NEUROPHYSIOLOGICAL INDICES OF MANDARIN LEXICAL TONE PROCESSING:

THE ROLE OF SENSORY MEMORY AND LONG-TERM LANGUAGE EXPERIENCE

Dissertation

by

Yan H. Yu

A dissertation submitted to the Graduate Program in Speech, Language and Hearing Sciences in
partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2013

This manuscript has been read and accepted
for the Graduate Program in Speech, Language and Hearing Sciences
in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Valerie L. Shafer

Date:____________________ ____
Chair of Examining Committee

Dr. Klara Marton

Date:____________________ ____
Executive Officer

Dr. Elyse S. Sussman

Dr. Douglas H. Whalen

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Neurophysiological indices of Mandarin lexical tone processing: The role of sensory memory and long-term language experience

by

Yan H. Yu

Advisor: Dr. Valerie L. Shafer

Language experience enhances discrimination of speech contrasts at a behavioral, perceptual level, as well as at a pre-attentive level, as indexed by event-related potential (ERP) mismatch negativity (MMN) responses. The enhanced sensitivity could be the result of changes in acoustic resolution and/or long-term memory representations of the relevant information in auditory cortex. To examine these possibilities, we used a short (ca. 600 ms) versus long (ca. 2600 ms) interstimulus interval in a passive, oddball discrimination task while obtaining ERPs. These ISI differences were used to test whether cross-linguistic differences in processing Mandarin lexical tone are a function of differences in acoustic resolution and/or differences in long-term memory representations. Mandarin and English listeners listened to bisyllabic nonword tokens that differed in lexical tone categories using a multiple oddball paradigm. The results revealed robust MMNs to both easy and difficult tone differences for both groups at short ISIs. At long ISIs, there was no change or an enhanced MMN in the Mandarin group, but reduced MMN amplitude in the English group. In addition, Mandarin listeners showed a larger late negativity (LN) discriminative response than the English listeners for tone contrasts in the long ISI condition. Lack of robust MMN and LN to tone contrasts in English listeners under the long ISI condition suggests that English listeners do not maintain long-term memory representations of lexical tones that are sufficient for discriminating phonemically different Mandarin tones. These results

support the claim that language experience modulates neural representation of lexical tones, largely at the phonemic level. They also suggest that the acoustic correlates of tone are fairly robust and easily discriminated at short ISIs, when the auditory memory trace is strong. At longer ISIs beyond 2.5 s language-specific experience is necessary for robust discrimination.

Acknowledgements

This dissertation would not have been possible without the help of so many people in so many ways. First and foremost, I am deeply indebted to my mentor and dissertation advisor, Dr. Valerie Shafer, for instilling in me the quality of a good scientist, for providing me her unfailing support and for her confidence in me during the past years. I have learned a tremendous amount from Dr. Shafer, and I sincerely hope to reflect the scholarly wisdom I have accrued from her in years to come.

I would like to thank Dr. Elyse Sussman and Dr. Douglas H. Whalen for serving on my committee and for their mentorship. My thanks in advance for their continuous guidance on disseminating the dissertation studies. I thank Dr. Joseph Dien for being my external reviewer.

It is a great opportunity to express my gratitude to all faculty and staff members in the graduate program of Speech, Language and Hearing Sciences at the Graduate Center of the City University of New York for the knowledge, skills and support that I have received during these years in the program. My special thanks to Drs. Klara Marton, Richard Schwartz, and Winifred Strange for their invaluable guidance for my second-level exam project. Under the aegis of Dr. Valerie Shafer, I had the opportunities to take part in several research projects, and I take this opportunity to thank all the collaborators and the great master minds that I have learned from by collaborating on these various projects. My thanks go to Drs. April Benasich, Chiara Cantiani, Judy Flax, Michelle MacRoy-Higgins, and Richard Schwartz for the discussions and guidance on the Autism project. Thanks also go out to Drs. Arild Hestivik, Yael Werth, and Baila Epstein for the collaborations on the “Past Tense” study. I will forever cherish all the fun and stressful moments on the “Baby Grant” project (a grant project on babies, actually) and other projects that I had worked with the past and present DNL members. I greatly appreciate all the help and

support from the DNL members and the volunteer members of the DNL. All these experiences not only paved the direction for this dissertation, but also made the journey a smoother and more enjoyable one.

During the course of this dissertation, I have received help from Drs. Joseph Dien and Nathan Maxfield regarding Principal Component Analysis, from Simon Henin and Ning Xu on Matlab, and from Luca Campenelli regarding analysis in R. My sincere thanks to them all.

This dissertation was supported by Rees Dissertation Fellowship from the Graduate Center, City University of New York. I would like to acknowledge the generous support.

Life won't be the same without the "chicken soup for the soul" and the exotic Midtown takeouts for the palate frequently from our NeuroStudy group. A very special thanks to the core members of our NeuroStudy group, Hia Datta, Jungmoon Hyun, Miwako Hisagi, and Monica Wagner for their support and friendship over the years. I also extend my thanks to all my friends at the DNL and the Speech-Language-Hearing department for their friendship.

My very special thanks to my previous mentor, Dr. Fredericka Bell-Berti, for cultivating my interest in speech science when I was a master-level graduate student, and for continuing to be my mentor in and out of academia.

Finally, I come to my most personal source of gratitude. Over the journey of my doctoral study, my husband, Jianxun, and my now 6.5-year-old son, Jeffrey, have been very supportive and understanding. I owe everything I have today to my father, Jiang Tao Yu, and my mother, Sheng Rong Tu, who showed me the true worth of hard work, and who always have unwavering faith and confidence in my ability and in me is what has shaped me to be the person I am today. I would also like to thank my sisters and brother for their love. I dedicate this dissertation to my family.

Table of Contents

List of Tables **Page**

List of figures **Page**

List of Tables for Vowel Study **Page**

List of figures for Vowel Study **Page**

List of Additional Tables **Page**

List of Additional Figures **Page**

Chapter 1. Overall Objectives and Design of the Dissertation and Main Hypotheses

The neurophysiology of lexical tone has been understudied despite the fact that lexical tone is used in more than half of the world languages. The general purpose of this dissertation was to address some of the unresolved issues regarding the neural representation of lexical tone processing. The primary goal of this dissertation was to examine the time-course of neural representation and processing of lexical tone under two different memory delay conditions at a largely automatic, attention-independent level of processing.

To do this, a passive oddball design was used to examine neural correlates of discrimination of lexical tones at two different interstimulus intervals (ISIs). Two tone contrasts (T3-T1 and T3-T2) and two vowel contrasts (/u/-/i/ and /u/ -/y/) were used in multi-token bisyllabic nonwords. The mismatch negativity (MMN) and the late negativity (LN) event-related potential ERP discriminative responses were expected to be elicited and modulated by an array of factors, including long-term linguistic experience, degree of sensory memory decay, and the acoustic distinctiveness of the stimulus contrast. It is possible that ISI interacts differently with different types of phonemic contrasts (Pisoni, 1973). For perception of differences in segmental phonemes, short ISIs may be sufficient to reveal the influence of linguistic experience because the acoustic-phonetic differences are brief in time (e.g., for stop consonants and for vowels) and/or acoustically minimal (in terms of spectral, duration or intensity differences). However, the perception of lexical tone unfolds across the longer time period of a full syllable, especially some tone contrasts for the target language, Mandarin Chinese. For these reasons, a longer ISI than previous studies (e.g., Chandrasakaran et al., 2007; Kaan et al., 2008) may be necessary to clearly reveal cross-linguistic differences in lexical tone processing. Another reason for using long ISI is to separate the contribution of acoustic/phonetic processing from phonological processing.

A passive task with repeated, frequent standards and multiple, less frequent oddballs was used while event-related potentials (ERPs) were obtained from 31 native speakers of English and 32 native speakers of Mandarin (who learned English after age 15 years). Memory for lexical tone patterns was examined using two stimulus presentation rates (fast rate: 900 ms per bisyllabic sound, equivalent to an interstimulus interval of 575 ms; slow rate: 3000 ms per bisyllabic sound, equivalent to an ISI of ca. 2675 ms). The two ISIs were chosen based on the behavioral literature (e.g., Pisoni, 1973, Werker and Logan, 1985) and a pilot study. Multi-token natural speech stimuli were used to increase the reliance on memory for higher-level abstract phonemic processing as opposed to lower level acoustic/phonetic processing. Behavioral discrimination and identification tasks were administered to participants after the ERP experiment to ensure that participants perceived the stimuli as expected.

This dissertation tested the following main hypotheses:

1. Mandarin lexical tone can be easily discriminated when the ISI is short for both English and Mandarin listeners because the acoustic correlates of the contrasting tones are sufficiently different to allow discrimination at a short delay.
2. At a long memory delay, listeners must rely on long-term memory representations.
    1) For Mandarin listeners, experience with lexical tone leads to maintaining information in the memory trace that allows for robust discrimination.
    2) For English listeners, lack of experience with tone results in insufficient information in the memory trace for good discrimination.

The neural measure of discrimination will reveal whether sufficient information is maintained in the memory trace to allow for discrimination without attention. Specifically,

1. Increased ISI will have a different effect for native and non-native listeners on the MMN and LN to speech contrasts because long-term phonemic representations will modulate the quality and durability of the auditory sensory memory trace;
2. The late negativity (LN) will be larger for Mandarin than English listeners because LN will reflect differences in tone contour shape, which require a longer time interval;
3. In terms of hemispheric effect, lexical tone processing in native speakers will involve more left hemisphere lateralization at the scalp potential level, and this will be reflected in MMN and/or LN amplitudes and latencies.

Chapter 2. Introduction

Mandarin Chinese is a tone language with more speakers than any other tone language. However, the neural mechanisms of Mandarin lexical tone have been understudied. In this chapter, a brief introduction of lexical tone will be provided, factors that affect lexical tone processing will be discussed, and previous studies on two neurophysiological measures of speech processing, mismatch negativity (MMN) and late negativity (LN), will be reviewed.

2.1. A brief introduction of lexical tone

Different from the perceptual term pitch, and acoustical term fundamental frequency (F0), tone is a linguistic term that describes language-specific phonological features, and it is measured acoustically using F0, which reflects the rate of the vocal folds vibration during the production of a syllable (Yip, 2002). A lexical tone is a pitch variation pattern that is associated with the basic meanings of individual words. A language is a tone language if changing the pitch of a word can result in a change in meaning of that word (Yip, 2002, p.1). It is estimated that 60-70% of the world languages are tonal languages. Most sino-Tibetan languages including most forms of Chinese, the vast majority of Niger–Congo languages including Zulu and Yoruba, all Khoisan languages in southern Africa, slightly more than half of the Athabaskan languages, such as Navajo in North America, and a large number of South and Central American languages are tonal. Mandarin Chinese, the most widely spoken language, is a tonal language. As in most tone languages, in Mandarin the smallest phonological unit that can bear a tone is a syllable (Fromkin, 1978, Xu & Wang, 2001; Yip, 2002). Lexical tones can be categorized as level tones or contour tones. A level tone has a relatively steady F0 level throughout the syllable when produced in isolation, while a contour tone is characterized by a significant shift in F0 from one level to another within the syllable. Lexical tone produced in connected speech may have F0 contours that differ drastically from the canonical forms due to co-articulation effect, intonation, and stress effect (Li, Le, & Qian, 2002; Xu, 1992, 1997, & 2006; Shih, 1988; Wong, 1999; Gu & Lee, 2007).While F0 functions as the dominant cue for tone recognition in isolation (e.g., Xu, 1997), other acoustic cues such as amplitude contour (e.g., Whalen & Xu, 1992; Kong & Zeng, 2006) or duration (e.g., Fu & Zeng, 2000; Guo et al., 2008) can contribute to Mandarin tone perception as well.

As displayed in Appendix Figure A1, Mandarin has one level tone and three contour tones in stressed syllables. Tone 1 (e.g., bi1,“逼”， ‘to force’) has a high and essentially level frequency contour. Tone 2 (e.g., bi2,“鼻” ‘nose’) has a dipping start and then changes into a rising contour at approximately 20% of the duration into the vowel. Tone 3 (e.g., bi3, “笔” ‘pen’ or “比” 'to compare') also has a dipping start and then changes into a rising contour at a point approximately 50% of the duration of the syllable; and Tone 4 (e.g., bi4,“壁”， ‘wall’) has a falling contour (Howie, 1976). Sometimes, linguists describe these tones in “tone letter” using two or three numbers to indicate the beginning and ending F0 of the tone (see Appendix Figure A1 for details). The F0 contours illustrated in Figure A1 reflects the F0 contours of tones produced in isolation (also called canonical F0).

2.2. Effect of interstimulus interval on behavioral responses

Speech discrimination is generally more challenging for nonnative listeners when the signal is presented under less optimal listening conditions (e.g., van Wijngaarden, Steeneken, &

Houtgast, 2002; Strange, 2011). A few behavioral studies have demonstrated that increased memory delay results in changes in discrimination and categorization of speech sounds (Pisoni, 1973; Werker & Logan, 1985). Specifically, these studies showed that at longer interstimulus intervals (ISIs), listeners discriminated and categorized speech sounds based on whether these sounds make meaningful (phonemic) distinctions in the native language (e.g., /b/ and /d/ in “big” vs. “dig”). At very brief ISIs, listeners were able to discriminate fine acoustic (phonetic) differences that may not be meaningful in the native language (e.g., dental [d] vs. retroflex [d], which is phonemic in Hindi, but not English). In these studies, comparisons of short versus long ISI in tasks requiring speech sound discrimination suggested that speech information could be stored in an auditory, phonetic and phonemic code, but that the auditory and phonetic codes decay rapidly. In contrast, the phonemic code is a long-term memory representation. Despite the differences between vowel and consonant, these few behavioral studies showed that when the ISI is very short (e.g., less than 500 ms), speech information can be coded at the acoustic/phonetic level, whereas when ISI is long (e.g., lengthened to 1500 ms, or 2 s), the information is maintained only at the phonemic level due to the decay of echoic sensory memory representations (Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda, & Keller, 1996; Pisoni, 1973; Werker & Logan, 1985). These studies mostly focused on segmental contrasts. It is less clear whether lexical tone will show the same pattern of processing as segmental information because the acoustic-phonetic properties of segmental elements are different from that of lexical tone. In a Thai lexical tone training study, Wayland and Guion (2004) examined native English (NE) and native Chinese (NC) listeners’ ability to identify and discriminate the mid- vs. low-tone contrast in Thai before and after auditory training under two ISI presentation rates (500 ms vs. 1500 ms), and found that the NC group outperformed the NE group in their ability to discriminate the two Thai tones under the 500-ms ISI condition before training and under both ISI conditions after training. No studies have examined the neural mechanisms involved in the decay of echoic sensory memory for phonemic, lexical tone contrast. The current study is designed to fill this gap.

2.3. Cross-linguistic speech perception model

A number of models in the speech perception field have been proposed to account for listeners' perception of non-native contrasts. The Automatic Selective Perception model (ASP)(Strange, 2011; Strange & Shafer, 2008) is of particular interest to this study. The Automatic Selective Perception (ASP) model proposes that native language listeners have a learned habit of listening that leads them to selectively weight phonetic cues. This selective weighting (or selective perceptual routines, SPRs) allows them to automatically recover the relevant acoustic-phonetic information needed to identify words in the L1. While listening to non-native (or second language) stimuli, their specific L1 SPRs interfere with nonnative speech processing, especially under non-optimal listening situations, such as situations in which attention and memory resources are limited, speaking rate is fast, or the stimulus contrasts are complex (Strange, 2011; Strange & Shafer, 2008). The present study will use a short and a long ISI conditions with a preattentative listening paradigm. The hypothesis is that the lack of attention support together with the long ISI condition will lead English listeners to fall back on their native language SPRs without attention or increased effort, therefore, the English listeners will encounter difficulties in perceiving lexical tone contrast.

2.4. Effect of interstimulus interval on neurophysiological discrimination responses

Auditory sensory memory is commonly assessed using behavioral tasks such as verbally repeating numbers, syllables, words or nonwords of increasing length, or recalling lexical items (Gathercole & Baddeley, 1990). These tasks require immediate and overt responses from subjects, as well as a level of language proficiency that allows understanding of the task instructions and expressive abilities to repeat the syllables. Thus, task performance is affected by attention, motivation, motor dexterity and other cognitive factors. The language-specific lexical and sublexical features of the stimuli also affect the results (e.g., Hulme, Maughan, & Brown, 1991). Consequently, these behavioral methods alone cannot easily tease apart the contributions from different cognitive and linguistic factors, and it is particularly difficult to address whether the experiential changes to speech perception are implemented at an early level of processing by affecting acoustic-phonetic resolution, or only at a later level of processing related to the phonemic representation. An event-related potential (ERP) method designed to elicit the mismatch negativity (MMN) allows us to address this question because it can index neural discrimination at a largely attention-independent level. The MMN is elicited to a stimulus change in a series of repeated stimuli and is seen as an increased negativity at fronto-central sites.

Studies have manipulated ISI to examine the period for which the preceding auditory context remains relevant to the deviant using pure tones (e.g., 1000 Hz versus 1200 Hz). Many of these studies have shown that lengthening the ISI between standard and deviant auditory tones leads to reduced MMN amplitude (Mäntysalo & Näätänen, 1987; Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1987; Böttcher-Gandor & Ullsperger, 1992; Sam, Hari, Rif, & Knuutila, 1993). This finding was often interpreted in terms of duration of sensory memory (Böttcher-Gandor & Ullsperger, 1992; Sams et al., 1993; Winkler, Schröger, & Cowan, 2001) because longer ISI leads to greater sensory memory trace decay for the standard stimulus. Thus, manipulating the ISI between stimuli allows an estimate of the short-term sensory memory duration for the standard stimulus (Mäntysalo & Näätänen, 1987; Näätänen, Paavilainen, Alho, Reinikainen, &Sams, 1987; Böttcher-Gandor& Ullsperger, 1992; Sams, Hari, Rif, & Knuutila, 1993). In adults, the MMN can be elicited with an ISI as long as 10 seconds when the stimuli are auditory tones that differ in frequency by 10 percent (e.g., Standard: 1000 Hz, Deviant: 1100 Hz in Sams et al., 1993). Short-term auditory memory decay may be more rapid in children. Gomes and colleagues found that no MMNs were elicited in children between six and 10 years of age when the intertrain intervals between 1000 and 1200 Hz tones (100 ms in duration) were increased to 8 s (Gomes, Sussman, Ritter, Kurtzberg, Cowan, & Vaughan, 1999). Čeponiene and colleagues found that when the stimuli were auditory tones (1000 & 1100 Hz), there was no MMN amplitude difference between the children with high and low nonword repetition (NWR) performance under either 350 ms or 2000 ms ISI condition. However, when the stimuli were speech (/baka/ & /baga/), MMNs were obtained only in the high performers, albeit the MMN amplitude was reduced in the long ISI condition (Čeponiene, Service, Kurjenluoma, Cheour, &Näätänen, 1999). Smaller MMNs compared to those in controls were found in parents of children with SLI, young children, or children with language impairment when the ISI was longer than 2 s for the selected stimulus differences (Barry, Hardiman, Line, White, Yasin, & Bishop, 2008; Glass, Sachse & von Suchodoletz, 2008; Grossheinrich, Kademann, Bruder, Bartling, & Von Suchodoletz, 2010). Only one speech study has examined the MMN response using an ISI longer than 1.5 second (Čeponienė et al., 1999). Little is known about how language experience affects the duration of the sensory memory trace. We hypothesize that increased ISI will have a differential effect on the MMN to speech contrasts for native and non-native listeners because long-term phonemic representations will modulate the auditory memory trace.

2.5. MMN as an index of speech processing in cross-linguistic studies

Numerous studies have used the MMN as an index of speech processing. Cross-linguistic studies have shown that experience with a native-language speech contrast that is phonemic (i.e., changes meaning) modulates MMN responses for consonants (e.g., Dehaene-Lambertz, 1997; Shafer, Schwartz & Kurtzberg, 2004; Sharma & Dorman, 1999, 2000) and vowels (e.g., Näätänen et al., 1997; Hisagi, Shafer, Strange, & Sussman, 2010; Szymanski, Yund & Woods, 1999; Winkler, at al., 1999a, 1999b). The overall larger MMN peak amplitude for the cross-category compared to within-category speech contrasts in native listener was hypothesized to reflect categorical, phonemic representation at a largely attention-independent level (e.g., Winkler et al., 1999a; Xi Zhang, Shu, Zhang, Li, 2010). More specifically, a cross-category contrast (phonemic) generates a larger MMN than a within-category contrast (acoustic-phonetic), keeping acoustic difference equal. The MMNs elicited to within-category comparisons are presumed to reflect auditory-phonetic sensory representation of the stimulus sequences. For example, the MMN was found to both a cross-category (phonemic) and within-category (acoustic-phonetic) vowel distinction for Finnish, but the MMN was larger for the cross-category than the within category distinction for the Finnish listeners (Näätänen et al., 1997). For Estonian listeners the MMN was robust for both vowel distinctions because both were cross-category in the Estonian language.

