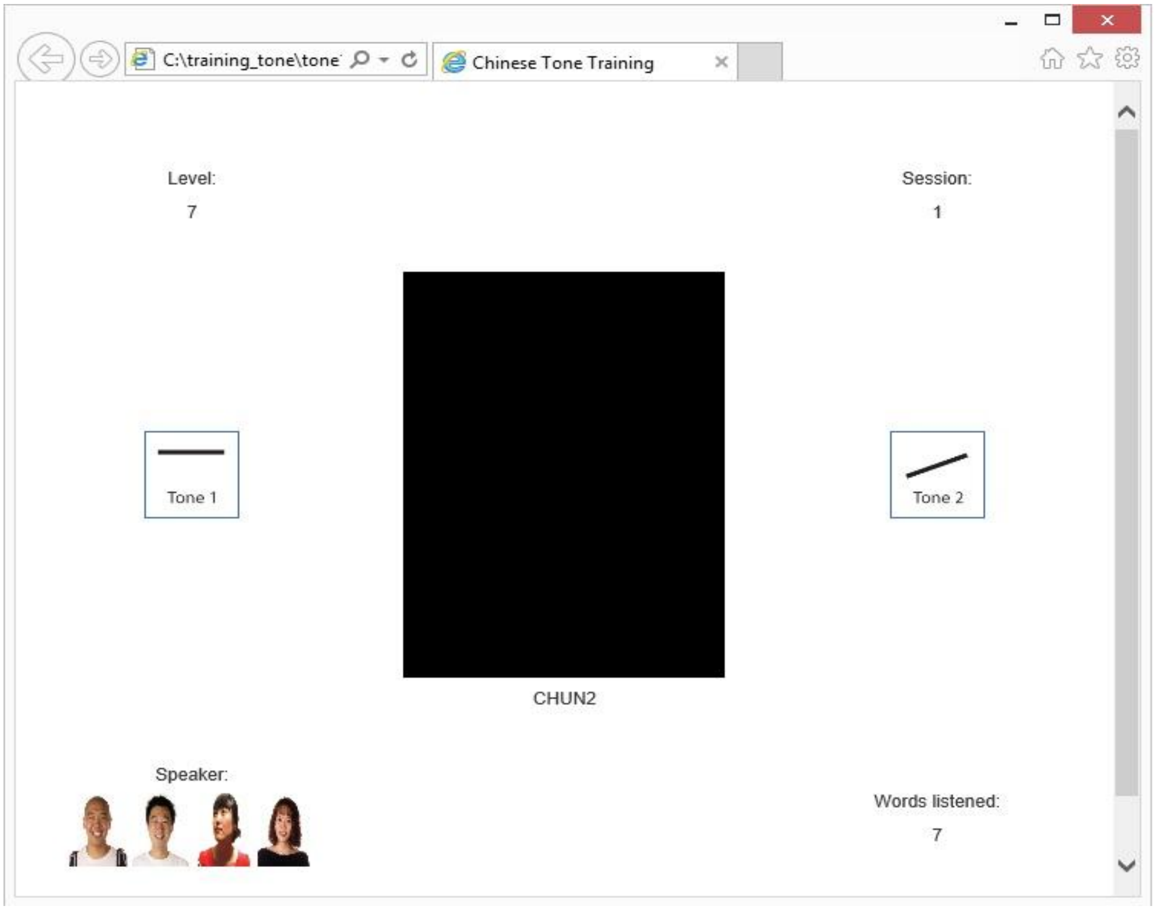
Instructions

Thank you for participating in this language training study. There are seven levels of difficulty for each tone contrast. At the end of training session, there will be a short quiz to test your learning.

Let us start with Level 1 for Tone 1 and Tone 2. After successful login, you will see two signs at the left and right side of your screen representing the flat tone — and the rising tone ╱ When you click on one of the two signs, Chinese words spoken in that chosen tone will be played. Pay attention to the dynamic pitch patterns and be patient in completing the trials.

Please confirm that you understand the training task before moving on. Click the login link below to get started.

Thank you for your cooperation. Have a great training session!