MMN is elicited by detection of a change to an otherwise invariant background of homogeneous events. More importantly, the MMN can be elicited to changes in an abstract auditory pattern, for example, a sequence of ascending tones varying in frequency that then change to descending tones (Korzyukov, Winkler, Gumenyuk, & Alho, 2003; Paavilainen, Saarinen, Tervaniemi & Näätänen, 1995; Paavilainen, Jaramillo, & Näätänen, 1998; Paavilainen, Valppu &Näätänen, 2001; Saarinen, Paavilainen, Schröger, Tervanjemi, & Näätänen,, 1992; Sussman, Ritter & Vaughan,1998; Tervaniemi, Maury, & Näätänen, 1994). Furthermore, MMN can be elicited by two or more deviants in the context of a common standard or a global pattern (called a multiple-oddball paradigm) (Sussman, Sheridan, Kreuzer & Winkler, 2003).

The phonemic identity of a speech stimulus is an example of an abstract auditory pattern, and recovery of native-language phonemic identity is hypothesized to occur automatically, and with minimal effort (Strange, 2011). The ability to rapidly and effortlessly extract meaningful phonetic information from a highly variable acoustic signal is critical for later stages of language processing. Strange (2011) argues that non-native speech perception suffers under tasks that pull attention away from or increase difficulty because non-native listeners need attention to extract the relevant non-native cues. Cross-linguistic studies of speech processing have typically used a passive task, in which attention was focused on a book or a video, and thus discrimination of non-native contrasts should be poor (see Näätänen, Paavilainen, Rinne, & Alho, 2007, for a recent review). The study by Hisagi and colleagues showed that with focused attention to a non-native Japanese vowel-duration contrast, English listeners showed more robust MMNs than when attention was directed away from the contrast (Hisagi, Shafer, Strange & Sussman, 2010).

2.6. MMN as an index of lexical tone processing

Most studies of cross-linguistic speech processing revealed less robust MMNs to non-native contrasts. However, the majority of these studies have focused on spectral and temporal cue differences for consonant and vowel contrasts. Studies focusing on the neural processing of lexical tone remain scarce.

One recent study examining MMNs elicited by lexical tones revealed the expected pattern of results, with larger MMNs to between- than within-category F0 differences in native, Mandarin listeners (Xi, Zhang, Shu, Zhang, & Li, 2010). However, other neurophysiological studies examining lexical tone processing have presented results that are less straightforward to interpret. Chandrasekaran and colleagues used an “easy” contrast (an acoustically more distinct pair: Tone 3 versus Tone 1) and a “hard” contrast (an acoustically less distinct pair: Tone 3 versus Tone 2)(Chandrasekaran, Krishnan, & Gandour, 2007). A language group difference was expected for the hard contrast, but to a lesser extent for the easy contrast because the hard contrast is clearly challenging for English listeners, whereas the easy contrast has been shown to be relatively easy in perception tasks for both Mandarin and English listeners (Wang, Spence, Jongman & Sereno, 1999).The results revealed no MMN amplitude difference between native Mandarin listeners and English listeners on the “hard” Mandarin Tone 2 (T2) versus Tone 3 (T3) contrast. This “hard” contrast (T2 versus T3) elicited a rather small negativity for both the Mandarin and English listeners. In addition, there was no difference between the MMN amplitude in the T1-T3 and T2-T3 conditions for the English listeners, whereas Mandarin listeners showed a larger MMN for the easy than the hard contrast. The authors suggested that the English listeners were focusing on different cues than the Mandarin listeners, and that resulted in equivalent MMN amplitude to the hard and easy contrasts. This finding was unexpected given that lexical tone is reported to be acoustically robust (compared to segmental contrasts), especially for the easy tone contrast.

The results of lexical tone training studies using Thai lexical tone were also unexpected (Kaan, Wayland, Bao, & Barkley, 2007; Kaan, Barkley, Bao, & Wayland, 2008). These studies used one common standard (mid-level) and two deviant (high-rising and low-falling) syllables. For the low-falling deviant condition (compared to a mid-level standard), the English listeners showed larger MMN than the Mandarin listeners before training, but no group difference after training. For the high-rising deviant condition (compared to the mid-level standard), native Thai, Mandarin and English listeners showed no significant MMN before training; following training, only the native Thai speakers showed significant MMNs (Kaan et al., 2008). The small/absent MMN to the high-rising deviant condition is surprising given very high behavioral discrimination accuracy from all three groups (Kaan et al., 2007). However, a language group effect was observed for a late negativity that the authors called the LN. A left lateralized LN was observed for the high-rising deviant condition for the English and Chinese groups. It is likely that the LN is actually an MMN to the mid-high tone contrast because the high-rising and mid-level tones do not diverge significantly in F0 until 300 ms later than for the low-falling compared to midlevel tone. In addition, some researchers have suggested that lexical tone speakers pay more attention to the contour of the fundamental frequency, whereas the nonnative speakers paid more attention to the onset or offset of the tone contour (Gandour & Harshman, 1978; Gandour, 1983; Lee & Nusbaum, 1993). In this case, the MMN to a contour, or offset would be later, particularly in Thai and Mandarin speakers, since the contour shape and F0 offset could not be evaluated without the complete signal.

2.7. Late negativity as an index of auditory discrimination

A late negativity (LN), often co-occurring with the MMN, has been found in increasing number (Hardiman, & Barry, 2010; Datta, Shafer, Morr, Kurtzberg, & Schwartz, 2010; Kaan, Wayland, Bao, & Barkley, 2007; Korpilahti, Krause, Holopanien, & Lang, 2001; Shestakova, Huotilainem, Čeponiene & Cheour, 2003; Shafer, Morr, Datta, Kurtzberg, & Schwartz, 2005; Ortiz-Mantilla, Choudhury, Alvarez, & Benasich, 2010). The LN usually follows MMN and peaks between 300 and 500 ms (Čeponiene, Cheour, & Näätänen, 1998; Hill, McArthur, & Bishop, 2004; Shafer et al., 2005). The neural sources and conditions leading to the LN have not been fully specified, in part because studies often ignore this component, even when it is elicited. Based on the relatively consistent temporal relationship, but different topographic features between MMN and LN, Čeponienė, and colleagues proposed that LN (called Late Difference Negativity in their paper) is less likely to be linked to sensory aspects of processing, but more likely to reflect higher order, albeit preattentative, cognitive processing of sound change (Čeponienė, Lepisto, et al., 2004). Shafer et al. (2005) found that children with SLI showed absence of MMN but presence of LN, and proposed that both MMN and LN reflect a trace-comparison process, but LN may reflect discriminative processing that is independent of stored phonological representations. Other accounts about the nature of LN include that LN is an index of formation of phonological representation (Barry et al., 2009; Bishop, Hardiman & Barry, 2010), or LN reflects increased processing demand (Addis, et al., 2010), and can only be observed in complex acoustic stimuli and word stimuli, but not for stimuli with simple acoustic structure (Korpilahti et al., 2001). Alternatively, the LN could be a late-occurring MMN, which is late because the difficult acoustic discrimination takes longer to process. In either case, LN may be observed in non-native listeners in the absence of an earlier MMN. In the case of the contour tone differences, as suggested above, an LN may be observed because longer processing time is necessary to register the contour shape.

2.8. Hemispheric differences in lexical tone processing

There is some evidence that the left hemisphere is dominant in lexical tone processing. Two competing hypotheses have been proposed to explain the hemispheric dominance during lexical tone perception. The functional hypothesis emphasizes that hemispheric dominance is task-dependent and it reflects the psychological function of the task. Left-hemisphere dominance is hypothesized for pitch processing that has a larger linguistic load, such as lexical tone processing. In contrast, processing of pitch patterns used in intonation was argued to have a more peripheral linguistic function (or more related to emotional prosody) and be right-hemisphere dominance (van Lancker, 1980; Wong, 2002). Alternatively, the acoustic hypothesis (cue-dependent hypothesis) claims that regardless of psychological function, all long time-constant pitch patterns are lateralized to the right hemisphere (Klouda, Robin, Graff-Radford, & Cooper, 1988; Sergent, 1982; Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002). Evidence for the functional hypothesis includes some recent neuroimaging and behavioral (e.g., dichotic listening, Wang, Jongman, & Sereno, 2001) studies that suggest that the right hemisphere is more involved in auditory/acoustic processing, whereas the left hemisphere shows greater activation in response to phonemic processing (Gandour et al., 2000, 2002, 2004, Gandour, 2006; Hsieh, Gandour, Wong, & Hutchins, 2001; Wang, Jongman, & Sereno, 2001; Wang et al., 2003; Wong, Parsons, Martinez, & Diehl, 2004). The left hemisphere dominance for lexical tone processing can also be predicted based on the proposal that the left hemisphere is dominant for well-practices abstract-category processing and the right hemisphere is dominant for episodic exemplar processing (Dien, 2009).

Examination of the topography of the MMN has not revealed consistent hemispheric differences for lexical tone processing. One study found evidence of right hemisphere (RH) dominance in native Mandarin listeners for both lexical tone and intonation processing based on a source estimation analysis (Ren, Yang, & Li, 2009), whereas another study observed larger amplitudes for the MMN over left hemisphere sites for between-category deviants (Xi et al., 2010). Some studies have reported lack of hemispheric differences between Mandarin and English listeners (e.g., Chandrasekaran et al., 2007a). This lack of consensus on hemispheric dominance for lexical tone processing may be partially due to methodological differences. Note that effects in electrode sites measured at the scalp over the left hemisphere do not necessarily indicate a left hemisphere source activation, thus, the ERP measures of hemispheric differences do not necessarily match the hemispheric differences observed in neuroimaging or behavioral studies.

2.9. The Present Study

The present study examined Mandarin lexical tone processing in native and non-native listeners under two different memory-delay conditions (short and long ISIs). ERPs were obtained to lexical tone contrasts in native English listeners and native Mandarin listeners in two experimental conditions using different ISIs (Stimulus Onset Asynchrony: 900 ms and 3000 ms, resulting in ISIs of ca. 575 and 2675 ms, stimulus duration: 291-355 ms). We used multi-token natural bisyllabic speech in a multiple oddball paradigm in order to generate an ecologically more valid task, which was more likely to encourage phonemic processing. We hypothesized that the acoustic correlates of Mandarin lexical tone are quite robust, and can be easily discriminated when the ISI is short for both English and Mandarin listeners. However, under conditions of lengthened ISI (> 2.5 s) and in a context of more complex stimuli, language-specific long-term memory representations would be needed for robust discrimination. Thus, only Mandarin listeners are expected to show strong MMNs. We also hypothesized that the amplitude of the MMN would reflect the difficulty of discrimination and that the latency of the MMN peak would reveal the time period needed for discrimination. Finally, we further hypothesized that a late negativity (LN) would be observed and would reflect differences in tone contour shape, and would therefore be larger for Mandarin than English listeners. These experiments have particular theoretical significance for building a model of lexical tone speech perception.

An additional question of interest was whether there are hemispheric differences in lexical tone processing. We hypothesized that lexical tone processing involves more left hemisphere activation, and MMN can be used as an index of this process. We predicted that a larger MMN difference would be seen in the Mandarin groups at left hemisphere than right hemisphere sites, especially for the long ISI condition. The hemispheric contribution to the LN has rarely been discussed, but in general, the topography of LN is similar to that of MMN. Therefore, we predicted that the LN difference would be more salient at the left than right hemisphere sites for the native Mandarin listeners and that the asymmetry would be reversed or absent for the native English listeners.

Chapter 3. EXPERIMENTAL PROCEDURES

3.1. Participants

Thirty-one monolingual adult native English speakers (16 participants in the short ISI condition, and 15 participants in the long ISI condition, age range: 20-42 years) with no (or little) exposure to tone languages and 32 adult native Mandarin speakers (16 participants in each ISI condition, age range: 21-40years) were included in the study. A total of five participants were excluded due to incomplete participation, or excessive noise in the ERP signal. All participants passed a hearing screening, had no history of neurological impairment, and no extensive music training as studies have shown that musicians have an advantage in lexical tone perception tasks (Alexandar, Wong & Bradlow, 2004; Wong, Skoe, Russo, Dees, & Kraus, 2007). Specifically, participants had no formal music training in the past ten years, and did not play any instruments on a regular basis. The two language groups were closely matched with respect to age and years of formal education. The handedness questionnaire adapted from the Edinburgh handedness inventory (Oldfield, 1971) by Cohen (2008) was administered to all the participants. Table 1 presents participant information. The participants were paid 10 dollars per hour for their voluntary participation. The study was approved by the human subject research institutional review board at the City University of New York.

3.2. Stimuli

Natural speech sounds containing both phonetically relevant and phonetically irrelevant acoustic variations were produced by a female native speaker of Mandarin, and digitized at a sampling rate of 22,050 Hz. The stimuli consisted of three nonsense bisyllabic word types (/gupa/, /gipa/ and /gypa/) with three tone variations (T1, T2, and T3) on the first syllable only;

the second syllable was always "pa" with T1. The final set of 11 stimuli consisted of two tokens for the T1 deviant /gu1pa/, two tokens for the T2 deviant /gu2pa/, three tokens for the standard T3 /gu3pa/, two tokens of standard T3 /gi3pa/ and two tokens of standard /gy3pa/. These eleven tokens were selected from a larger set of recordings that was piloted extensively. /gi3pa/ and /gy3pa/ were not included in the current ERP analysis because these tokens have a dual function serving as tone standard and vowel deviant stimuli concurrently.

Phonetically irrelevant acoustic differences (e.g., overall amplitude, overall duration and voice onset time of the stop consonants among others) were equivalently distributed across each tone category, according to measurements from Praat 4.1 and Sound Forge (version 8). The average duration of the stimuli is 330.5 ms (range: 291 -355 ms, SD = 19.7), and the average intensity of the stimuli is 70.2 dB (range: 66.8 – 72.5 dB, SD = 1.9dB).

Table 2 presents the acoustic characteristics of the experimental stimuli, including relevant measurements of the target contrast differences (e.g., the 10-point normalized pitch values and the onset and offset pitch values for each vowel, the duration and pitch contour and average intensity of the transition part from the offset of the first syllable to the onset of the second syllable). The F0 of the second syllable (/pa1/) is also affected by the previous syllable in that /pa1/ has the highest F0 when following T1, and the lowest F0 when following T3. It is possible that listeners may detect this allophonic effect, and use the F0 height of the second syllable 'pa1' as an additional cue for decoding the tone type of the first syllable. If this is the case, then we expect to see a MMN to the second syllable as well as the first syllable, for the deviant tones.

3.3. Procedure

During the ERP and behavioral experiment, participants were seated in a sound- and electrically-shielded booth. Stimuli were presented through two loudspeakers placed above and in front of the participants. The total duration of the experiment lasted between 2.5 and 3.5 hours including preparation and multiple breaks.

3.3.1. Behavioral experiments: Tone discrimination and identification

The discrimination task was conducted on the same stimuli after the ERP session. A total of 33 trials including three practice trials were presented. Each trial consisted of a train of five stimuli (four standard followed by a deviant), and participants were asked to judge whether the final stimulus was the same or different from the previous four stimuli. This design was chosen to mimic the ERP design in the ISI between trials, but it required fewer total trials, and allowed time for a response. After the discrimination task, a three-alternative forced choice (3AFC) tone identification task was presented. In this identification task, one stimulus was presented at a time, and participants were asked to press a button (Button 1, 2 or 3) to decide whether the first syllable of the sound was Tone 1, Tone 2, or Tone 3. Six practice trials plus 30 test trials were presented. Participants had 3 seconds to provide a response.

3.3.2. ERP experiments

A passive oddball paradigm was used in which attention is directed away from the stimuli via a movie with the sound muted. Twenty blocks with 103 stimuli in each block were presented with an interblock interval of 20 seconds. Three syllabic types in which the first syllable was Tone 3 constituted the standard trials and had the following percentages: three tokens of /gu3pa/

occurred on 62.2%, two tokens of /gy3pa/ on 9.7% and two tokens of /gi3pa/ on 9.7% of the trials. The tone deviants were two tokens of Tone 1 /gu1pa/ on 9.7 % of the trials and two tokens of Tone 2 /gu2pa/ on 9.7% of the trials. A total of 200 deviant trials were delivered per category. This multi-deviant paradigm has been successfully used in previous studies (Muller, Widmann & Schröger, 2005; Nousak, Deacon, Ritter, & Vaughan, 1996; Sussman, Winkler, Kreuzer, Saher, Näätänen, & Ritter, 2002). Only the ERPs from the standard /gu3pa1/tokens were included in the analysis for the standard category (to match the deviant Tone 1 and Tone 2 on vowel /u/).

A stimulus onset asynchrony (SOA) of 900 ms (an average ISI of 575 ms, range 545-609 ms) was used for the short ISI condition, and of 3000 ms (average ISI of 2675 ms, range of 2645 to 2709 ms) for the long ISI condition.

3.3.3. ERP recording and offline analysis

The electroencephalogram (EEG) was sampled at 500 Hz (filtering bandwidth of 0.1-100 Hz) from 65 scalp sites using Geodesic sensor nets, referenced to the vertex electrode (Cz). Data were further filtered offline with a 15 Hz, low pass filter, and re-referenced using average reference. The EEG was time-locked to the onset of stimuli and was segmented offline into 1000 ms epochs including a 200 ms pre-stimulus baseline. After artifact rejection, the majority of participants had over 75% of trials included in the individual average data. Means and standard deviations (in parentheses) for the three stimulus types are: /gu1pa/: 177(15), /gu2pa/: 177(15) and /gu3pa/: 408(35). Mixed model ANOVAs for the deviant conditions and standard conditions showed that there were no group differences in terms of the number of trials included in the analyses (ps > 0.10), and no trial difference between the two deviant types although the standard conditions have more trials than the deviant conditions.

Chapter 4. Data analyses

4.1.Behavioral analysis

Based on the categorical nature of the results for both discrimination and identification tasks, we applied a mixed-effects logistic regression model on the raw response accuracy for each trial from each subject using language, ISI and tone type as the fixed effects and subject as a random effect. Mandarin, short ISI and Tone 3-Tone 1 contrast were used as the baseline category for the discrimination data, and Mandarin, short ISI and Tone 3 were used as the baseline category for the identification data. Factors were entered into the models in a hierarchical way. In Model 1, we only included language as a predictor, and then in Model 2, we added ISI, and in Model 3, we added tone type, then we entered all interactions one at a time into the models. We compared the overall fit of the models using ANOVAs to find the difference in the deviance statistics. The final model was selected based on the value of the deviance for the model and the significance value of its associated chi-square statistics ($p < 0.05$).

4.2.“Composite” Fz measures for MMN and LN

As found in previous studies, MMN was generally largest at Fz. We built a model of frontocentral activity from 16 sites as follows. We first calculated the Pearson’s correlation coefficients between each of the 64 channels and Fz for each stimulus condition and each language group. The 16 adjacent fronto-central sites (Electrodes 3,4,5,8,9,13,16,17,18,43,54,55, 57, 58, 62, and 65) showed correlations of greater than 0.87 to Fz. Averaging across these sites reduces the contribution of independent noise sources at each electrode site to the signal of interest (Parra et al., 2004), and it also reduces the inter-subject variations in the typography of the ERP to speech (Zevin, et al., 2010). Thus, the average of these 16 sites was used as the dependent measures (called “composite Fz”) (see Figure 1). Note that inferior-posterior sites (including the mastoids and sites near P7, P8) were highly negatively correlated with Fz (Pearson r values of -0.79 or above), which is consistent with the topography of MMN.

Composite Fz was used to examine whether the ERPs to the standard and deviant stimuli differed significantly. Mixed model ANOVAs were performed with stimulus (standard and deviant) and time (five levels between 100 and 350 ms for MMN and four levels between 350 and 550 ms for LN) as within-subject factors, and ISI (short and long) and language (English and Mandarin) as between-subject factors. Step-down ANOVAs were performed to follow up significant interactions.

To compare the ISI effect and deviant type effect, four-way ANOVAs with Language group (English, Mandarin), ISI (short, long) as between-subject variable, and deviant stimulus type (T2, T3) as within-subject variables were undertaken separately for the early time-interval (five intervals from 100-350 ms) and the later time intervals (four intervals from 350-550 ms). Amplitudes of the subtraction waves (deviant minus standard) were used as the dependent measure in this analysis.

4.3.MMN and LN amplitude and topography analyses

The dependent measure in these analyses was the amplitude of the subtraction waves (deviant minus standard). Peak amplitudes were used to compare topography of MMN and LN responses. To derive the peak amplitude at the relevant sites, a 100-ms time window around the group mean peak latency (+/- 50 ms on each side) was selected and divided into five consecutive time intervals of 20 ms each. The maximum amplitudes at each of the six sites (frontal: F3, F4; central:C3, C4; inferior: LM, RM) were chosen from the five average values over the 100 ms time window (five 20-ms time intervals). Our focus was to examine right versus left hemispheric dominance; therefore, we used a frontal-mastoid model (two frontal sites: F3, F4 and two mastoids: LM, RM) and a centro-mastoid model (two central sites: C3, C4 and two mastoids: LM, RM) in the mixed model ANOVAs with site (inferior/superior) by hemisphere (left/right) as within-group factors and language and ISI as between-group factors. All the subtraction waveforms for each participant were first normalized to a mean of zero and standard deviation of one for these analyses.

We also examined whether each group resembled their counterpart language or ISI group in topography using Pearson's correlation coefficients across all 65 channels (related to Global Dissimilarity Index by $r = 1\text{-}(\mathrm{GDI}^2)/2$) (Shafer et al., 2011; Skrandies, 1990). Three consecutive time windows between 150 and 300 ms, 400 and 550 ms for MMN and LN were chosen, respectively, for the analyses.

For all ANOVAs, degrees of freedom were adjusted using Greenhouse-Geisser correction (ε) for comparisons with more than one degree of freedom in the numerator and were reported as corrected p-values. The uncorrected degrees of freedom, F-values, corrected p-values and the epsilon (ε) values if applicable were reported.

Chapter 5. RESULTS

5.1.Behavioral discrimination and identification results

5.1.1.Behavioral discrimination results

Both Mandarin and English listeners performed the discrimination task with greater than chance level accuracy under the short and long ISI conditions. Table 3 describes the average and standard deviations of discrimination accuracy. Figure 2 displays the hit and false alarm rates for each group/condition. Two technical error cases (Subject 407 from the English long ISI group and 119 from the Mandarin short ISI group) were removed from the model-fitting process because both subjects reported that they failed to follow instructions and pressed the buttons the opposite way during the task. In the generalized linear model for logistic regression, the best-fit model included three significant predictors and one significant interaction. The significant fixed effects included tone type effect (beta estimate = -1.68, beta SE = 0.22, $z = -7.68$, $p < 0.001$) with T3/T2 contrast being harder to discriminate than T3/T1, and language effect (beta estimate = 1.93, beta SE = 0.43, $z = 4.57$, $p < 0.001$) with lower accuracy in the English group. A language by ISI interaction was also significant (beta estimate = -1.21, beta SE = 0.56, $z = - 2.18$, $p = 0.03$) with lower accuracy in the English long ISI condition.

5.1.2.Behavioral identification results

Table 4 provides a summary of average and standard deviation of the identification accuracy. There was a striking difference in the response patterns of the two language groups. The Mandarin listeners identified all three tone types well-above chance (> 33%) while the English listeners were at or below chance level for T1, and much lower accuracy for T2 and T3 than the Mandarin groups. The results from fitting a mixed-effects logistic regression model revealed that the only significant predictor (fixed effect) is language, in which the English group had lower accuracy than the Mandarin group (beta estimate = - 2.51, beta SE = 0.36, z = - 7.06, p < 0.001).

5.2. ERP Results

Figure 3 displays the grand mean ERPs to the standard and deviant stimulus and subtraction waveforms for the composite Fz.

5.2.1. Standard Tone 3 versus deviant Tone 1: Fronto-central, 100-350 ms:

Results from the ANOVAs are listed in Table 5. In the 100-350 ms time window, there was a main effect of stimulus type (p = 0.001), and interactions of stimulus by language (p = 0.02), time by stimulus by ISI interaction (p = 0.01), and time by stimulus by language (p = 0.01). The main effect indicated that the deviant (T1) was more negative than the standard (T3). Post hoc tests following up the stimulus by language interaction revealed that this pattern was significant only for the Mandarin group. Because we were interested in whether an MMN was present for each of the four language/ISI subgroups, step-down ANOVAs were performed for each group with stimulus and time as factors. Both Mandarin groups showed significant main effects of condition (ps < 0.05). The English short ISI group showed a significant time by stimulus interaction (p < 0.001), with post-hoc tests revealing a significant difference for the150-200 ms time window. No significant differences were observed for the English long ISI group for the 100-350 ms time window. The MMN amplitudes derived from the subtraction (deviant minus standard condition) are presented in Table 6.

5.2.2. Standard Tone 3 versus deviant Tone 2: Frontocentral, 100-350 ms

The four-way mixed model ANOVAs are shown in Table 7 and revealed a main effect of stimulus type ($p < 0.001$) and a stimulus by time interaction ($p < 0.001$). Post hoc tests indicated greater negativity of the deviant than the standard for the time points between 150-350 ms. Two-way ANOVAs on each of the four groups revealed a main effect of stimulus type ($ps < 0.01$) and stimulus by time interactions ($p < 0.01$) for the Mandarin short and long ISI subgroups. Post hoc tests revealed greater negativity of the deviant from 200-350 ms for the short ISI and from 150-350 ms for the long ISI condition. The English short ISI group, showed a stimulus main effect ($p < 0.001$) and an interaction of stimulus and time ($p < 0.05$), with greater negativity of the deviant from 200-350 ms. No significant differences were found in this interval for the English group in the long ISI condition. Table 8 presents the MMN amplitude across time for T3/T2 contrast at the Fz composite site.

5.2.3. Late negativity at composite Fz: Standard Tone 3 versus deviant Tone 1

The ANOVA results are displayed in Table 9. Between 350 and 550 ms, a main effect of stimulus type ($p < 0.001$), a time by stimulus interaction ($p < 0.001$) and a stimulus by time by ISI interaction ($p = 0.03$) were significant. Post hoc tests showed that from 350-400 ms, the deviant was more positive than the standard, whereas from 400-550 ms the deviant was more negative than the standard.

Analyses focusing on the groups separately showed a stimulus main effect for the Mandarin long ISI group ($p = 0.02$), and interactions of condition and time for all four groups ($p < 0.001$). Post hoc test revealed significant negativity of the deviant between 450-550 ms for the short ISI and between 400 and 500 ms for the long ISI for Mandarin groups; for the English short ISI group, the deviant was more positive than the standard between 350 and 400 ms, but more negative than the deviant between 450 and 500 ms. For the English long ISI group, the deviant was more negative than the standard between 400 and 500 ms. Table 10 shows the mean and SD of the subtraction waveforms for the interval of 350-550 ms for each group.

5.2.4. Late negativity (LN) at composite Fz: Standard Tone 3 versus deviant Tone 2

Table 12 displays means and standard deviations for the subtraction (deviant – standard) in time windows of 350-550 ms. For T3/T2 contrast, the four-way ANOVA (shown in Table 11) revealed a main effect of stimulus type ($p = 0.04$) with the deviant (T2) being more negative than the standard (T3), and an interaction of stimulus type and time ($p < 0.001$). Post hoc tests following the two-way interaction showed that the deviant is more negative than the standard between 400 and 550 ms.

ANOVAs carried out for each group separately revealed significant stimulus by time interactions for all groups ($ps < 0.05$) except the Mandarin short ISI group. For the Mandarin long ISI group, the deviant is more negative than the standard between 400 and 500 ms. For the English short ISI group, greater positivity of the deviant between 350 and 400 ms, followed by greater negativity between 450 and 550 ms was shown in post-hoc tests. For the English long ISI group, post hoc tests showed a significant difference between the deviant and standard only between 350 and 400 ms, with the deviant being more positive than the standard.

In summary, increased negativity in the early time interval (100-350 ms) was observed for both language groups and both deviant conditions for the short ISI condition. In contrast, for the long ISI condition, only the Mandarin group showed negativity in this early interval. In the later time window (400-550 ms), the T1/T3 contrast showed greater negativity for the deviant than standard for all groups except the Mandarin short ISI group. In contrast, for the English short and long ISI groups, the response to the T2/T3 contrast was more positive for the deviant than the standard between 350 to 400 ms, and for the response to the T1/T3 was more negative between 450 and 550 ms.

5.3. ISI and deviant type comparisons using composite Fz in MMN and LN time windows

5.3.1. Time window 100- 350 ms

ANOVAs using the subtraction waves (deviant minus standard) revealed main effects of language (F (1, 59) = 6.177, p = 0.01), and stimulus type (F (1, 59) = 9.590, p = 0.002), and interactions of time by stimulus type (F (4, 236) = 14.1, $p < 0.0001$, $\varepsilon = 0.87$) and time by deviant stimulus type by ISI (F (4, 236)= 5.544, $p < .001$, $\varepsilon = 0.87$). Post hoc tests for the main effects showed that the Mandarin listeners had larger negativity than the English listeners, and that the T2 deviant generated a larger negativity than the T1 deviant. Post hoc tests following time by deviant type interaction revealed that T2 was more negative than T1 between 200 and 350 ms. Post hoc tests following the three-way interactions indicated that T2/T3 is more negative than T1/T3 from 200 to 250 ms in the short ISI condition, and from 250 to 350 in the long ISI condition.

5.3.2. Time window350- 550 ms

The results from the ANOVA using the subtraction waves showed main effects of deviant type (F (1,59) = 7.83, p = 0.007), significant interactions of time by ISI (F (3, 177) = 3.66, p = 0.02, $\varepsilon = 0.77$), and time by deviant type (F (3, 177) = 3.382, p = 0.04, $\varepsilon = 0.56$). Post hoc tests showed that the subtraction wave in the long ISI condition was more negative than that in the short ISI condition between 350 and 500 ms, and the subtraction wave was larger for T3/T1 than T3/T2 condition. Post-hoc tests did not show significant differences in amplitude for the different time windows. No main effect or interaction involved the language variable.

5.4. Topographic features of MMN and LN

Figure 4 (for Tone 1) and Figure 5 (for Tone 2) depicts the subtraction waves at three frontal (F3, Fz and F4) and three central sites (C3, Cz and C4) where MMN has been most often reported. The two mastoids (LM and RM) and Oz illustrate inversion in polarity in relation to the frontocentral sites.

5.4.1. MMN topography

The five-way ANOVA (language by ISI by deviant type by hemisphere by site) results showed a main effect of site ($F(1,59) = 152.1$, $p < 0.001$), an interaction of deviant type by site ($F(1,59) = 9.67$, $p = 0.003$), and a five-way interaction of all variables ($F(1,59) = 6.22$, $p = 0.015$). Step-down analyses were done for each ISI condition and each deviant type. Main effect of site was present in all the analysis, in which the superior sites were more negative than the mastoid sites.

Under the short ISI condition for the fronto-mastoid model, main effects of site were found for both T3/T1 and T3/T2 contrast ($F(1,30) = 46.7$, $p < 0.0001$ for T3/T1, and $F(1,30) = 185.8$, $p < 0.0001$) for T3/T2. A main effect of language for the T3/T2 was also present with the Mandarin group showing more positive responses ($F(1,30) = 5.11$, $p = 0.03$). Examination of the figures suggested that the greater positivity of the mastoids in the Mandarin group led to this result. For the centro-mastoid model, site main effects were found for both T3/T1 and T3/T2 contrast ($F (1,30) = 78.1$, $p < 0.0001$ for T3/T1, and $F (1,30) = 176.5$, $p < 0.0001$) for T3/T2. A language main effect for the T3/T2 was also present with the Mandarin group, showing more positive responses($F (1,30) = 7.90$, $p = 0.008$). A site by hemisphere interaction was found ($F (1,30) = 4.21$, $p = 0.05$), but post hoc tests did not confirm where the difference occurred.

Under the long ISI condition, for the fronto-mastoid model, the T3/T1 contrast showed a main effect of site ($F(1, 29) = 18.5$, $p < 0.001$), and a site by language interaction ($F(1, 29) = 5.085$, $p = .032$), however, pairwise post hoc tests between groups for superior/inferior site did not show significant differences. A main effect of site was present for T3/T2 contrast ($F(1, 29) = 45.6$, $p < 0.001$). For the centro-mastoid model, a main effect of site was present for both T3/T1 and T3/T2 (T3/T1 contrast: ($F(1, 29) = 36.0$, $p < 0.001$; T3/T1 contrast: ($F(1, 29) = 117.0$, $p < 0.001$).

In summary, for the short ISI condition, a main effect of language was observed for the T2/T3 contrast for both the fronto-mastoid and centro-mastoid model. The Mandarin group showed relatively more positive responses than the English listeners, suggesting more contribution from the mastoid. A site by hemisphere interaction was present for the centro-mastoid model under the short ISI T3/T2 condition. A site by language interaction was present for the T3/T1 contrast in the fronto-mastoid model for the long ISI condition.

5.4.2. LN topography

Under the short ISI condition, main effects of site were significant for the fronto-mastoid models (short ISI T3/T1: $F(1,30) = 80.6$, $p < 0.001$; short ISI T3/T2: $F(1,30) = 74.8$, $p < 0.001$) and centro-mastoid model (short ISI T3/T1: $F(1,30) = 99.7$, $p < 0.001$; short ISI T3/T2: $F(1,30) = 42.7$, $p < 0.001$). There was a three-way interaction among site, hemisphere and language (F(1,30)

= 4.97, $p = 0.03$) for T3/T1 contrast of the centro-mastoid model, but post hoc tests did not reveal any specific difference.

Figure 5 showed a striking difference between the Mandarin and English group at the two mastoid sites for the short ISI T3/T2 conditions. The Mandarin group has much larger positivity than the English group at the two mastoid sites. This difference was confirmed statistically, as a language main effect was present in the short ISI fronto-mastoid model for T3/T2 contrast ($F(1,30) = 11.6$, $p = 0.002$) with the Mandarin group showing larger positivity.

Under the long ISI condition, main effects of site were significant for both the fronto-mastoid models (long ISI T3/T1: $F(1,29) = 39.3$, $p < 0.001$; long ISI T3/T2: $F(1,29) = 25.4$, $p < 0.001$) and centro-mastoid model (long ISI T3/T1: $F(1,29) = 62.4$, $p < 0.001$; long ISI T3/T2: $F(1,29) = 57.4$, $p < 0.001$). The only significant interaction was hemisphere by language for T3/T2 contrast within the centro-mastoid model ($F(1,29) = 8.32$, $p = 0.007$). The Mandarin group showed greater negativity over the left than right hemisphere, whereas the English group showed greater negativity over the right than left hemisphere.

5.4.3. Pearson's *r* correlation analysis on the topography differences

Table 13 summarizes the result from the Pearson's *r* correlation analysis on the topography differences across language and ISI groups. Comparisons within language groups in the early and late time windows revealed strong correlations ($df = 63$, $r > 0.316$, $p < 0.01$) between the two English groups across all intervals for T1/T3, and for T2/T3 significant correlation only for the early intervals (200-300 ms).The two Mandarin groups showed high correlations ($df = 63$, $r > 0.40$, $p < 0.001$) for both tone contrasts from 150-250 ms, but only for T1/T3 for the later intervals (400-550 ms). Comparison for the two language groups receiving the short ISI condition reveals strong correlations ($r > 0.40$) from 150-200 ms, but no positive correlations for the later time intervals. For the two language groups in the long ISI condition, T1/T3 topography was strongly correlated ($r > 0.40$) from 150-300 ms and from 450-500 ms, and T2/T3 was strongly correlated from 200-250 ms, and from 450-550 ms ($r = 0.65$). The largest dissimilarity was observed for the two short ISI groups for the later time windows. A strong negative correlation was observed for T2/T3 from 400-450 ms, where the Mandarin listeners showed large positivity at the mastoid/inferior sites and the English listeners showed either larger negative response (at site OZ) or very small (if any) positivity at the mastoid sites. In addition, no correlation for T1/T3 was found for this time window between the Mandarin and English short ISI groups.

In summary, the Mandarin group had larger responses than the English group at the mastoid sites for the early intervals under both short and long ISIs for the T2/T3 deviant condition, and for the late time intervals for the short ISI T2/T3 deviant condition. A larger left than right hemisphere late negativity was observed for the Mandarin group in the long ISI T2/T3 deviant condition. Comparing topographic differences using all 65 sites, we found the greatest topographic differences lie between 400 and 450 ms time interval for T2 across ISI and language conditions. T1/T3 generally showed higher correlates between groups and between ISI conditions than the T2/T3 contrast.

Chapter 6. DISCUSSION

6.1. Summary of main findings

The current study was designed to extend our understanding of the neural correlates of lexical tone processing by introducing greater stimulus complexity and by allowing for trace decay. Comparing the neural discrimination across different ISI conditions and across native and non-native listeners allowed us to examine the nature of the long-term memory representations for lexical tone in Mandarin listeners. As predicted, the Mandarin groups showed similar or larger amplitude mismatch responses (presumably MMN and LN) in the long ISI conditions compared to the short ISI condition. In contrast, the English groups showed differences across the ISI conditions. Specifically, at the long ISI there was no difference between the standard and deviant in the early time interval (which we will call the MMN). For the late time interval, a difference (which we will call the LN) was present only for the T3/T1 contrast. With regards to topography, the Mandarin groups showed larger MMN and LN responses than the English groups from the mastoid sites for the T2/T3 deviant contrast. In addition, the Mandarin groups revealed more left-lateralized responses than the English listeners. The topographic differences were greater for the T2/T3 conditions within the LN time window. These results will be examined in greater detail below.

6.2. MMN and auditory sensory memory duration

We had predicted that native-language experience would allow robust brain discriminative responses for lexical tone in the face of decay of the immediate memory trace. Our findings are consistent with this claim. The Mandarin listeners showed robust negativity in the long ISI conditions while the English listeners showed no negativity under this condition. The negativity observed in the early time interval (150-350 ms) is consistent in timing and topography with MMN reported in previous studies. Cross-linguistic studies of lexical tone discrimination have shown better than chance discrimination of lexical tones in simple discrimination tasks by non-native listeners (e.g., Gandour, Wong, Hutchins,1998; Burnham et al., 1996). The behavioral discrimination result from our short ISI experiment corroborates the previous findings. However, lexical tone training studies indicate that lexical tone is not easy to master for listeners of languages that do not have tone (e.g., Wayland & Li, 2008), especially for some difficult tones such as Mandarin T2 and T3 (Wang, Spence, Jongman, & Sereno, 1999). Indeed, in the present study, the behavioral results of the participants revealed poorer discrimination of the T2/T3 than the T1/T3 contrast, and generally poor identification of the three tones in a forced-choice categorization task. In natural language processing, identification of the lexical tone category often occurs without the benefit of an immediately preceding model. Furthermore, L2 speech perception is worse under conditions of increasing task difficulty or increased noise (Strange, 2011). Native language listeners, however, show lesser decline in performance under taxing conditions, supporting the claim that they are highly automatic at extracting the relevant information. Strange (2011) called this first language (L1) automatic speech processing as selective perceptual routines (SPRs). Our findings add to this model by demonstrating that the SPRs to recover lexical tone in native tone language speakers are stored in long-term memory and can be accessed without focused attention, similar to segmental information (e.g., Hisagi, et al., 2010).

Studies using oddball paradigms designed to elicit MMN have manipulated ISI to examine the duration of sensory memory in adults and children and have found that, in general, longer ISI leads to less prominent or no MMN (for pure tone, Barry et al., 2008; Gomes et al., 1999;

Grossheinrich, et al., 2010; Pekkonen et al., 1996; for speech, Čeponiene, et al., 1999). These studies provided ERP evidence that the duration of the auditory sensory memory store is limited. The majority of these studies used pure auditory tones of contrasting frequencies as stimuli, and thus, the results of ISI manipulations primarily reflect auditory sensory memory decay because it is less likely that long-term representations of pure tones are stored without relevant training (Hedger , Heald, & Nusbaum , 2013).

Only one ERP study has used an ISI longer than 1.5 s in experiments using speech contrasts. Čeponiene and colleagues used a consonant contrast in bisyllabic nonwords (/baga/ and /baka/) and found that children with good phonological memory showed a smaller MMN for a long (2 s) ISI compared to a short (350 ms) ISI (Čeponiene, et al., 1999). In another pair of studies, which was not designed to directly examine ISI differences, typically developing children showed an earlier MMNs to a long 250-ms vowel contrast ([ɛ] vs. [I]) presented using a short ISI of 350 ms compared to a short 50 ms-version of the same contrasts presented with a longer ISI of 550 ms (Shafer, et al., 2005; Datta et al., 2010). Thus, the stimulus duration and/or ISI could have led to this pattern of findings. These three studies were designed to examine individuals who are likely to have poor working memory. The study by Čeponiene and colleagues (1999) showed no MMN to the speech contrasts for children with poor phonological memory (assessed using a non-word repetition task). In addition, both children with good and poor phonological memory showed robust MMNs to a tone contrast at the short and long ISIs in this study (Čeponiene et al., 1999). In the studies by Shafer and colleagues, many of the children with specific language impairment (SLI) did not exhibit a robust MMN to the short vowel/long ISI condition, but almost all of the children with SLI in the second study showed robust MMN to the long vowel/short ISI condition (Datta, et al., 2010). In addition, a study by Barry et al. (2008) found that parents of children with language impairment showed greater attenuation of the MMN to pure tone contrast (1000 Hz versus 1200 Hz) at the longer ISI, than seen for adults with no family members with SLI. Taken together, these studies suggest that long-term memory representations of speech allow for more robust discrimination of speech contrasts in participants with good language skills.

The present findings provide a better understanding of the effect of language-specific speech information on sensory memory. Specifically, robust representation of the lexical tone information is maintained as a memory trace for 2.5 s, without focused attention by native listeners. Absence of a MMN in English listeners in the long ISI condition suggested that they did not have long-term memory representations of the F0 patterns that were sufficient to allow for discrimination. English listeners' long-term representations of F0 may encode information that is necessary for lexical stress or for sentence level prosody, but these representations would likely weight F0 information in a manner that is insufficient to support lexical tone perception.