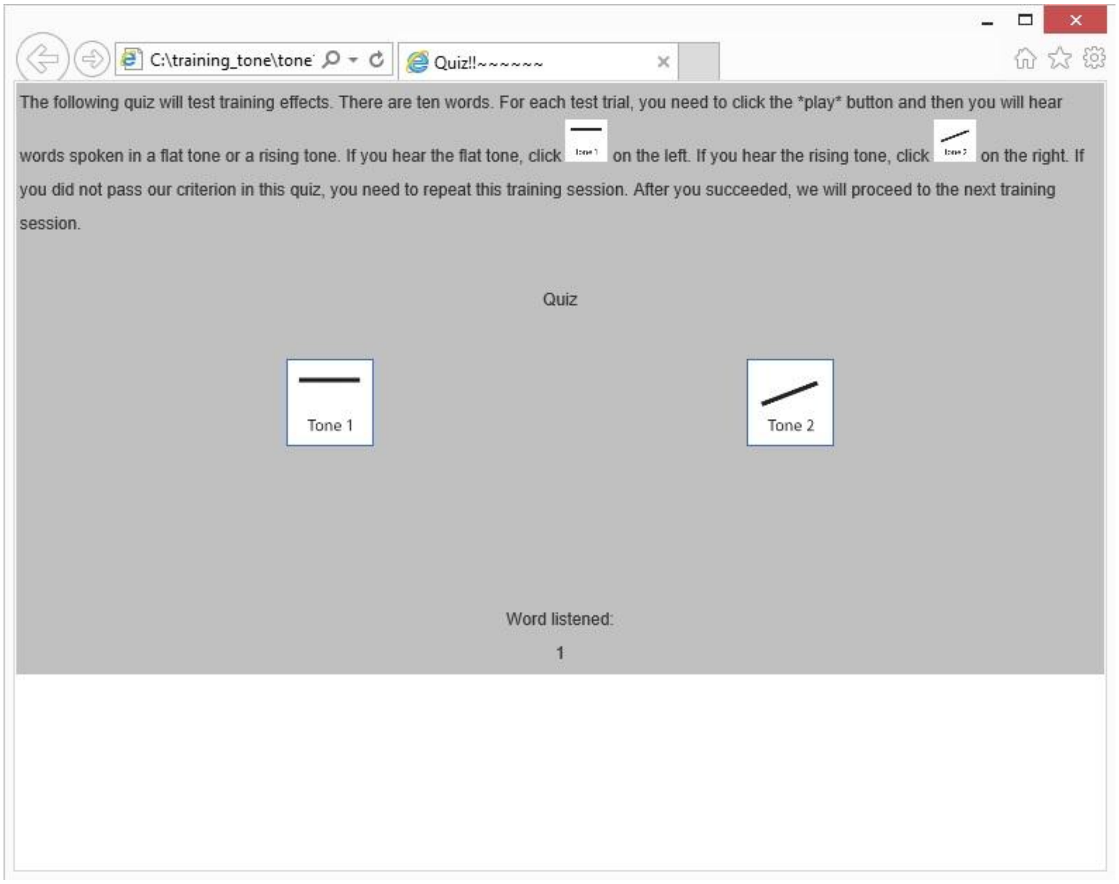
Level: 7

Session: 1

CHUN2

Speaker:

Words listened: 7



The following quiz will test training effects. There are ten words. For each test trial, you need to click the *play* button and then you will hear words spoken in a flat tone or a rising tone. If you hear the flat tone, click Tone 1 on the left. If you hear the rising tone, click Tone 2 on the right. If you did not pass our criterion in this quiz, you need to repeat this training session. After you succeeded, we will proceed to the next training session.

Quiz

Word listened:

1

Appendix E

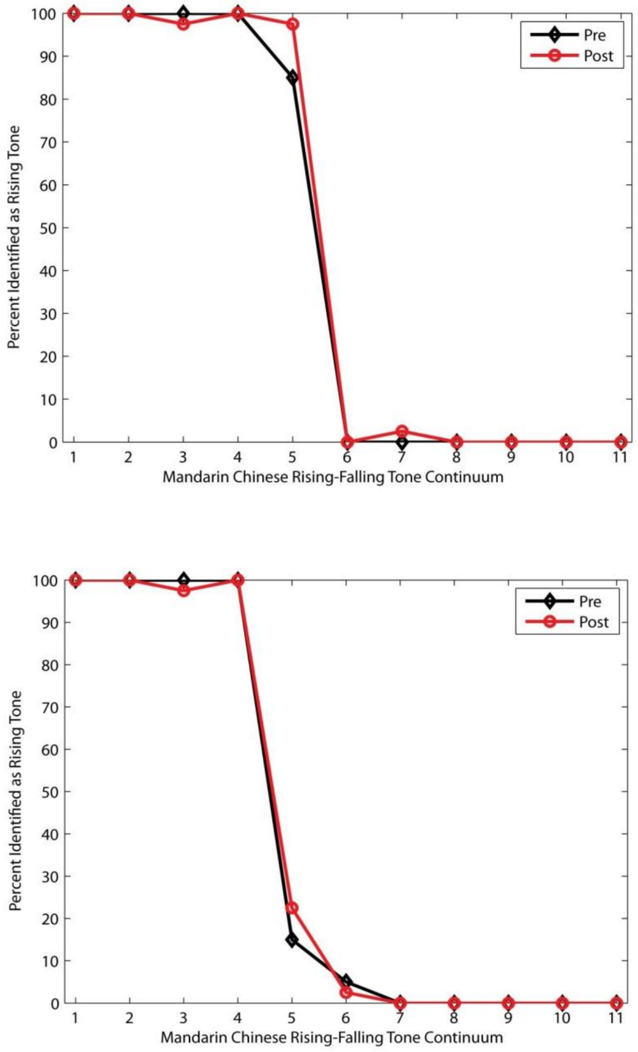


Figure 17. Identification functions of the two excluded participants. This figure depicts the identification functions of the excluded participants for the speech tone continuum both pre- (black diamonds) and post- (red circles) training. Because they exhibited near-perfect categorical perception prior to training, their data were excluded from analysis.

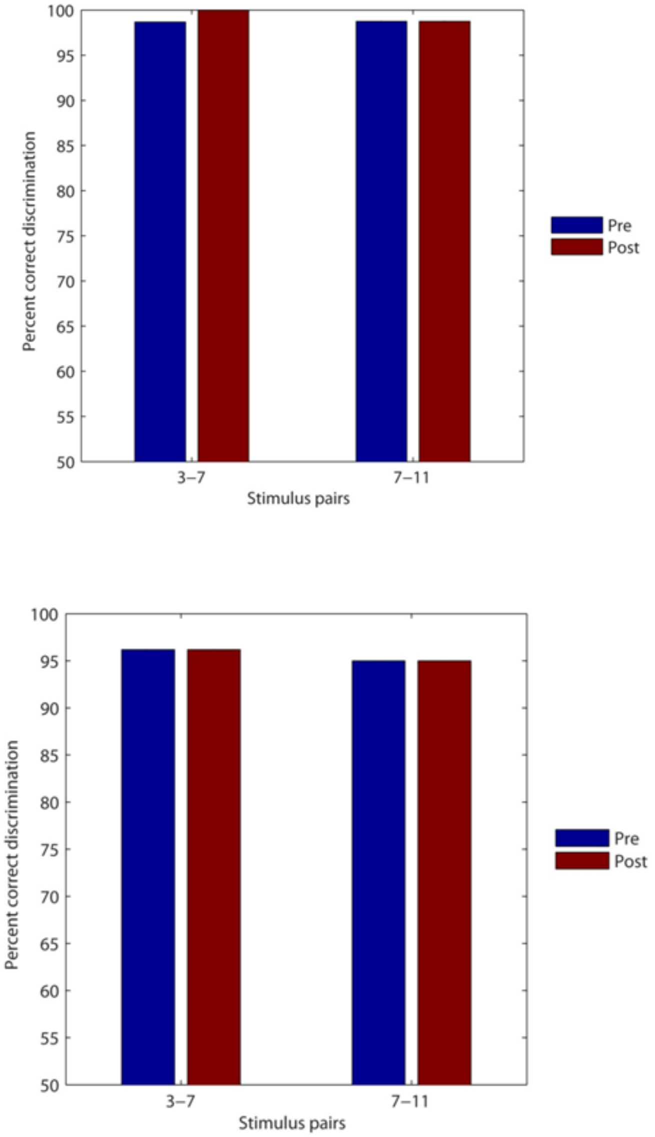


Figure 18. Discrimination data from the two excluded participants. This figure depicts percent correct discrimination of the excluded participants for across- and within-category pairs (3-7 and 7-11 respectively) from the speech continuum. Because they exhibited near-perfect categorical perception prior to training, their data were excluded from analysis.