The relationship between ISI and duration of sensory memory remains to be further explored given that findings across studies do not converge on how long an auditory memory trace lasts (e.g., Mäntysalo & Näätänen, 1987: 4s; Böttcher-Gandor & Ullsperger, 1992: 10 s; Sam et al., 1993: 9 s; Winkler et al., 2002: 30 s). Our study shows that experience with a speech sound influences the apparent time course of sensory memory delay. Presumably, the time-course of sensory memory decay is invariant, and what leads to apparent differences in this decay rate is the nature of long-term memory representations that modulate a specific sensory memory trace. Thus, experience with non-speech information should also influence this time course. In addition, the degree of difference will influence the apparent decay rate. A very large acoustic difference between two stimuli (e.g., 1000-1100) may give the appearance of longer maintenance of information in memory. The use of finer acoustic differences will provide more specific information regarding the nature and time-course of sensory memory decay.

In sum, the more parsimonious hypothesis is that the decay rate of sensory memory is the same across all acoustic stimuli and what differs is the amount of information in the stimulus and whether the listeners has repeated experience with the sounds, leading to long-term representations that can be used to update sensory memory.

6.3. MMN and higher level phonemic processing versus lower level acoustic/phonetic processing

In this study, no early negativity consistent with the MMN was observed for the English group under the long ISI condition. In contrast, both Mandarin groups and the English short ISI group showed presence of MMN in this early time interval. Most ERP studies on vowel and consonant discrimination have used an ISI of less than 1000 ms, and often reported that native language experience generates a larger MMN (Winkler et al. 1999, Shafer et al., 2004, see Näätänen, Paavilainen, Rinne, & Alho, 2007 for a review). Based upon the ERP and behavioral studies on speech perception (e.g., Werker & Logan, 1985; Pisoni, 1973; Burnham et al., 1996) and Cowan’s memory model, under a short ISI, the neural responses reflect the acoustic/phonetic level of processing; but processing at short ISIs can also be modulated by phonemic information. Longer ISIs will lead to more dependence on the phonemic representations in the long-term memory. Specifically, discrimination can rely on resolution of acoustic-phonetic information, but also can make use of phonological (long-term) information. For segmental contrasts (vowels and consonants), cross-language differences may have been observed at these short ISIs because the contrasts are difficult and long-term representations increase the saliency of language-specific cues (see Hisagi, et al., 2010). However, for syllable-level contrasts, such as F0, the acoustic-phonetic information may be sufficient to allow for robust, pre-attentive discrimination at short ISIs, even for non-native listeners. Thus, cross-linguistic differences for lexical tone are less pronounced even with some decay over a 1000 ms ISI. Even at the longer ISI, nine out of the 15 English participants showed robust MMNs, albeit significantly smaller in amplitude than for native listeners. Phonological priming research suggested that stored lexical representation in the subject's long-term memory can be accessed within 500 ms (see Hamburger & Slowiaczek, 1996, for a review). Long-term memory representations are likely to be available irrespective of ISI; however, their influence becomes more apparent under longer ISIs that results in considerable sensory memory decay.

In this study we used several strategies, in addition to increasing ISI, to preclude the possibility of using acoustic-phonetic processing alone. By using multiple oddballs and more than one token per stimulus type, the number of repeating identical tokens was greatly reduced, which would force greater reliance on more abstract patterns; also, the use of natural speech and bisyllabic nonwords resulted in an ecologically more valid task and a context that is more likely to preclude reliance on acoustic/phonetic cues alone. Such a paradigm taps into phonemic processing that is based on one's long-term language experience to a greater extent than the conventional MMN paradigm, where a single oddball deviant token and a short ISI are used. The finding of no MMN peak amplitude reduction in the long ISI Mandarin group suggests that, with the support from language-specific long-term memory representation, sensory memory for lexical tone is sustained beyond 2.5 s. Lack of MMN in the long ISI English group provided evidence for that the sensory memory trace for F0 that is based on acoustic/phonetic information decays to a point where it is insufficient to allow for robust discrimination, without support from a long-term representation. This sensory memory trace for lexical tone patterns decays within 3 s, which is much shorter than found for studies using pure auditory tones. By comparing lexical tone discrimination using a short versus long ISI, we are able to glean the differential contribution of long-term memory representation from that of echoic sensory memory trace responses in speech discrimination. Further study is needed to examine the nature of long-term memory representations of English listeners. It is likely that English listeners maintain information that allows discrimination of English suprasegmental patterns such as sentence prosody and lexical stress, but that this information is insufficient to allow lexical tone discrimination.

6.4. MMN responses to different lexical tone contrast pairs

Our results show that under the short ISI condition, the English and Mandarin language groups differed the most for the more difficult T3/T2 contrast at the mastoid sites, and under the long ISI condition, the Mandarin group showed robust MMN responses for both deviant conditions while clear MMN was observed in the English listeners for T3/T1 only. This result is consistent with other behavioral data (Gandour, 1978; Wang et al., 1999) and our behavioral data showing that there is a striking difference in the English groups for discriminating the two tone contrasts. The English groups were much poorer in discriminating T3/T2 than T3/T1, while Mandarin listeners showed discrimination accuracy of over 90% for both T3/T2 and T3/T1.According to Burnham (1986), and expanded by Strange (2011), contrast salience depends upon the size of acoustic change, as well as listener’s experience with the phonetic contrast. Behavioral performance in the present study suggests that the T3/T2 might be a “fragile” contrast for English listeners, but not necessarily so for Mandarin listeners. However, this result is at odds with the findings of the Chandrasekaran et al. (2007a) study. In Chandrasekaran et al. (2007a), the two language groups differed under the “easy” T3/T1 condition, but not the “hard” T3/T2 condition. MMNs to T1/T3 and T2/T3 for English listeners were equally small and comparable to the Chinese T3/T2 condition. Chandrasekaran and colleagues used a rather short ISI similar to our short ISI condition. Our results from the short ISI condition (SOA = 900 ms) show that there is no language group difference at the six frontal and central sites for these contrasts but there was a difference at the mastoid sites. The mastoid positivity is the opposite pole of the negativity found at frontocentral sites. It is not clear whether there was a language effect at the inferior sites in the Chandrasekaran study for the hard T2/T3 contrast since no results including the inferior sites were presented. In addition, they used linked mastoid as a reference, which will minimize the amplitude of the response at nearby inferior sites whereas this study used average reference. It is possible that the MMN amplitudes to the tone contrasts observed at frontocentral sites serves to index the lower level acoustic-phonetic processing, but the mastoid sites more clearly measure differences at the phonemic level. Thus, larger MMN at frontocentral sites for T3/T1 than for T3/T2 in English listeners reflects larger acoustic differences in this pair. In the Chandrasekaran study, larger T3/T1 MMN was found for the Mandarin compared to the English listeners, suggesting that this can be attributed to language experience effects. Perhaps this effect was the result of the linked mastoid reference. Also, it is important to keep in mind that language experience differences were found in the later time interval in our study and that it is possible that the LN reflects the change detection process for the contour shape. This will be discussed further in the next sections.

6.5. MMN and topography in lexical tone processing

In the current study, Mandarin listeners showed larger responses than the English listeners at the mastoid sites, especially for the T3/T2 contrast. In addition, the mastoid responses were larger over the left than right site for Mandarin listeners, but over the right compared to left sites for English listeners, particularly in the long ISI condition for the late (LN) time window. The left hemisphere difference is consistent with the account of lexical tone hemisphere dominance (Gandour et al., 2000; Gandour et al., 2002; Hsieh et al., 2001; Klein et al., 2001; Wang et al., 2001; Wang et al., 2003). Neural imaging studies and some behavioral studies suggested larger left hemisphere involvement in lexical tone processing (for Mandarin: Hsieh et al., 2001; Klein et al., 2001; Wang et al., 2003; for other tone languages: Felder et al.,2009; Gandour et al., 2000; Gandour et al., 2002; dichotic listening: Wang et al., 2001). However, neurophysiological studies on lexical tone have not revealed clear hemispheric dominance. Many studies examined only Fz, or two or three of the frontal sites (e.g., Chandrasekaran, et al., 2007; Kaan, et al., 2007; Xi et al., 2010) and no studies have examined the mastoid sites. Ren and colleagues performed a low-resolution electromagnetic tomography (LORETA) analysis and claimed right hemisphere (RH) dominance in native Mandarin listeners for both lexical tone and intonation processing (Ren et al., 2009). Xi et al (2010) also reported marginally larger MMNs at the right than the left frontocentral scalp sites for both within- and across-category deviant conditions, but the difference between the across-category deviant condition and the within-category deviant condition was evidenced in the left hemisphere (F3) only. The authors took this hemisphere by condition interaction as evidence supporting the dichotomous view of right hemisphere processing for lower-level acoustic processing and left hemisphere for higher level categorical processing (Gandour et al., 2004; Tong, Gandour, Talavage, et al., 2005; Zattorre & Gandour, 2008). However, Gandour and colleagues also pointed out that this dichotomous view has to be scrutinized because brain laterality may depend on interactions among language experience, modality features and task features. Just because the deviant and standard are categorical contrasts, it does not guarantee phonemic processing, especially when it is presented in a passive auditory paradigm with a short ISI.

It is also important to keep in mind that the left mastoid asymmetry could indicate triggering of the left hemisphere language network (Hickok & Poeppel, 2007) rather than tone discrimination per se. The stimuli used in the current study were word-like and consistent with Mandarin phonology. Thus, it is possible that triggered lexical access processes for native listeners.

6.6. LN and lexical tone processing

In the present study, LN was observed for all groups and conditions with the exception of the long ISI, hard contrast for the English listeners. Mandarin listeners generally showed larger LN responses than the English groups across both short and long ISI conditions. Specifically, for the short ISI condition the Mandarin group showed a larger LN (reflected as positivity at the inferior sites) for the T3/T2 contrast than that observed in the English group. In addition, Mandarin listeners showed enhanced LN at the left frontocentral sites for the long ISI condition relative to the short ISI condition and relative to the English listeners in the long ISI condition. This result diverges from findings of some previous studies in which a larger LN is sometimes seen in the less experienced group (e.g., family of specific language impairment or SLI in Addis et al., 2010; late bilingual learners in Ortiz-Mantilla et al., 2010) or impaired listeners (e.g., children with SLI in Shafer et al., 2005), although, in Shafer, et al. (2005), both typical and SLI children showed robust LN.

Studies have suggested that the LN, as well as MMN, indexes discrimination regardless of the specific functional account. These suggestions include re-orienting (Shestakova et al, 2003; Schröger & Wolff, 1998; Ortiz-Mantilla et al., 2010), and further processing that is independent of phonological representation (Shafer et al., 2005). LN may reflect a different process than the MMN, or in some cases, it may be a late MMN.

It is possible that the mechanism indexed by an LN in a long ISI condition differs from those indexed by LN in a short ISI listening context, and that the long ISI context automatically recruited more higher-level cognitive resources. An additional possibility is that the LN is the MMN to the tone contour, which necessarily is later in time because the difference cannot be computed until the end of the stimulus. Even if this is the case, however, the enhancement of negativity needs to be explained. We also found positivity in the early portion of this 300-500 ms interval for the short ISI conditions. It is possible that this positivity is a P3a orienting response that partially overlaps with the LN (Gumenyuk, Korzyukov, Escera, et al., 2005). In this case the apparent enhancement of the LN at long ISIs may be due to absence of the overlapping P3a.

There is only one study other than ours that has examined the LN responses to lexical tone deviance (Kaan et al., 2007). Unfortunately, the results presented in the training study by Kaan et al. (2007) are difficult to interpret. For example, Mandarin listeners showed decreased late negativity to an untrained tone, and there was no change in response to a trained tone contrast. In their study, a high-rising tone / mid tone contrast did not generate an MMN in Thai, Chinese or English listeners, but elicited a late negativity. In contrast, a low-tone deviant generated MMN, but no late negativity. Examining the tone contour of stimuli in Kaan et al. (2007), it appears that the significant acoustic difference between the standard and the high-rising deviant occurs almost 300 ms after stimulus onset, thus it is possible what the authors are calling the LN for the high-rising deviant condition is actually the MMN to the contour change.

The current study used bisyllabic stimuli, and the second syllable is always /pa1/ with Tone 1 (a high level tone). However, due to coarticulation effects, the F0 contour for /pa1/ is affected by the tone status of the preceding syllable and had the highest values when preceded by T1 context (e.g., gu1pa), the lowest values for T3 context (e.g., gu3pa), and the intermediate values when preceded by T2 (e.g., gu2pa). It is possible that an F0 difference also contributes to the negativity we observed in the 350-550 ms time window. The pitch difference on the second syllable /pa1/ is a within-category distinction for both Mandarin and English listeners, therefore, it is more likely to generate similar discrimination responses from the two language groups.

In summary, the LN seen to lexical tone contrasts may really reflect a later MMN-type process to the contour shape. LN is probably more likely to be observed to complex speech contrasts than simple, pure tones because speech contrasts (and in particular, natural speech) can be discriminated in more than one way. Specifically, spectral, duration and intensity cues may all contribute to discriminating a natural speech contrast, and the time course of discrimination is likely to differ for each of these dependent on degree of difference and the timing of the difference.

Chapter 7. CONCLUSION

This is the first study that examined the neural mechanisms involved in the decay of echoic sensory memory for phonemic, lexical tone contrast. Our study illustrates that the sensory memory trace elicited by the suprasegmental F0 contrast decays within 3 s to an extent that will not support lexical-tone discrimination, without the support of language-specific, long-term memory representations. In addition, the timing of the neural discriminative response is dependent on the nature of the stimulus difference. Contour tone patterns need a longer time-course for discrimination, as reflected by language group differences in the LN discriminative response. The finding that more left lateralized responses seen at the mastoid site for the Mandarin groups and more bilateral responses for the English group provided new neurophysiological evidence that tone processing is left lateralized at the scalp potential level when it is distinctive in the language and right lateralized or bilateral at the scalp potential level when it is not. In addition, this study showed that the use of natural speech in a multiple oddball paradigm demonstrated robust language group differences. These findings suggest that this paradigm could be used to effectively examine questions of second-language learning and language disorders.

Table 1. Participants. Age, sex and handedness information of the four groups of participants (2 interstimulus interval conditions by 2 language background conditions).

Participant group	Age (range, SD)	N(Sex)	Handedness
English Long ISI	28.4 (20-42, 6.6)	15(8M, 7F)	1 LH,14 RH
English Short ISI	29.8 (22-41, 5.6)	16(7M,9F)	1 LH, 15 RH
Mandarin Long ISI	29.1 (23-40, 5.3)	16(9M,7F)	all RH
Mandarin Short ISI	25.9 (21-36, 4.5)	16(8M, 8F)	1 ambidextrous, 15 RH

Table 2. Acoustic characteristics of the experimental stimuli

Stimuli	F0 (Hz)				Formant(Hz, mean values)							
	gV	**Pa**	**Onset (gV)**	**Offset (gV)**	**gV_F1**	**gV_F2**	**gV_F3**	**gV_F4**	**Pa_F1**	**Pa_F2**	**Pa_F3**	**Pa_F4**
gu1pa	186	194	190	182	340	1158	2794	4209	653	1394	2607	4140
gu1pa	214	214	219	207	364	1300	2902	4307	638	1487	2697	3967
gu2pa	174	202	169	175	339	1247	2799	4126	744	1509	2760	4120
gu2pa	166	184	161	167	349	1194	2758	4176	664	1472	2737	4328
gu3pa	140	167	155	142	345	1013	2622	4031	722	1538	2681	4022
gu3pa	142	168	155	136	345	1102	2630	4049	766	1465	2648	3961
gu3pa	143	171	155	141	341	1154	2688	3986	778	1478	2730	3989
gy3pa	138	166	136	134	312	1786	2599	3958	698	1495	2787	3993
gy3pa	142	186	151	137	331	1776	2465	4041	694	1418	2924	3984
gi3pa	208	175	290	264	326	1947	2666	3991	798	1603	2798	4145
gi3pa	196	173	157	135	344	2055	2617	4039	929	1581	2732	3767

Table 3. Behavioral discrimination accuracy(1 = 100% accuracy, and 0.50 is at chance accuracy)

Mean (SD)	T3/T3	T3/T1	T3/T2
Mandarin-Short	0.95(0.15)	0.97(0.06)	0.93(0.15)
Mandarin-Long	0.96(0.08)	0.97 (0.07)	0.94(0.10)
English-Short	0.87(0.13)	0.92(0.14)	0.67(0.29)
English-Long	0.89(0.27)	0.85(0.28)	0.61(0.31)

Table 4. Behavioral identification accuracy (1 = 100% accuracy, and 0.33 is at chance accuracy)

	T1: Mean(SD)	T2: Mean(SD)	T3: Mean(SD)
Eng_Long	0.23(0.29)	0.57(0.22)	0.16(0.20)
Eng_Short	0.29(0.34)	0.51(0.28)	0.49(0.33)
Mand_Long	0.73(0.39)	0.83(0.24)	0.85(0.26)
Mand_Short	0.82 (0.30)	0.82(0.18)	0.87(0.18)

Table 5. ANOVA results comparing standard T3 (/gu3pa/) with deviant T1 (/gu1pa/)for the MMN time window between 100 and 350 ms (Stim = stimulus condition, Man=Mandarin, Eng = English, S = short ISI, L = long ISI, Lang = language, T = time, n.s. = not significant)

Group	Effect	MMN(100-350 ms)			
		DF	F	P	ε(G-G)
All	Stim	1,59	10.9	0.001	
	Stim*Lang	1,59	5.33	0.02	
	Stim*ISI	1,59	0.08	n.s.	
	Stim*Lang*ISI	1,59	0.32	n.s	
	T*Stim	4,236	2.23	n.s	
	T* Stim *Lang	4,236	3.367	0.01	0.90
	T* Stim *ISI	4,236	3.50	0.01	0.90
	T*Stim*Lang*ISI	4,236	1.67	n.s	
ManS	Stim	1,15	5.421	0.03	
	T* Stim	4,60	0.872	n.s	
ManL	Stim	1,15	10.2	0.006	
	T* Stim	4,60	1.602	n.s.	
EngS	Stim	1,15	0.450	n.s.	
	T* Stim	4,60	7.433	< 0.001	0.64
EngL	Stim	1,14	0.121	n.s.	
	T* Stim	4,56	1.831	n.s.	

Table 6. MMN amplitudes for Tone 3/Tone 1 condition at Composite Fz

	100-149 ms	150-199	200-249	250-299	300-349
EEL	**-0.13**	**-0.16**	**0.09**	**0.11**	**-0.11**
EEL(SD)	*0.26*	*0.51*	*0.47*	*0.52*	*0.64*
EES	**0.03**	**-0.41****	**-0.15**	**0.11**	**0.15**
EES(SD)	*0.40*	*0.41*	*0.39*	*0.58*	*0.47*
MML	**-0.33***	**-0.17**	**-0.34****	**-0.49***	**-0.18**
MML(SD)	*0.46*	*0.39*	*0.43*	*0.67*	*0.56*
MMS	**-0.17**	**-0.25***	**-0.33**	**-0.10**	**-0.22**
MMS(SD)	*0.40*	*0.39*	*0.70*	*0.45*	*0.48*

*p< 0.05; ** p < 0.01; *** p < 0.001

Table 7. ANOVA results comparing standard T3 (/gu3pa/) with deviant T2 (/gu2pa/) for the MMN time window between 100 and 350 ms (Stim = stimulus condition, Man=Mandarin, Eng = English, S = short ISI, L = long ISI, Lang = language, T = time, n.s. = not significant)

Group	Effect	MMN(100-350 ms)			
		DF	F	P	ε(G-G)
All	Stim	1,59	44. 2	<0.001	
	Stim*Lang	1,59	3.79	0.056	
	Stim*ISI	1,59	0.46	n.s.	
	Stim*Lang*ISI	1,59	3.95	0.052	
	T*Stim	4,236	13.3	<0.001	0.85
	T* Stim *Lang	4,236	1.225	n.s	
	T* Stim *ISI	4,236	2.105	0.081	0.85
	T*Stim *Lang*ISI	4,236	0.389	n.s	
ManS	Stim	1,15	24.34	<0.001	
	T* Stim	4,60	9.37	<0.001	0.84
ManL	Stim	1,15	11.96	0.003	
	T* Stim	4,60	4.05	0.01	0.79
EngS	Stim	1,15	18.8	<0.001	
	T* Stim	4,60	3.36	0.04	0.57
EngL	Stim	1,14	1.63	n.s.	
	T* Stim	4,56	1.90	n.s.	

Table 8. MMN amplitudes for Tone 3/Tone 2 condition at Composite Fz

	100-149 ms	150-199	200-249	250-299	300-349
EEL	**0.05**	**-0.05**	**-0.33**	**-0.12**	**-0.04**
EEL(SD)	*0.31*	*0.37*	*0.52*	*0.48*	*0.41*
EES	**-0.06**	**-0.27**	**-0.42*** **	**-0.48*** **	**-0.39** **
EES(SD)	*0.43*	*0.39*	*0.43*	*0.46*	*0.47*
MML	**-0.19**	**-0.34* **	**-0.64*** **	**-0.57*** **	**-0.46* **
MML(SD)	*0.54*	*0.47*	*0.62*	*0.59*	*0.74*
MMS	**0.01**	**-0.22**	**-0.40*** **	**-0.42*** **	**-0.57*** **
MMS(SD)	*0.35*	*0.28*	*0.39*	*0.39*	*0.40*

*p< 0.05; ** p < 0.01; *** p < 0.001

Table 9. ANOVA results comparing standard T3 (/gu3pa/) with deviant T1 (/gu1pa/) for the LN time window between 350 and 550 ms (Stim = stimulus condition, Man=Mandarin, Eng = English, S = short ISI, L = long ISI, Lang = language, T = time, n.s. = not significant)

Group	Effect	LN (350-550 ms)			
		DF	F	P	ε(G-G)
All	Stim	1,59	11.6	0.001	
	Stim*Lang	1,59	0.58	n.s	
	Stim*ISI	1,59	0.836	n.s	
	Stim*Lang*ISI	1,59	0.001	n.s	
	T*Stim	3,177	47.0	< 0.001	0.86
	T* Stim *Lang	3,177	0.045		
	T* Stim *ISI	3,177	3.241	0.03	0.86
	T*Stim *Lang*ISI	3,177	2.04	n.s.	
ManS	Stim	1,15	2.136	n.s	
	T* Stim	3,45	12.04	< 0.001	0.79
ManL	Stim	1,15	6.92	0.02	
	T* Stim	3,45	13.52	< 0.001	0.84
EngS	Stim	1,15	1.337	n.s	
	T* Stim	3,45	12.97	< 0.001	0.62
EngL	Stim	1,14	2.49	n.s	
	T* Stim	3,42	14.1	< 0.001	0.65

Table 10. LN amplitudes for Tone 3/Tone 1 contrast at Fz composite

	350-399 ms	400-449	450-499	500-549
EEL	0.19	-0.46*	-0.56**	-0.33
EEL (SD)	*0.52*	*0.79*	*0.69*	*0.72*
EES	0.34*(+)	-0.13	-0.38***	-0.17
EES(SD)	*0.57*	*0.4*	*0.35*	*0.44*
MML	0.19	-0.46*	-0.67***	-0.20
MML(SD)	*0.63*	*0.68*	*0.51*	*0.3*
MMS	0.19	-0.2	-0.33*	-0.45**
MMS(SD)	*0.63*	*0.56*	*0.61*	*0.48*

*p< 0.05; ** p < 0.01; *** p < 0.001

Table 11. ANOVA results comparing standard T3 (/gu3pa/) with deviant T2 (/gu2pa/) for the LN time window between 350 and 550 ms (Stim = stimulus condition, Man=Mandarin, Eng = English, S = short ISI, L = long ISI, Lang = language, T = time, n.s. = not significant)

Group	Effect	LN (350-550 ms)			
		DF	F	P	ε(G-G)
All	Stim	1,59	4.37	0.044	
	Stim*Lang	1,59	3.65	0.064	
	Stim*ISI	1,59	0.008	n.s	
	Stim*Lang*ISI	1,59	2.84	n.s	
	T*Stim	3,177	20.3	< 0.001	0.72
	T* Stim *Lang	3,177	0.236	n.s	
	T* Stim *ISI	3,177	2.69	n.s.	
	T*Stim *Lang*ISI	3,177	0.616	n.s	
ManS	Stim	1,15	1.505		
	T* Stim	3,45	3.387	0.06(n.s.)	
ManL	Stim	1,15	3.998	0.06	
	T* Stim	3,45	6.965	0.002	0.74
EngS	Stim	1,15	1.132	n.s	
	T* Stim	3,45	10.8	< 0.001	0.84
EngL	Stim	1,14	1.68	n.s.	
	T* Stim	3,42	4.92	0.01	0.66

Table 12. LN amplitudes for Tone 3/Tone 2 contrast at Fz composite

		LN for T2 at Composite Fz		
	350-399 ms	400-449	450-499	500-549
EEL	**0.29*(+)**	**-0.07**	**-0.18**	**0**
EEL (SD)	*0.41*	*0.26*	*0.47*	*0.48*
EES	**0.27**(+)**	**-0.14**	**-0.23***	-0.23*
EES(SD)	*0.34*	*0.49*	*0.42*	*0.37*
MML	**0**	**-0.41***	**-0.53****	**-0.15**
MML(SD)	*0.7*	*0.79*	*0.61*	*0.37*
MMS	**0.17**	**-0.16**	**-0.18**	**-0.27****
MMS(SD)	*0.64*	*0.57*	*0.44*	*0.36*

*p< 0.05; ** p < 0.01; *** p < 0.001

Table 13. Pearson's Correlation Coefficient *r* between language and ISI groups on the ERP topographies of the subtraction waves for early and late peak time window. Significant correlations are in bold (df = 65-2, $r > 0.316$, $p < 0.01$), and strong correlations are underlined (df=65-2, $r > 0.40$, $p < 0.001$).

	150-200		200-250		250-300		400-450		450-500		500-550	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Eng Short vs Mand Short	**0.66**	**0.52**	**0.34**	**0.4**	-0.001	**0.45**	0.06	**-0.65**	0.29	-0.05	0.07	**0.42**
Eng Long vs Mand Long	**0.55**	0.10	**0.56**	**0.65**	**0.56**	**0.33**	**0.39**	0.17	**0.56**	**0.47**	**0.35**	**0.43**
Eng Short vs Eng Long	**0.58**	0.15	**0.44**	**0.54**	**0.51**	**0.51**	**0.64**	-0.20	**0.79**	0.28	**0.62**	0.27
Mand Short vs Mand Long	**0.48**	**0.48**	**0.48**	**0.58**	**0.37**	0.22	**0.83**	-0.09	**0.76**	0.14	**0.50**	**0.34**

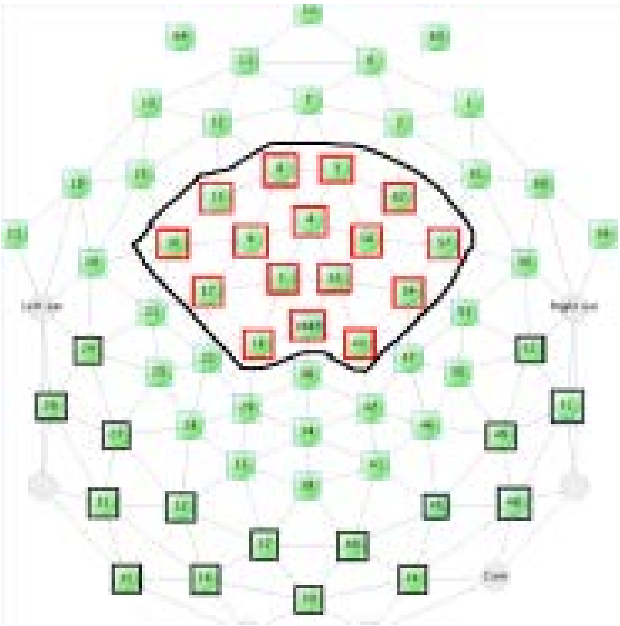


Figure 1. The Fz composite site was built using the average of the 16 frontocentral sites (highlighted with red squares). These 16 sites are highly correlated with Fz. The inferior 16 sites (in blue squares) are negatively correlated with the superior frontocentral sites.

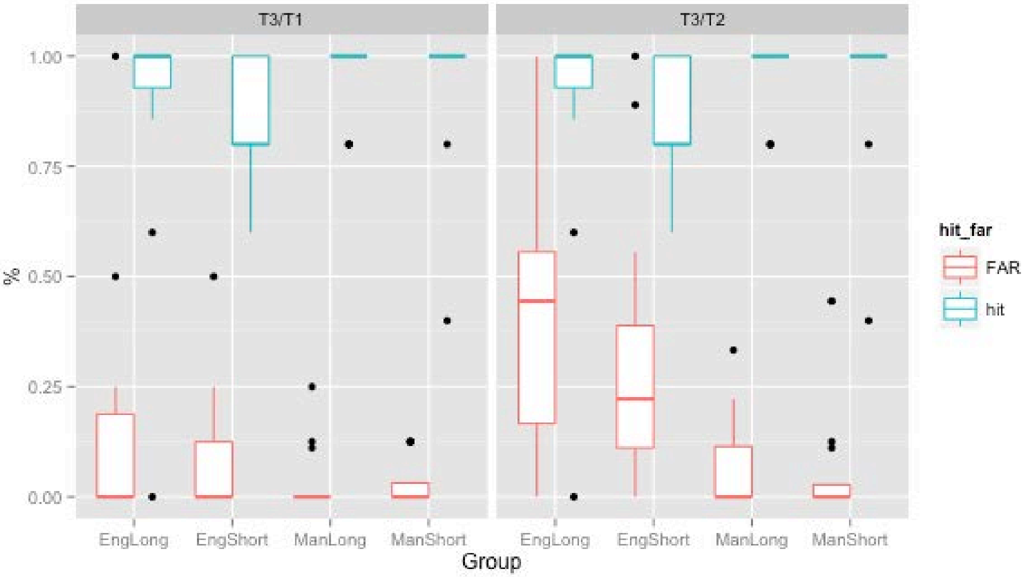


Figure 2. The hit rates and false alarm rates for tone discrimination.

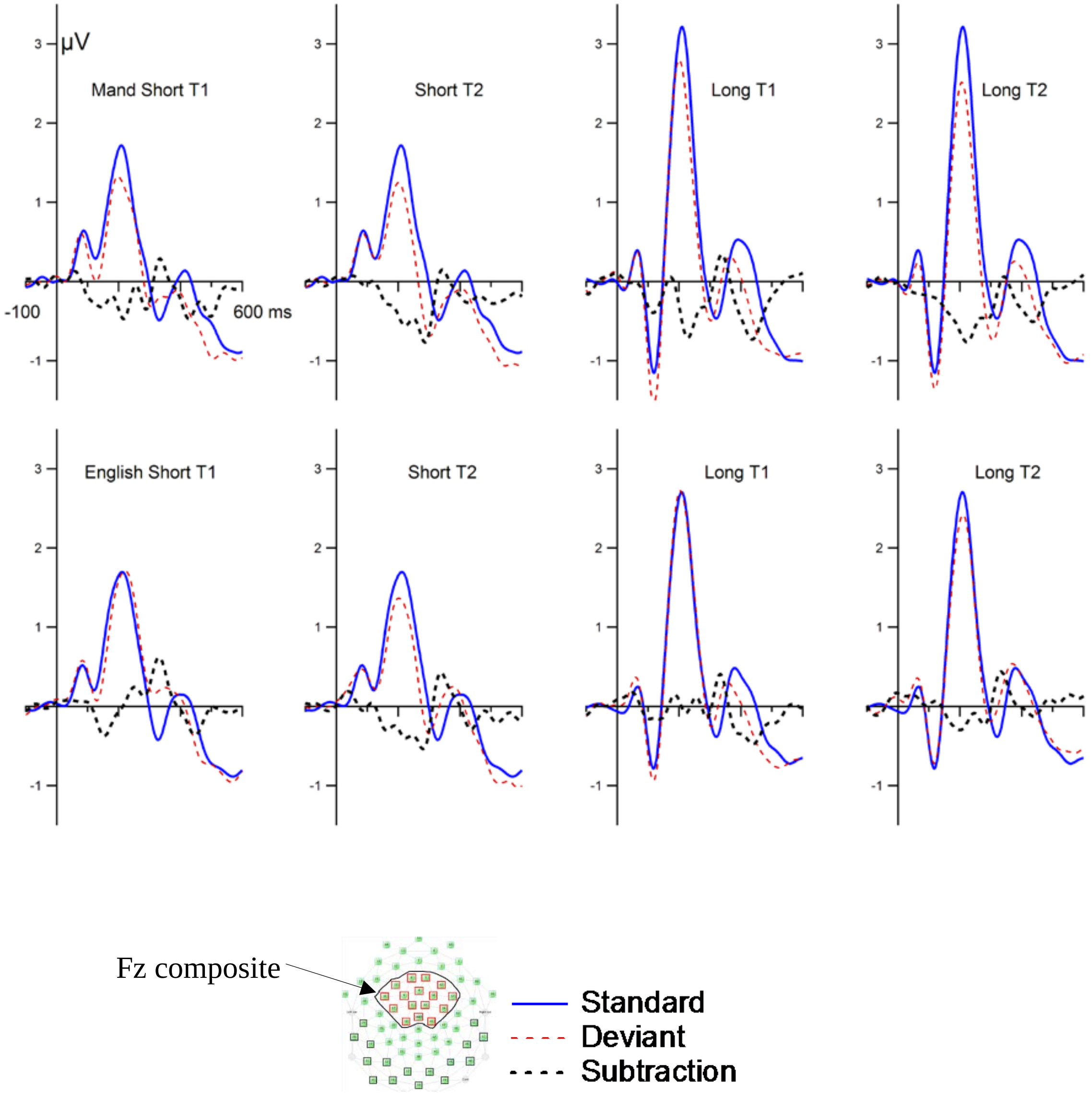


Figure 3. Grand average ERPs to the standard and deviant stimulus and subtraction waveforms for the composite Fz. The top panel shows Mandarin short ISIgroupT1 and T2 conditions (left), and long ISI group T1 and T2 conditions (right). The bottom panel shows English short ISI group T1 and T2 conditions (left), and long ISI group T1 and T2 conditions (right).

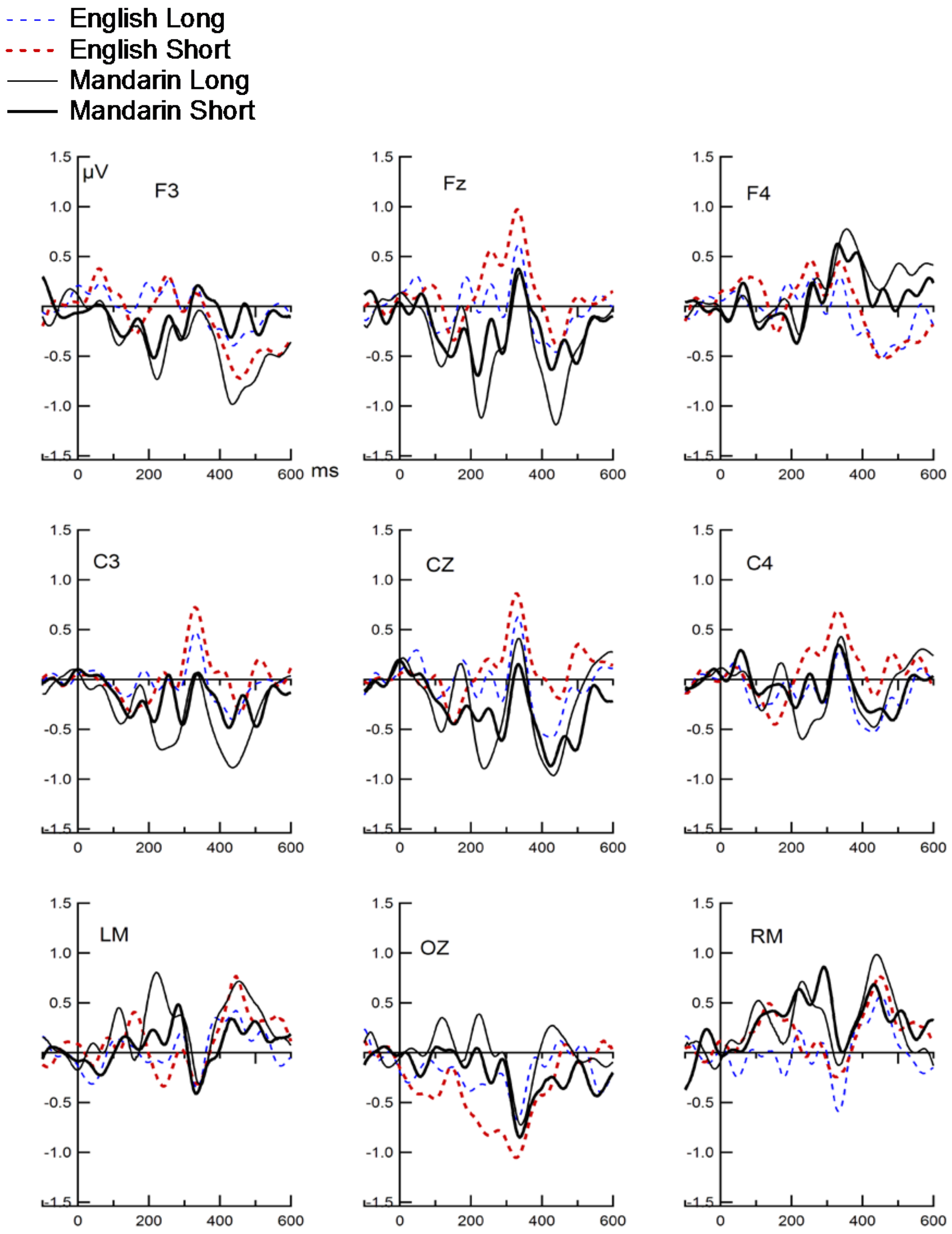


Figure 4. Subtraction wave for Tone 1 versus Tone 3 at nine sites for all four groups.

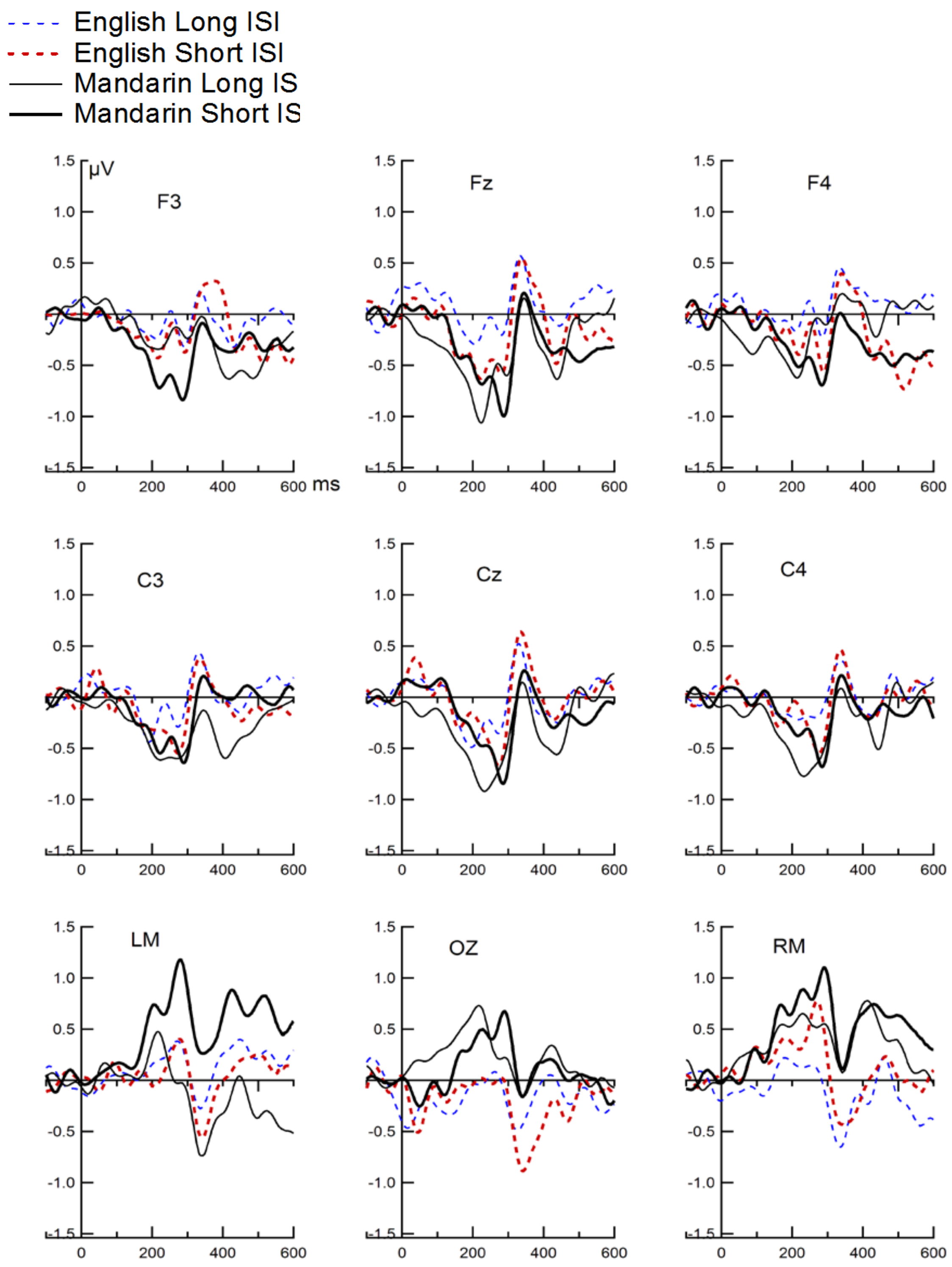


Figure 5. Subtraction wave for Tone 2 versus Tone 3at nine sites for all four groups.

Appendix One: Vowel Study

The duration of auditory sensory memory for vowel processing: Mismatch negativity and late negativity

Vowel Abstract

Speech perception behavioral research suggests that rates of sensory memory decay are dependent on stimulus properties at more than one level (e.g., acoustic level, phonemic level). The neurophysiology of sensory memory decay has rarely been examined in the context of speech processing. In a lexical tone study, we showed that long-term memory representation of lexical tone modulates the decay rate of sensory memory for these tones. Here, we tested the hypothesis that long-term memory representation of vowels modulates the rate of auditory sensory memory decay in a similar way to that of lexical tone.

We measured auditory sensory memory using an oddball paradigm involving electrophysiological components called the mismatch negativity (MMN) and late negativity (LN). Sensory memory effects were assessed by varying the inter-stimulus interval (ISI) between the standard and deviant. Event-related potential (ERP) responses were recorded from native Mandarin and native American English participants under short and long ISI conditions (short ISI: an average of 575 ms, long: an average of 2675 ms). The standard (/gupa/) and deviant (/gipa/ and /gypa/) stimuli are multiple bisyllabic natural speech nonwords containing two vowel contrasts, one which is phonemic for both English and Mandarin listeners, and a second which is

phonemic only for Mandarin listeners. Behavioral discrimination and identification tasks followed the ERP measures.

We found that the MMN effects were present for all groups and both deviant conditions except the English short ISI /gipa/-/gupa/ contrast, and the LN effects were present for all groups and both deviant conditions except that no LN for either language groups under the short ISI /gypa/-/gupa/ condition. In terms of memory trace decay effect, the Mandarin group showed larger MMN than the English group regardless of ISI for the /gypa/ deviant condition. As expected, the MMN amplitudes in the long ISI conditions were reduced comparing to that in the short ISI conditions for /gypa/-/gupa/ contrast. Surprisingly, for the /gipa/-/gupa/ contrast, the MMNs were larger in the long ISI conditions than in the short ISI conditions. Behavioral discrimination results showed that there was a steep decrement in performance for English listeners in discriminating the non-native contrast, /gypa-gupa/, especially in the long ISI condition comparing to the discrimination performance in the /gipa –gupa/ condition.

This study provided new evidence that native language experience affects different ERP components (here, MMN and LN) in different ways. By using different ISIs, we demonstrated that native language experience plays a role in echoic sensory memory trace maintenance.

V1. Introduction

Mismatch negativity (MMN), an event-related potential (ERP) component, can be elicited to a change (called the deviant) to a frequently heard standard stimulus. The MMN is considered a neural index of human auditory sensory (echoic) memory representation (Haenschel,Vernon, Dwivedi, Gruzelier, & Baldeweg,, 2005; Näätänen, Gaillard, & Mäntysalo, 1978). The repetition of the standard stimulus results in the establishment of a memory trace to which each successive stimulus is compared. The strength and durability of this sensory memory trace, as reflected by the MMN, is affected by a number of factors, including the number of standard stimulus repetitions before the deviant (Imada, Hari, Loveless, McEvoy, & Sams, 1993; Javitt, Grochowski, Shelley, & Ritter 1998; Sams, Alho, & Näätänen, 1983), the acoustic distinctiveness of the standard-deviant contrast (e.g., Sams, Paavilainen, Alho &Näätänen, 1985), the linguistic status of the contrast (e.g., Näätänen, et al., 1997), and the rate of stimulus presentation (e.g., Schröger, 1996). More specifically, a smaller MMN amplitude has been observed to a smaller magnitude of stimulus change in tone frequency, duration or intensity (e.g. Amenedo & Escera, 2000;Lang, Nyrke, Market, Ek, Aaltonen, Raimo, & Näätänen, 1990; Rinne, Särkkä, Degerman, Schröger & Alho, 2006; Sams, Paavilainen, Alho, & Näätänen, 1985;Näätänen, 1985; Tiitinen, Sinkkonen, May, & Näätänen, 1994). Several studies have also shown smaller or absent MMNs for non-native compared to native phonetic contrasts that are phonemic only for the native group (Dehaene-Lambertz, 1997; Näätänen, et al., 1997; Winkler, Lehtokoski, Alku, et al., 1999; Sharma & Dorman, 1999). MMN also decreases in amplitude when the ISI between standard and deviant auditory stimuli is increased (Barry, Hardiman, Line, White, Yasin, Bishop, 2008;Čeponienė, Cheour, & Näätänen, 1998; Čeponienė, Service, Kurjenluoma, Cheour, & Näätänen, 1999; Gomes, Sussman, Ritter, Kurtzberg, Cowan, & Vaughan, 1999; Javitt, et al., 1998;Sam et al., 1993; Schröger, 1996).

V1.1. Sensory memory decay as measured by MMN

Sensory memory decays in a nonlinear fashion, but this decay rate is dependent on a number of factors. In healthy adults, MMN can be elicited with an ISI as long as 10 seconds when the stimuli are auditory pure tones that are different in frequency by 10 percent (e.g., Sams et al., 1993). Speech perception research suggests that rates of sensory memory decay are dependent on stimulus properties at more than one level. For example, at the acoustic level, a “simpler” (steady-state) vowel has an advantage over a “complex” (brief and transitional) consonant in terms of the rate of decay (Pisoni, 1973). At a phonemic level, a between-category but not within category consonant contrast differing equally on an acoustic scale, can be retained for successful behavioral discrimination at an ISI of 1.5 seconds (Werker & Logan, 1985). In the main paper of this dissertation, we found that native speakers of English failed to show an early negativity (i.e., MMN) to a Mandarin lexical tone contrast, which is phonemic in Mandarin, but not English, when the ISI was greater than 2.5 seconds; however, English listeners did show MMN to the lexical tone contrast using a short ISI of approximately 500 ms. To our knowledge, there is no study that has directly examined the duration of the neuronal trace of auditory sensory memory for vowels, and only one for consonants (Čeponienė, Service, Kurjenluoma, Cheour, &Näätänen,

1999). Čeponienė and colleagues (1999) found that two groups of 7-9-year-old children with high and low phonological memory skills, as measured by nonword repetition (NWR), showed very similar MMN responses to auditory tone changes (1000 Hz versus 1100 Hz) under both short (350 ms) and long (2000 ms) ISI conditions. However, the good nonword repeaters differed from the low repeaters in terms of the MMN amplitude for a consonantal voicing contrast (/baga/-/baka/). The MMN amplitude was greatly reduced in high repeaters under the long ISI condition compared to the short ISI condition, and no MMN was observed for either short or long ISI condition in the low repeaters.

A goal of the current paper is to examine how the rate of stimulus presentation affects the formation and durability of the neural sensory memory trace to vowel, as measured by MMN.

V1.2. Late Negativity

An oddball paradigm, in which one stimulus is frequently repeated (standard) and the other is rare (the deviant) is used to establish a standard sensory memory trace and to introduce the stimulus change that will elicit neural discriminative responses. This paradigm, when presented using a passive task, often generates two types of mismatch responses: the classic MMN that is generally observed in an early time window between 100 and 300 ms following the onset of stimulus difference and a late negativity (LN) (also called late discriminative response or MMN2) observed in a later time window (300-500 ms) in relation to the stimulus difference. The MMN and LN may co-occur or may be present independently of each other. The MMN has been extensively studied (see Näätänen, Paavilainen, Rinne, & Alho, 2007 for a review). The LN has less frequently been reported, but an increasing number of studies suggest that it is a fairly common mismatch response (Barry, Hardyman, Bishop, 2009; Bishop, Hardyman, Barry, 2010; Datta, Shafer, Morr, Kurtzberg, & Schwartz, 2010; Kaan Wayland, Bao, & Barkley, 2007; Korpilahti, Krause, Holopainen, & Lang, 2001; Ortiz-Mantilla, Choudhury, Alvarez, & Benasich, 2010; Shestakova, Huotilainen, Čeponienė, & Cheour, 2003; Shafer, Morr, Datta, Kurtzberg, & Schwartz, 2005). Based on the relatively consistent temporal relationship, but different topographic features between MMN and LN, Čeponienė, and colleagues proposed that LN (called Late Difference Negativity in their paper) is less likely to be linked to sensory aspects of processing, but more likely to reflect higher order, albeit preattentative, cognitive processing of sound change (Čeponienė, Lepisto, et al., 2004). However, it is not clear whether this higher order cognitive processing of sound corresponds to the extraction of a perceptual representation, or is just a further step of a discrimination process after the initial sensory processing, as indexed by MMN.

V1.3. Non-native speech perception

A number of models in the speech perception field have been proposed to account for listeners' perception of non-native contrasts. Two speech perception models, the perceptual assimilation model (PAM) (Best, 1994; Best & Tyler, 2007) and the Automatic Selective Perception model (ASP)(Strange, 2011; Strange & Shafer, 2008) are of particular interest to this study. The perceptual assimilation model (PAM) accounts for how the first language (L1) system constrains the perception of nonnative phones that are completely unfamiliar to the listeners

(Best, 1994; Best & Tyler, 2007). PAM proposed that naïve listeners perceptually assimilate a nonnative phone to “the most articulatorily-similar native phoneme”, when the listeners can find a match (good or poor) in their L1. PAM provides specific prediction regarding non-native perception with respect to the goodness of fit of the contrast to a non-native category. For example, native English listeners perceived the French front rounded vowel /y/ as the back rounded vowel /u/(Levy & Strange, 2008). English rounded vowels are all back vowels, which may account for this pattern of assimilation. Of particular interest to this study are American English listeners’ expected patterns of assimilation of Mandarin vowels. Similar to French, Mandarin has a high front rounded vowel /y/, but also the front vowel /i/ and back vowel /u/. English listeners should demonstrate two-category perceptual assimilation (TC) when discriminating Mandarin tokens of /u/ and /i/; that is, they should assimilate the Mandarin /u/ and /i/ tokens into the English /u/ and /i/ phoneme categories, respectively. In contrast, English listeners are expected to assimilate tokens of the Mandarin front-rounded vowel /y/ into the English /u/ vowel category. Their perception of this vowel should display a discrimination pattern of category-goodness (CG) or single category (SC) assimilation. That is, the English listeners will perceive one sound as being less different from the prototypical English vowel /u/ than the other (CG pattern), or they will perceive both Mandarin /y/ and /u/ tokens as equally good examples of English /u/.This prediction derives from the evidence that American English learners of French often perceived the French /y/ tokens as American English /u/ (Levy & Strange, 2008).

The Automatic Selective Perception (ASP) model proposes that native language listeners have a learned habit of listening that leads them to selectively weight phonetic cues. This selective weighting (or selective perceptual routines, SPRs) allows them to automatically recover the relevant acoustic-phonetic information needed to identify words in the L1. While listening to non-native (or second language) stimuli, their specific L1 SPRs interfere with nonnative speech processing, especially under non-optimal listening situations, such as situations in which attention and memory resources are limited, speaking rate is fast, or the stimulus contrasts are complex (Strange, 2011; Strange & Shafer, 2008). The PAM and ASP models are compatible with each other, with PAM focusing on the specific perceptual patterns found for language pairs and ASP focusing on the nature of the processing mechanism.

V1.4. ERP studies of non-native speech processing

Studies using MMN support the PAM and ASP models. Of particular relevance to this study are experiments that have shown a larger MMN to a vowel contrast for one group for which the two vowel stimuli are perceived as two different phonemes compared to the other group, for whom the two vowels are perceived as members of one phoneme category (Lipski, Escudero, & Benders, 2012; Näätänen et al., 1997; Nenonen, Shestakova, Houtilainen, & Näätänen, 2005; Peltola, Kujala, Tuomainen, Ek, Aaltonen, & Näätänen, 2003; Winkle et al., 1999; Ylinen, Streinikov, Huotilainen, & Näätänen, 2009). For example, Näätänen and colleagues (1997) found that Estonian listeners showed a larger MMN to the Estonian /õ/ versus /e/ vowel contrast than Finnish listeners. The vowel /õ/ is phonemic in the Estonian, but not the Finnish language. In addition, the contrast between the Finnish vowels /ö/ and /e/ generated a larger amplitude MMN than the Estonian /õ/ versus /e/ contrast in the Finnish listeners, even though the latter contrast was acoustically greater. Another study showed a similar pattern using asymmetric cross-language design (Winkler, Lehtokoski, Alku, et al., 1999). They examined the MMN responses to both a within-category contrast (falling within the /e/ category) and cross-category contrast (/é/ and /e/ for Hungarian listeners, and/ɛ/ and /æ/ for Finnish listeners). With a constant SOA of 1.2 s, both within- and across-category vowel contrasts elicited the MMN responses, but the MMN amplitude was larger in across- than within-category contrasts in both Finnish and Hungarian groups.

The LN has been observed in a few studies using speech contrasts, but its modulation via language experience has not been systematically explored. Barry et al. (2009) examined the relationship between MMN, LN and individual adults' nonword repetition (NWR) ability and found that there was no difference in the MMN responses between the good and poor NWR performers, but LN was larger in one of the four deviant conditions for the good compared to poor NWR performers. In contrast to the findings of Barry et al. (2009), several studies observed that LN is larger to speech contrasts in second language learners (e.g., Ortiz-Mantilla et al., 2010) in children with language impairment (e.g., Shafer, et al.,2005), or in other clinical groups (e.g., Addis, Friederici, et al., 2010). Shafer and colleagues suggest that the larger LN observed in children with language impairment may reflect inaccurate feature weighting (Shafer et al., 2005). In the lexical tone study, we found larger LN in the native Mandarin listeners than in the English listeners for Mandarin lexical tone contrasts. The LNs were either the same or larger amplitude in the long ISI condition than the short ISI condition in the Mandarin listeners. It remains unclear whether LN is actually a second MMN, or a separate component of its own. Further study on LN will help elucidate the nature of LN. In addition, no study, other than the experiment reported in the lexical tone study of this dissertation, has examined how memory trace decay affects LN.

V1.5. Objectives of the study

The focus of this paper is on the modulation of MMN and LN under different ISI conditions for two vowel contrasts, one which is phonemic for both English and Mandarin listeners, and a second which is phonemic only for Mandarin listeners. We used an average ISI of 575 ms (range 545 -609 ms) for the short ISI condition and an average ISI of 2675 ms, (range of 2645 to 2709 ms) for the long ISI condition (see Chapter 2). We hypothesized that native language experience modulates vowel perception at both behavioral and preattentative neuronal level, and the degree of success and ease in discriminating a nonnative vowel contrast depends on the relative relationship between the L1 and the nonnative language. Based on the PAM and ASP models and previous findings, we predicted that native English listeners would demonstrate difficulty perceiving the /u/-/y/ contrast because they might assimilate the /y/tokens into the /u/ phoneme category, and, thus, fail to discriminate tokens of /y/ from tokens /u/. In contrast, they would show good perception of the /i/ versus /u/ contrast, seen as good discrimination and categorization of tokens of /i/ and /u/ into different categories. In terms of the ASP model, the L1 automatic selective perception routines will select information that is relevant to the L1, and for English, vowel roundness will be automatically selected to indicate a back vowel. This will lead to the automatic perception of Mandarin tokens of /y/ as English /u/, and result in a smaller MMN than found for the Mandarin /i/ versus /u/ contrast. With a short ISI the acoustic-phonetic cues that allow for category goodness judgments may allow better than chance discrimination of Mandarin tokens of /y/ and /u/ by English listeners and less of a difference should be found between English and Mandarin listeners. In contrast, with a long ISI, the automatic SPRs will select the relevant cues and only these will be supported by long-term memory representations. Thus, behavioral perception and the MMN brain discriminative response will be poorer at the long than short ISIs and poorer for English than Mandarin listeners for the /y/ versus /u/ contrast. No differences in behavioral perception or MMN are expected for the /i/ versus /u/ contrast. Finally, we predicted that a reduction in LN amplitude may not be observed in nonnative listeners since a robust LN has been found for populations with language impairment (see Shafer et al., 2005) or those who are second language learners (Ortiz-Mantilla et al., 2010).

V2. EXPERIMENTAL PROCEDURES

V2.1. Participants

Data from a total of 63 adult participants using a between-subject design were included. The 31 native speakers of English (16 participants in the short ISI and 15 participants in the long ISI condition, age range: 20-42 years) had little or no exposure to any tone languages. The 32 native speakers of Mandarin (16 participants in each ISI condition, age range: 21-40 years) were all from Mainland China, and all moved to the United States no earlier than high school years. Data from a total of five participants (two English and three Mandarin) were excluded from the analysis due to incomplete participation (N = 2), or excessive noise in the data, defined by retaining less than 50% of trials after artifact reject (N = 2) or no clear obligatory components (N = 1) The participants were the same as in Chapter 2(Paper 1). Table V1 describes participant's age, sex and handedness.

V2.2. Stimuli

The stimuli consisted of eleven tokens of bisyllabic nonsense words produced by a native Mandarin female speaker, as described in the lexical tone study of this dissertation. This part of the dissertation focuses on the vowel deviant conditions. The standard stimuli consisted of three tokens of /gupa/ with a low-rising tone (Tone 3 in Mandarin) on the first syllable, and two tokens of /gupa/ with a high-level tone on the first syllable (Tone 1), and two tokens of /gupa/ with a rising tone (Tone 2) on the first syllable. Only the three tokens of /gupa/ with the same tone (Tone 3) as the vowel standard stimuli were included in the ERP analysis and averaged to create the standard response. The two deviant types (two tokens each) were /gypa/ and /gipa/. Vowel /u/ and /i/ are phonemically contrastive in both English and Mandarin. The front-rounded vowel /y/, however, is found only in Mandarin. American-English listeners were expected to perceive /y/ as the back-rounded /u/, as found in studies examining American-English listeners' perception of the French /y/ (Levy and Strange, 2008).

V2.3. Paradigm

Participants were seated in a sound-attenuated and electrically shielded booth for a passive listening MMN paradigm and behavioral tasks. E-Prime software (Psychology Software Tools, Pittsburgh, PA, USA) was used for stimulus presentation and behavioral data collection. A 65-channel Geodesic sensor net was used for ERP data collection.

Twenty blocks with 103 stimuli in each block were presented with an interblock interval of 20 seconds. The vowel and lexical tone deviant stimuli were interspersed within the sound stream of common standard stimulus type /gu3pa/. Table V2 provides examples of the stimulus sequences used in the study. A total of 200 deviant trials per category with a probability of 9.7% were presented. So the three tokens of common standard /gu3pa/ occurred with a probability of 62.2%, two tokens of vowel standard /gu1pa/ and two tokens of vowel standard /gu2pa/ occurred 9.7% each, and two tokens of deviant /gy3pa/ and two tokens of deviant /gi3pa/ occurred 9.7% each. The /gu1pa/ and /gu2pa/ stimuli were not included in the analysis as they have a dual role of serving as the vowel standard and lexical tone deviant in the experiment.

A vowel behavioral discrimination task (four standard and a fifth/final deviant) followed the ERP measurement. Participants were asked to determine whether the stimulus in the final position of the sequence of five stimuli was the same or different from the previous four stimuli. The ISI between stimuli was the same as for the ERP paradigm, that is, short ISI for participants receiving the short ISI ERP condition and long ISI for those receiving the long ISI ERP condition. A three-alternative choice identification task with “gipa”, “gupa” or “gypa” as the alternatives was presented last. Three practice trials plus 30 test trials were presented in the vowel discrimination task, and six practice trials plus 30 test trials were presented in the vowel identification task.

All the stimuli were presented free-field with a comfortable listening level of 70.2 dB (SD= 1.9 dB). The entire experiment lasted 2-2.5 hours for the short ISI experiment, and 3-3.5 hours for the long ISI experiment. Breaks were given halfway through the ERP experiment and whenever the participant requested. The participants were reimbursed 10 dollars per hour for their time. The study was approved by the institutional research board of City University of New York and all participants provided informed consent.

V2.4. ERP recording and offline processing

The continuous EEG was time-locked to the onset of the stimuli. The EEG was recorded with a band pass of 0.1 -100 Hz, and a sampling rate of 500 Hz from 64 scalp sites using a Geodesic sensor net with the vertex electrode (Cz) as the reference. For offline processing, the EEG was refiltered using a band-pass filter of 0.3-15 Hz, and segmented into 1200 ms epochs, including a 200 ms pre-stimulus baseline. Following artifact rejection, epochs were averaged for each subject and stimulus category. The data were re-referenced using the average of all 65 sites and baseline corrected. After artifact removal, on average, 178.4 trials (89%; SD = 15.7) for the gipa deviant condition, 177.8 trials (89%; SD = 16.8) for the /gypa/ deviant condition and 408.3 (86%, SD = 35.8) trials for the /gupa/ standard condition were included in the individual average.

V3. Data analysis

V3.1. Behavioral analysis

For both discrimination and identification task, we applied mixed effects logistic regression model using language, ISI and vowel type as fixed effect and subject as random effect. The response accuracy was the outcome variable. Vowel type is nested under subject, and the Mandarin group, the short ISI and the /gipa/-/gupa/ contrast served as the reference level for the discrimination task. For the identification task, the Mandarin group, short ISI and /gupa/ stimulus condition served as the reference level for the major model, and the /gipa/ stimulus condition served as the reference level for the sub-model comparing /gipa/ and /gypa/ stimulus effect. All three predictor variables (ISI, language, and tone) and all interactions were entered into the models using forward modeling. We compared the overall fit of the models using ANOVA to find the difference in the value of the deviance for the model. The final model was selected based on the change in the deviance statistics and the significance value of its associated chi-square statistics ($p < 0.05$).

V3.2. "Composite" FzCz measures

To reduce inter-subject variation in the topography of the ERP to speech (Zevin, et al., 2010), and to reduce the contribution of independent noise sources at each electrode site to the signal of interest (Parra et al., 2004), we built a model of frontocentral activity from six sites around Fz and Cz as follows. We chose Fz and Cz as pivotal sites because MMN is known to have a frontocentral topography, and visual inspection of the data shows that Fz and Cz do indeed have the consistently largest MMN amplitude across participants. We then calculated the Pearson’s correlation coefficients between Fz and each of the 64 channels and Cz and each of the 64 channels for individual stimulus conditions within each language/ISI group (5 stimulus conditions by 2 language groups by 2 ISI conditions) within the 120 to 320 ms time window (five 40-ms time-windows). The four sites that had the highest average correlation across conditions with Fz and Cz were selected (mean correlation = 0.92, SD = 0.06). Thus, the six sites used in the model were Fz/4, 5, 9, 55, 58 and Cz/65. Figure V1 depicts the electrode placement.

V3.3. ERP analyses

We first examined whether significant negativity was obtained in the MMN and LN time range by conducting repeated measure mixed ANOVAs using the mean amplitude of four consecutive 20-msec time windows from 100 to 180 ms for the /gipa/-/gupa/ contrast, and 160 to 240 ms for the /gypa/ -/gupa/ contrast to test for presence of the MMN. Likewise four consecutive 20-msec time windows from 380 to 460 ms were selected to test for presence of the LN. Four-way mixed ANOVAs with ISI and language as between-subject variables, and time and stimulus condition (standard, deviant) as within-subject variables for each standard-deviant pair were conducted on these intervals. In a second set of analyses designed to test for language group and ISI condition differences in the time range of interest, we used the difference waves (deviant minus standard) in mixed ANOVAs with ISI and language as between-subject variables and time and deviant type as within-subject variables.

To directly compare the amplitude and latency differences of the early negativity for the two deviant types, the amplitude and latency of the most negativity interval of the subtraction waves (deviant minus standard) were used. The maximum values were selected using the four, down sampled 20-millisecond time intervals between 100 and 240 ms. This analysis allowed more precise examination of the latency of the response independently of amplitude and of the amplitude of the response independently of latency. Three-way ANOVAs with language group (English, Mandarin), ISI (short, long), and deviant stimulus type (gipa, gypa) were performed separately for these latency and amplitude measures.

The Greenhouse-Geisser correction was applied whenever the degree of freedom in the denominator is larger than one. Uncorrected degree of freedom, corrected p-value and epsilon value (ε) were reported. Two-way interactions were examined using Tukeys' HSD post-hoc tests.

V4. RESULTS

V4.1. Behavioral discrimination results

Figure V2 depicts the scores for accuracy by language, ISI and deviant type. Table V3 presents the results for the mixed effects logistic regression. The results indicated that deviant stimulus type significantly predicted discrimination accuracy ($p < 0.001$), with the /gypa/ condition having lower accuracy than /gipa/ condition. Neither language nor ISI itself significantly predicts discrimination accuracy, but an interaction of deviant stimulus type and language ($p < 0.001$) and an interaction of language and ISI ($p < 0.05$) both significantly predicted discrimination accuracy. The combination of the /gypa/ condition and an English background yielded a lower accuracy; likewise, the combination of long ISI and an English background resulted in lower accuracy.

V4.2. Behavioral identification results

Figure V3 summarizes the behavioral identification task results. All groups labeled each category as intended well-above chance level (> 33.3%). Table V4 presents the parameter information from the results for the mixed effects logistic regression. Mandarin, short ISI and /gupa/ served as the reference level. Both /gipa/ and /gypa/ tokens showed significantly lower accuracy than /gupa/ tokens ($ps < 0.001$). Language was also a significant predictor ($p < 0.001$) with the English group having lower accuracy. ISI was not a significant predictor. A second model was constructed to further examine the difference between /gipa/ and /gypa/. This model excluded the /gupa/ stimulus condition. This analysis showed that language significantly predicted identification accuracy ($p < 0.001$) and ISI was not a significant predictor, as in the model including /gupa/. A significant interaction was obtained for language and condition. For participants with an English background, the /gypa/ tokens were categorized with significantly lower accuracy ($p < 0.001$).

V4.3. ERP results

Figure V4 displays the grand mean ERPs to the standard and deviant stimuli and subtraction (deviant minus standard) waveforms for the composite FzCz site.

V4.3.1. Early time intervals: 100-240 ms

Standard vowel /gupa/ versus deviant vowel /gipa/

Table V5 summarizes the statistical findings for these analyses. The four-way ANOVA on the 100-180 ms time interval revealed a main effect of ISI ($p < 0.001$) in which the responses in the long ISI conditions were generally more negative than those in the short ISI conditions, and a main effect of stimulus ($p = 0.009$) with the deviant being more negative than the standard. Significant interactions of stimulus by time ($p < 0.001$) and stimulus by time by ISI ($p < 0.05$) were also found. Post hoc tests following the interaction of stimulus by time showed that the deviant was more negative than the standard in the time window of 120-140 ms. The three-way interaction of stimulus by time by ISI was followed up by a step-down analysis on each ISI/language group, because we were interested in whether there was a difference between the standard and deviant stimulus for each of the four subgroups. ANOVAs were performed for each subgroup with stimulus and time as factors. For the Mandarin short ISI group, a main effect of stimulus ($p < 0.05$) and a time by stimulus interaction ($p < 0.05$) were found. Post hoc tests following the time by stimulus interaction revealed that the deviant was more negative than the standard between 120-140 ms, and 160-180 ms. For the Mandarin long ISI group, greater negativity was present from120-140 ms ($p = 0.02$). No difference between the standard and deviant was observed for the English short ISI group between 100 and 180 ms. For the English long ISI group, greater negativity of the deviant than the standard was present from120 to140 ms ($p = 0.02$).

Standard vowel /gupa/ versus deviant vowel /gypa/

Table V6 provides statistics for the four-way mixed ANOVAs of the signal integrated between 160 and 240 ms. The results showed a significant main effect of ISI ($p < 0.001$), in which the response in the short ISI condition had larger negativity than in the long ISI condition, and a main effect of stimulus ($p < 0.001$) in which the deviant was more negative than the standard. Significant interactions of stimulus by language ($p < 0.01$), and stimulus by time ($p < 0.01$) were also found. Post hoc tests following up the stimulus by language interaction showed that the deviant was different from the standard in both language groups; post-hoc test following up the stimulus by time interaction indicated that the deviant was more negative than the standard in all time windows between 160-240 ms. To examine whether the MMN was present for the /gupa/ -/gypa/ contrast in all language/ISI groups, two-way ANOVAs on each of the four groups were conducted using stimulus and time as factors. The deviant was more negative than the standard for all four subgroups ($ps < 0.05$). Post hoc tests following significant interactions of stimulus and time revealed that MMN was present in all four time intervals between 160 and 240 ms for the Mandarin long ISI group, and between 180 and 240 ms for the English long ISI group. No stimulus by time interaction was observed for the short ISI English group.

In summary, increased negativity of the deviant compared to the standard was observed for both groups at both ISIs for the /gypa/-/gupa/ contrast, for both groups at the long ISI for the /gipa/-/gupa/ contrast and for the Mandarin group at the short ISI for the /gipa/-/gupa/ contrast. Only the English group in the short ISI condition did not show an early increased negativity to the /gipa/ deviant. For the /gypa/–/gupa/ contrast, the Mandarin listeners in both ISI conditions demonstrated broader negativity encompassing all four time intervals. In contrast, for the English long ISI group, the increase in negativity is only significant from 180-240 ms for the /gypa/-/gupa/.

V4.3.2. Late Time Interval: 380 – 460 ms

Standard vowel /gupa/ versus deviant vowel /gipa/

Table V7 shows the results of the four-way mixed ANOVAs. There was a significant main effect of stimulus type ($p < 0.001$). The deviant was generally more negative than the standard. Significant interaction of time by stimulus was also observed ($p = 0.01$). Post hoc tests showed that the deviant was more negative than the standard in all four time windows. The difference between deviant and standard appeared to be larger in the short ISI groups than in the long ISI groups, however, ANOVA result showed that the ISI main effect is not significant ($p = 0.08$). No significant main effect or interactions including group were found. The interactions that did not include stimulus were not of interest and thus were not examined further.

It was important to determine whether each group showed a robust difference between the deviant and standard stimulus in the two ISI conditions. Thus, ANOVAs with stimulus and time as factors were conducted for each of the four subgroups. Main effects of stimulus were obtained for all four groups (ps < 0.01). An interaction of stimulus by time was only observed for the Mandarin short ISI group. Post hoc tests showed that the two stimulus types differed in all four time windows, but the largest difference was observed between 360 and 380 ms.

In summary, increased negativity of the deviant was present for all four language/ISI groups to the /gipa/ deviant between 360-440 ms.

Standard vowel /gupa/ versus deviant vowel /gypa/

Table V8 provides a summary of the statistical analyses for the late time intervals. The four-way mixed ANOVA revealed a main effect of stimulus ($p < 0.01$) with the deviant /gypa/ being generally more negative than the standard /gupa/. Significant interactions of stimulus by time ($p = 0.02$), time by language ($p = 0.04$), and stimulus by time by ISI ($p < 0.01$) were also observed. Post hoc tests following the stimulus by time interaction revealed that the deviant was more negative than the standard in all four time intervals. Figure V4 shows that the differences were largest between 360 and 380 ms. Step-down tests following the three-way interaction of stimulus, time and ISI by examining the two language groups under each ISI condition separately revealed that under the short ISI condition, the deviant was more negative than the standard between 380-420 ms (F(3,90) = 3.086, p = 0.03, ε = 0.64), and under the long ISI condition, the deviant was more negative than the standard in all four time intervals (F(1,29) = 7.81, p = 0.009). The ISI main effect was approaching significant (p = 0.052) with the short ISI groups in the trend of being more negative.

ANOVAs for each language group in each ISI condition were carried out to determine whether there were significant differences for the subgroups. These analyses revealed no significant main effects or interactions for either language group in the short ISI conditions. A time by stimulus interaction was observed for the Mandarin group in the long ISI condition (p = 0.03). Post hoc test indicated that the standard and deviant differed between 380-440 ms. A main effect of stimulus type was found for the English long ISI group (p = 0.02), with the deviant stimulus being more negative than the standard. Table V8 provides a summary of the statistics for these analyses.

In summary, increased negativity of the deviant stimulus was found in general, but the follow-up analyses showed that it was robustly present only in the two long ISI groups.

V4.3.3. ISI and deviant type effect on the latency and amplitude of the early negativity

Table V9 provides the mean amplitudes and standard deviations of the subtraction waves in the MMN time window. Table V10 describes the details of the statistical analyses. In terms of the latency of the greatest negativity in the early time frame, the only significant effect found was a deviant stimulus type main effect (p <0.001). The peak latency of the /gypa/ condition was later than that of the /gipa/ deviant condition (/gypa/ mean = 191 ms, SD = 28 ms; /gipa/ mean = 131 ms, SD = 32 ms). In terms of the amplitude of the difference wave, there were main effects of language group, and stimulus, and an interaction of stimulus by group. Larger negativity was observed in the waveforms of the Mandarin listeners (/gipa/ mean = - 0.70, SD = 0.566; /gypa/ mean = - 1.19, SD = 0.66) than the English listeners (/gipa/ mean = - 0.60, SD = 0.569; /gypa/ mean = - 0.73, SD = 0.568). In addition, the subtraction wave for the /gypa/ condition generated larger negativity than that for the /gipa/ condition. Post hoc tests following up the stimulus type by group interaction revealed that the amplitude of the subtraction waves for the /gipa/-/gupa/ condition did not differ for the groups. In contrast, the Mandarin listeners had larger negativity than the English listeners for the /gypa/-/gupa/ contrast. No ISI main effect or interaction was observed.

V4.3.4. ISI and deviant type effect on the latency and amplitude of the late time interval

Table V11 showed group average and standard deviant of the subtraction waves within the LN time window. Table V12 provides the results of the statistical analyses. In terms of latency effect, the only significant effect was ISI (p = 0.03), with the short ISI conditions generated longer LN latency (short ISI /gipa/ mean = 407 ms, SD = 20 ms; short ISI /gypa/ mean = 416 ms, SD = 21; long ISI /gipa/ mean = 402 ms, SD = 22.3 ms; long ISI /gypa/ mean = 401 ms, SD = 23.9 ms). The only effect that was related to amplitude was a main effect of stimulus type, with

/gipa/ deviant type generated larger negativity than the /gypa/ type (/gipa/ mean = - 0.85μV, SD = 0.62μV; /gypa/ mean = -0.53μV, SD = 0.73μV; p < 0.001).

In summary, larger negativities were observed for the /gipa/ than for the /gypa/ condition in general. LN latency was longer in the short ISI condition than in the long ISI condition.

V4.4. Token differences within each category

Two tokens per deviant category and three tokens in the standard /gupa/ category were used in the ERP experiment. Figure V5 depicts the ERP waveforms from each token per category from the Mandarin short ISI group. It appeared that /gipa/ token 1 generated robust negativity in the N1 time window, while /gipa/ token 2 was more negative than the standard in the P2 time window. This difference in the amplitudes of N1 and P2 for the two tokens may account for the smaller negative shift when averaging across the ERPs to these two tokens when presented as deviants. For the two /gypa/ tokens, the difference between the deviant and standard was more similarly distributed from P1 to N2 region for both tokens.

To examine whether each token was labeled consistently across participants, individual ANOVAs (token, ISI and language) were performed for each of the three stimulus types using arcsine transformed accuracy scores. Only effects related to token variable are reported here. There was a main effect of token (F (1,59) = 104, p < 0.000) and group (F(1,59) = 10.8, p = 0.002) for the /gipa/ category. Post hoc test revealed that the Mandarin listeners had higher identification scores than the English listeners labeling /gipa/ as intended, and both language groups labeled /gipa/ token one as intended more often than /gipa/ token two. A main effect of token for the /gupa/ category ($F(1,59) = 7.30$, $p < 0.01$), and an interaction of token by group for the /gypa/ category ($F(1,59) = 5.05$, $p < 0.05$) were found. Post hoc tests revealed that the Mandarin listeners had similar identification scores for the two gypa tokens, but the English listeners categorized one token more often as /gypa/ than the other token. Table V13 presents the identification probability for each token.

V5. DISCUSSION

V5.1. Main findings

We examined behavioral and neurophysiologic correlates of vowel processing under different memory trace decay conditions, allowing us to explore the role of long-term memory representation in this process. We hypothesized that having a native phonemic representation in long-term memory would ameliorate trace decay, and thus allow better behavioral discrimination and more robust brain responses even under unattended conditions. Results from the behavioral discrimination experiment supported this hypothesis. Indeed, Mandarin listeners could behaviorally discriminate both of the native vowel contrast pairs with near-ceiling accuracy, regardless of ISI. Similar to Mandarin listeners, English listeners demonstrated high accuracy discriminating /gipa-gupa/ phonemic contrast under both ISI conditions. However, there is a steep decrement in performance in discriminating the non-native contrast, /gypa-gupa/, especially in the long ISI condition. A similar pattern was found for identification, with English listeners performing similarly to the Mandarin listeners for categorizing the /gupa/ tokens, but substantially worse for categorizing /gypa/ tokens. In addition, the English listeners show much poor categorization of one of the /gipa/ tokens compared to the Mandarin listeners.

The ERP results also support the hypothesis that language-specific memory traces modulate auditory memory trace decay. The increased negativity to the deviant conditions in the early time window (100-240 ms) was consistent in timing and topography with the MMN. This MMN was observed for Mandarin listeners at long and short ISIs for both deviant types and for English listeners for both deviant types at the long ISI, but only for the /gypa/-deviant at the short ISI. As predicted, the Mandarin groups showed larger MMN than the English groups regardless of ISI for the /gypa/ deviant condition. There was no language group difference in terms of the MMN peak amplitude or latency for the /gipa/ condition, as expected.

For the LN, we had predicted that the LN would not be attenuated for the non-native group. Significant negativity was evidenced in the later time window between 380 and 460 ms for all four language/ISI groups for the /gipa/-deviant conditions, but only for the long ISI /gypa/-deviant conditions. The frontocentral topography and latency of this negativity were consistent with that of LN reported in previous studies (e.g., Shafer et al., 2005; Ortiz-Mantilla et al., 2010). The prediction regarding the LN was somewhat upheld, although it was actually larger for the non-native English listeners at the long ISI for both deviant types. For the Mandarin listeners, the amplitude of the LN did not change with increasing ISI. Below, these results will be discussed in greater detail.

V5.2 Duration of auditory memory and language experience

The current study showed that extensive experience with a contrast type modulates the duration of auditory memory for the contrast information. Given that all ISI/language groups showed significant MMN except the English short ISI /gipa/ condition, it appears that the duration of auditory sensory memory for speech, or vowel contrast, specifically, is longer than reported in the behavioral literature (e.g., 1.5 s for consonantal contrast in Werker & Logan, 1985; 2 s for short vowels in Pisoni, 1973), but shorter than the auditory memory for nonspeech sound such as pure auditory tone. (e.g., > 10 s in Sams et al., 1993). Behavioral paradigms differ from the ERP paradigms in that behavioral judgments are dependent on a number of factors including attention skills, listener's response strategies, tradeoff between accuracy and reaction time. Another factor that affects the ISI effect disparity between consonant behavioral studies (e.g., Werker & Logan, 1985; Pisoni, 1973), vowel behavioral studies (e.g., Pisoni, 1973), the lexical tone study in chapter 2, and the current vowel ERP study is that there are some intrinsic differences among consonant, vowel and lexical tone categories. Early studies have found that vowels are perceived in a more continuous manner, much like nonspeech whereas consonant discrimination showed a pattern of marked discontinuity at points along the acoustic continuum that correspond to behavioral identification changes (Fry, Abramson, Eimas & Liberman, 1962; Pisoni, 1973). In another word, within-category vowel differences are more available for behavioral discrimination than within-category consonant differences. Pisoni (1973) found discrimination accuracy dropped continuously for within-category vowel contrasts as a function of ISI between 0.25 and 2 s, and within-category consonant discrimination accuracy stayed low regardless of ISI changes. The current study showed that within-category vowel discrimination is poor with an ISI of 2.6 s. This indicates that there is a difference between vowel perception and pure auditory tone perception. Vowel, as a speech category, is acoustically more complex on one hand, and overlearned and language-specific on the other hand. These differences may explain that findings from the pure tone frequency differences may not apply to vowel perception although both vowel and pure tone frequency contrast have the feature of steady-state acoustic differences.

We found that ISI clearly affected the LN amplitude. The LN for /gypa/-/gupa/ contrast was absent under the short ISI conditions, but present under the long ISI conditions. We propose that the presence of LN is associated with significant sensory memory trace decay. When ISI goes beyond 2.5 s, the degradation of the echoic sensory memory imposes a greater cognitive demand, and recruitment of long-term memory or other compensatory strategies have to be employed to achieve discrimination. More discussion of LN was presented later.

We chose the ISI of 2.6 s as the long ISI condition to examine the duration of the auditory sensory memory based upon several considerations. First of all, the behavioral discrimination literature reported that an ISI of 1500 ms is sufficient to shift speech processing mode from the phonetic to the phonemic mode. To our knowledge, studies that have focused on the duration information in sensory memory and using an MMN paradigm have not used speech as stimuli except two studies (Čeponienė et al., 1999; lexical tone study of this dissertation). Based upon results from these studies, the sensory memory duration in healthy adults lasts longer than 3 s for pure auditory tones that differ in frequency, ranging from 10% to 50%. However, the duration of sensory memory for various speech information has rarely been examined. Speech is acoustically and functionally much more complex than pure tone, and it has a more direct relationship with language experience. The current study provided information regarding sensory memory decay for vowel contrasts in terms of both behavior and neurophysiology. The findings confirmed that native language experience allows for maintenance of sufficient information for discriminating the /i/ vs. /u/ contrasts. In contrast, the information maintained for the /y/ stimuli was insufficient to allow the non-native group to discriminate it from Mandarin /u/. The neurophysiology added insight to this finding by showing that when SOA lengthens to 3 s or when the contrast is difficult, vowel processing relies more on the long-term memory representation, and LN appears to be a more sensitive index than MMN.

In healthy adults, an MMN can be present under conditions with an ISI up to about 10- to 30 seconds, but the MMN was shown to reduce in amplitude as the ISI increased (Böttcher-Gandor & Ullsperger, 1992; Mäntysalo & Näätänen, 1987; Sams, Hari, Rif & Knuutila, 1993; Winkler et al., 2002). These studies provided important data on human being's basic auditory sensory memory processing mechanism. However, pure auditory tones differ from complex stimuli in a number of ways. In particular, the relevance of pure tone stimuli (when not part of a melody) is quite different from other complex auditory stimuli, such as speech. The current study revealed that differences in experience with speech information (native versus non-native) modulate the time course of sensory memory decay for this information in the system indexed by MMN. It will be important in future studies to examine memory trace decay of other types of complex (environmental sounds) and/or relevant auditory information (e.g., music) to determine whether the decay rate for well-learned speech categories is comparable to other well-learned auditory categories.

V5.3. Stimulus variable: acoustic-phonetic differences versus phonemic differences

The current study found a striking difference in both behavioral and ERP responses between the native phonemic contrasts and the nonnative phonetic contrast (/gupa/ versus /gypa/ for the English group) using natural speech tokens. Specifically, higher discrimination accuracy for /gipa-gupa/ contrast than /gypa-gupa/ contrast for the English group, and larger MMN for the Mandarin groups than for the English groups for the /gypa –gupa/ contrast were obtained. These findings are in line with the studies using synthesized speech (e.g., Winkler et al., 1999). Studies have shown that human listeners have some degree of sensitivity to the acoustic differences between nonnative phonemic contrasts (Miyawaki, Strange, et al., 1975) even though they might be “mistuned” with regards to the nonnative language because they pay attention to a different set of acoustic cues that are relevant for the first language (Iverson et al., 2003). Many cross-language studies have used synthetic speech to allow strict control over the variance of acoustic parameters. However, natural speech produced by human speakers is notoriously variable, and the multiple acoustic parameters that are exploited by native listeners to differentiate the phonemic categories are still not entirely understood. The perceptual patterns observed for highly controlled synthetic speech may not reflect the reality of everyday speech perception. Thus, a more ecologically valid task, as promoted by Strange and Shafer (2008), is the use of natural speech. In the current study, we used two tokens per deviant category, and three tokens for the standard category of naturally produced bisyllabic nonwords. These tokens were selected from a large pool of recordings based on careful listening and detailed acoustic analysis of formant frequencies, F0 contour, overall amplitude and duration, duration and amplitude of each segment within the syllable, and to ensure that phonetically irrelevant acoustic variability was not highly correlated with the phonetically relevant acoustic variability, but to allow for the natural variability found in everyday speech. By implementing such natural speech, we hoped to tap into phonemic processing to a greater extent than found in processing synthetic speech without introducing so much variability that the “noise” masked the contrast of interest. The similarity in behavioral and MMN patterns of response indicate that use of natural speech can be effective. Both deviant types /gi/ and /gy/ are illegal syllables in Mandarin, and /gipa/, /gypa/ and /gupa/ are all nonwords in both Mandarin and English. So phonotactic probabilities are controlled in the stimuli.

V5.4. Language experience and behavioral responses

The behavioral findings of the current study were consistent with previous studies, in that they revealed reliance of phonemic levels of processing at longer ISIs. Previous studies using behavioral methods have proposed that speech perception may be influenced by several different factors, such as psychoacoustic auditory, language-general phonetic, and language-specific phonemic factors (Burnham et al., 1996; Pisoni, 1973; Werker & Tees, 1984; Werker & Logan, 1985) depending on the rate of sound presentation. Using an AX discrimination task, Werker and

Logan (1985) found that when two stimuli were presented with an ISI of 250 ms, American-English (AE) listeners was able to discriminate two different CV syllable tokens that both fell within a single Hindi phonetic category, or across the two non-native categories of dental and retroflex stop consonants. At a longer ISI of 1500 ms, poor performance was observed for American (AE) listeners to the cross-category Hindi contrast. But Hindi listeners maintained good categorization performance. Werker and Logan (1985) interpreted these results as evidence of engaging three different levels of perception. They proposed that under conditions of high stimulus uncertainty and memory load listeners rely on language-specific categories, while in less demanding task conditions (e.g., low memory demand) discrimination and categorization of speech information can be based on language-general phonetic properties.

Our behavioral discrimination experiment differs slightly from the AX paradigms used by previous studies (Burnham et al., 1996; Pisoni, 1973; Werker & Tees, 1984; Werker & Logan, 1985). We adopted a modified version of our ERP oddball paradigm ($A_1A_1A_2A_1X$ or $A_1A_2A_1A_1X$) for the purpose of examining correlations between behavioral and neurophysiological responses. Instead of using 500 ms versus 1500 ms, we used an SOA of 900 ms versus 3000 ms (equivalent to ISI of about 575 ms and 2675 ms, on average). Therefore, our long ISI condition is significantly longer than the ISI of 1500 ms. The reason for using a longer ISI is based on the result of our pilot studies by looking at the ERP response. It appears that an ISI of 1500 ms is inadequate to observe a difference in the ERP responses, suggesting a dissociation between behavioral and neurophysiological measures under certain conditions. Furthermore, we are comparing vowel differences instead of consonant or lexical tone difference in this study, and the inherent differences among different phonetic categories may account for the ISI needed to observe clear memory demand difference.

In general, our behavioral discrimination results support the previous findings in that under long ISI conditions, listeners have to rely on their native phonemic categories for discriminating speech contrast. In our study, both the Mandarin and English listeners discriminated the /gipa-gupa/ contrast with similar high accuracy under both short and long ISI conditions. However, for the /gypa-gupa/ contrast, the English listeners have lower accuracy under the short ISI condition, and more than 25% of participants were at chance level under the long ISI condition. This pattern suggests that English listeners, indeed, were using language-general phonetic information for discrimination under the short ISI condition, but could not employ this information at the longer ISI. Thus, at the longer ISI, English listeners appear to have assimilated the /y/ vowel into the /u/ phonemic category. This finding was predicted and is consistent with Best’s Perceptual Assimilation Model (Best, 1994; Best & Tyler, 2007). Further examining the identification results showed that /gipa/ was only occasionally mislabeled as /gupa/ (4% and 15% of total errors for the Mandarin and English groups, respectively) in either language group, while labeling /gypa/ as /gupa/ accounted for 96% of total errors for the English listeners, and only 57% of the errors were due to mislabeling /gypa/ as /gupa/ in the Mandarin listeners. So the behavioral identification result provided further evidence of the /y/ -/u/ assimilation patterns in American English listeners. All the stimuli were produced by a native Mandarin speaker, the high discrimination and identification accuracy for /gipa/ and /gupa/ suggested the presence of two-category (TC) assimilation for the English listeners based on the PAM (Best, 1994; Best &

Tyler, 2007). The findings in this study supported and strengthened the PAM by providing data that tap into phonemic perception to a greater extent than previous behavioral speech perception studies have achieved.

V5.5. Language experience and MMN responses

Our findings are consistent with previous studies that have shown smaller MMNs for non-native compared to native listeners of a contrast (Aaltonen et al., 1987; 1994; Dehaene-Lambertz, 1997; Dehaene-Lambertz & Bailet, 1998; Philips, Marrantz, McGinnis, Pesetsky, Wexler, & Yellin, 1995; Shafer, Schwartz, & Kurtzberg, 2004; Sharma & Dorman, 1999, 2000; Sussman, Kujula, Halmetoja, Lyytinen, Alku, & Näätänen, 2004; Winkler et al., 1999; see Näätänen et al., 2007 for a review). Based on these studies, we expected a language group difference for the /gypa/-/gupa/ contrast under the short ISI condition because previous studies have shown that MMN is larger for phonemic contrast than for within-category phonetic contrast. We did not know whether this pattern of findings would be maintained, or possibly accentuated, when ISI was lengthened to 2.5 s for vowel contrasts because no study has examined this question previously. Consistent with previous research, our results did show that the Mandarin listeners had larger MMNs than the English listeners across both ISI conditions for this Mandarin-only phonemic contrast. In addition, the English listeners showed a significant MMN for /gypa-gupa/ contrast under both ISI conditions. This finding suggests that English listeners could detect the phonetic differences between these vowels even thought they did not make use of this information for discrimination on the long ISI condition. In other words, the durability and quality of the memory trace underlying MMN generation was adequate to allow for discrimination of the /y/-/u/ contrast, but at a behavioral level, English participants were less successful making use of this information.

We included /gipa/-/gupa/ vowel contrast in this study and expected robust MMN for both language/ISI groups given that this is a phonemic contrast in both English and Mandarin. MMN was present earlier and smaller than that for the /gypa/ condition, and were absent in the English short ISI condition. Increased negativity was found for the later (LN) time interval for both groups. Several reasons might account for such a finding. First, we used natural speech and the acoustic properties of the speech may have resulted in overlapping obligatory responses that masked the MMN (see Fig.V5). A second possibility was that there was too much latency variability among individuals as latency jitter is known to cause problems in the averaged data (Luck & Hillyard, 1990; Luck, 2005, p.135). From Figure V5, we can see that the difference between the standard /gupa/ and deviant /gipa/ token A is mostly in the N1 time window, whereas for the second /gipa/ token, the difference was mostly in the P2 time window. Therefore, in the averaged data, the MMN effect is reduced. When we use the peak amplitude taken from the four consecutive time interval of 20 ms, and compared this value with hypothetic zeros using paired t-test, MMN is significant for the English short ISI groups ($p < 0.000$) as well as for the other groups. A third possibility is that one of the two tokens was a poorer exemplar of /i/ as one token was identified as /gipa/ more often than the other token for all the groups, and the gap between the two identification rates were the largest in the English short ISI group (see table

V13, identification rate of 0.88 for token 1 and 0.54 for token 2). We examined categorization of the separate tokens and found that there is a difference in labeling behavior for the two /gipa/ tokens and the two /gupa/ tokens in general, and the two /gypa/ tokens in the English listeners only. The Mandarin listeners categorized /gipa/ token two as intended with 75% probability, and while the English listeners only 55%. Both language groups labeled /gipa/ as intended well above chance level (> 33%), and all groups had very high discrimination accuracy for /gipa/-/gupa/ contrast (> 88%). Thus, we do not consider this might be the major reason. Furthermore, as expected, the two language groups do not differ under the long ISI condition for /gipa-gupa/ contrast, both show significant LN, and this result suggested that the memory traces underlying discrimination were similar for common phonemic contrast between English and Mandarin listeners.

V5.6. Language experience and LN responses

The discriminative ERP response to the speech contrasts in the current study included LNs for most conditions. The late negativity (LN) is often generated in an oddball paradigm, following the MMN response, but it is less often analyzed and discussed. Barry, Bishop and colleagues reported reduced LN in adults and children with language impairment who scored lower in nonword repetition task, and they claimed that LN is an index of formation of phonological representation (Barry et al., 2009; Bishop, Hardiman & Barry, 2010). However, these findings were at odds with some other studies in several respects. Unlike many studies of disordered groups, both Barry et al. (2009) and Bishop et al. (2010) did not find any difference in MMN responses between a poor and good repeater group, or between children with or without SLI. The LNs in these two studies were smaller in the less proficient language users/learners. In contrast many studies of language/learning differences have found reduced MMN and larger LN in the less proficient language groups (e.g., family history of language impairment: Addis, Friederici, et al., 2010; children with SLI: Shafer, et al., 2005; second language learners: Ortiz-Mantilla et al., 2010). Therefore, the nature of LN still requires explanation.

In this study, we observed LN for /gipa-gupa/ condition for all language/ISI groups irrespective of presence/absence of MMN, and LN presence for /gypa-gupa/ under the long ISI conditions, and robust LNs were not present for the /gypa-gupa/ contrast under the short ISI conditions for either language groups, but in these cases robust MMNs were present. One possible explanation, considering all the previous findings, is that LN reflects additional recruitment of cognitive resources for further processing of the sound contrast, as it can be generated independent of the amplitude of MMN. Across studies, LN is sometimes absent when MMN is robust, especially when the task is easy, such as in this study. Short ISI conditions are easier than the long ISI conditions due to less memory trace decay, which may explain the absence of LN for /gypa/ deviant type for the two short ISI groups. One neuronal network operation principle is to minimize the cost (Bullmore & Sporn, 2012). Based on this principle, it is feasible to propose that MMN and LN represent a two-stage sequential processing. If the processing is sufficient and adequately automatic during the early processing time window as indexed by MMN, then no further processing is necessary, thus no LN will be elicited. For a difficult contrast, or challenging perceptual condition, the early automatic discrimination may or may not take place, and LN represents the recruitment of additional resources.

In Shafer et al. (2005), children with SLI showed good behavioral discrimination, but significantly poor identification, furthermore, children with SLI in this study also showed reduced/absence of MMN, but presence of LN. In this study, we expected /i/-/u/ contrast to be easier for all listeners than the /y/ - /u/ contrast, thus less robust LN at least in the short ISI conditions. However, the /i/-/u/ contrast may not be as easy as we predicted because the stimuli were produced with the Mandarin accent by a native Mandarin speaker who is a late English L2 learner, thus the /gipa/ in this study may not represent a prototypical token of English /i/, and one token of /gipa/ was less consistently labeled as /gipa/ by both the Mandarin and English listeners. The high discrimination accuracy and low identification rate in our behavioral results from the English listeners pattern with the behavioral results from the children with SLI in Shafer et al. (2005), the absence of MMN but presence of LN ERP responses in the /gipa-gupa/ conditions for the English short ISI group pattern with that of children with SLI as well. Shafer et al. (2005) proposed that the increased LN may indicate the recruitment of compensatory strategies in children with SLI. Similar to this proposal, Addis, Friederici and colleagues (2010) have investigated three generations of a German family with language impairment with an oddball paradigm and other psychometric and genetic tests, and found that the youngest family members showed absence of MMN and presence of LN, while the adult family members showed both MMN and an enhanced LN, compared to the presence of MMN but absence of LN in the control adults. Addis and colleagues proposed that LN reflects increased processing demand. The results from the present study supports the proposal that LN is associated with increased processing load and is therefore observed in the long ISI conditions and for difficult contrasts.

One alternative explanation is that LN is simply a later MMN. In this case, presence of two MMNs indicates that discrimination of the stimulus contrasts is based on different acoustic properties that result in timing differences for the response. Manipulation of the acoustic properties of stimuli, as well as other factors that increase task difficulty (e.g., background noise) will be necessary to help chose between these explanations. However, based on the finding that LN was present for the easy tone and vowel /i-u/ contrast, and absent for the hard tone and /y-u/ contrast under short ISI conditions, we proposed that it is unlikely that LN observed in this study is just a later MMN.

V6. Conclusion

This is the first neurophysiological study examining the role of long-term memory representations for vowel processing under condition of auditory sensory memory trace decay. We did find evidence supporting the existing literature that native language experience enhances speech discrimination at the pre-attentive level. We also found new evidence that native language experience affect different ERP components (here, MMN and LN) in a different way. Reduced MMN and presence of LN as a result of auditory sensory memory trace decay for vowel contrasts were evidenced. By using different ISI, we clarified that native language experience plays a role in echoic sensory memory trace maintenance. The pattern of results in the present study can be used to evaluate specific hypotheses concerning mechanisms underlying MMN dysfunction in memory disorders such as schizophrenia, children with phonological working memory deficit, or language proficiency changes in general.

Appendix Two: Bibliography for vowel study

Aaltonen, O., Eerola, O., Lang, A.h., Uusipaikka, E. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *The Journal of the Acoustical Society of America*, *96*(3), 1489-1493.

Aaltonen, O., Niemi, P., Nyrke, T., Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology, 24,* 197-207.

Addis,L., Friederici, A.D., Kotz, S.A., Sabisch, B., Barry, J., Richter, N., Ludwig, A.A., Rubsamen, R., Albert, F.W., Paabo, S., Newbury, D.F., Monaco, A.P. (2010). A locus for an auditory processing deficit and language impairment in an extended pedigree maps to 12p13.31-q14.3. *Genes, Brain, and Behavior, 9*(6), 545-561.

Amenedo, E., Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioral perception. *The European Journal of Neuroscience, 12*(7), 2570-2574.

Barry, J.G., Hardiman, M.J., Bishop, D.V. (2009). Mismatch response to polysyllabic nonwords: A neurophysiological signature of language learning capacity. *PLoS One, 4*(7), e6270.

Barry, J.G., Hardiman, M.J., Line, E., White, K.B., Yasin, I., Bishop, D.V. (2008). Duration of auditory sensory memory in parents of children with SLI: A mismatch negativity study. *Brain and Language, 104*(1), 75-88.

Best, C.T.(1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 171-204). Timonium, MD: York Press.

Best, C.T., & Tyler, M. C. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.S. Bohn & M.J. Munro (eds.), Language experience in second language speech learning: In honor of James Emil Flege, Amsterdam/Philadelphia: John Benjamins, 13-34.

Bishop, D.V., Hardiman, M.J., & Barry, J.G. (2010). Lower-frequency event-related desynchroization: a signature of late mismatch responses to sounds, which is reduced or absent in children with specific language impairment. *The Journal of Neuroscience, 30*(46), 15578-15584.

Böttcher-Gandor, C., Ullsperger, P. (1992). Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiology, 29*, 546-550.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews. Neuroscience, 13*(5), 336-349.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: evidence for a linguistic mode of processing. The 4th International conference on Spoken Language Processing (ICSLP 96), Philadelphia, USA, *4,* 2514-2517.

Čeponienė, R., Cheour, M., &Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children : Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology, 108*(4), 345-354.

Čeponiene R., Service, E., Kurjenluoma, S., Cheour, M., &Näätänen, R. (1999). Children's performance on pseudoword repetition depends on auditory trace quality: Evidence from event-related potentials. *Developmental Psychology, 35*(3), 709-720.

Čeponienė, R., Lepistö, T., Soininen, M., Aronen, E., Alku, P., & Näätänen, R. (2004). Event-related potentials associated with sound discrimination versus novelty detection in children. *Psychophysiology, 41*(1), 130-141.

Čeponienė, R., Service, E., Kurjenluoma, S., Cheour, M., & Näätänen, R. (1998). Children's performance on pseudoword repetition depends on auditory trace quality: Evidence from event-related potentials. *Developmental Psychology, 35*(3), 709-720.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8*(4), 919-924.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *Neuroreport, 9*(8), 1885-1888.

Datta, H., Shafer, V.L., Morr, M.L., Kurtzberg, D., & Schwartz, R.G. (2010). Electrophysiological indices of discrimination of long-duration, phonetically similar vowels in children with typical and atypical language development. *Journal of Speech and Hearing Research, 53*(3), 757-777.

Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H.G. Jr. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Developmental Psychology, 35*(1), 294-302.

Fry, D.B., Abramson, A.S., Eimas, P.D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*, 171-189.

Grier, J.B. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin, 75*(6), 424-429.

Haenschel, C., Vernon, D.J., Dwivedi, P., Gruzelier, J.H., Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience, 25*(45), 10494-10501.

Imada, T., Hari, R., Loveless, N., McEvoy, L., & Sams, M. (1993) Determinants of the auditory mismatch response. *Electroencephalography and Clinical Neurophysiology, 87*(3), 144-153.

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), B47-57.

Javitt, D.C., Grochowski, S., Shelley, A.M., & Ritter, W. (1998). Impaired mismatch negativity (MMN) generation in schizophrenia as a function of stimulus deviance, probability, and interstimulus/interdeviant interval. *Electroencephalography and Clinical Neurophysiology, 108*(2), 143-153.

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.

Korpilahti, P., Krause, C.M., Holopainen, I., Lang, A.H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language, 76*(3), 332-339.

Lang, H.A., Nyrke, T., Market, Ek M, Aaltonen O, Raimo I, &Näätänen R. (1990). Pitch discrimination performance and auditory event-related potentials. In: Brunia CHM, Gaillard AWK, Kok A, Mulder G, Verbaten MN, editors. Psychophysiological brain research, vol. 1. Tilburg, the Netherlands: Tilburg University Press; 1990. p. 294–8.

Levy, E.S. & Strange, W. (2008). Perception of French vowels by American English adults with and without French language. *Journal of Phonetics, 36,* 141-157.

Lipski, S.C., Escudero, P., & Benders, T. (2012). Language experience modulates weighting of acoustic cues for vowel perception: an event-related potential study. *Psychophysiology, 49*(5), 638-650.

Luck, S.J. (2005). *An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience*). Gazzaniga, M.S. (Ed.). Cambridge, MA: The MIT Press.

Luck, S.J. & Hillyard, S.A. (1990). Electrophysiological evidence for parallel and serial processing during visual search. *Perception & Psychophysics, 48*(6), 603-617.

Mäntysalo, S., Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology, 24*(3), 183-195.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native Japanese and English. *Perception & Psychophysics, 18,* 331-340.

Näätänen, R., Gaillard, A.W., Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(2), 313-329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vaino, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 118*(12), 2544-2590.

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.

Ortiz-Mantilla, S., Choudhury, N., Alvarez, B., & Benasich, A.A. (2010). Involuntary switching of attention mediates differences in event-related responses to complex tones between early and late Spanish-English bilinguals. *Brain Research, 1362,* 78-92.

Peltola, M.S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters, 352*(1), 25-28.

Philips, C., Marrantz, A., McGinnis, M., Pesetsky, D., Wexler, K., & Yellin, E. (1995). Brain mechanisms of speech perception: a preliminary report. In Schutze, T., Ganger, J.B., & Broihier, K. (Eds.) *Papers on language processing and acquisition* (pp. 153-191). Cambridge, MA: MIT Press.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253-260.

Rinne, T., Särkkä, A., Degerman, A, Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research, 1077*(1), 135-143.

Sams, M., Ahlo, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology, 17*(1), 41-58.

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory trace persists about 10 seconds: Neuromagnetic evidence. *Journal of Cognitive Science, 5*(3), 363-370.

Sams, M., Paavilainen, P., Alho, K, & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology, 62*(6), 437-448.

Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience, 8,* 527-539.

Schröger, E., & Wolff, C. (1998).Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport, 9 (15),* 3355-3358.

Shafer, V.L., Schwartz, R.G., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Brain Research. Cognitive Brain Research, 18*(3), 242-254.

Shafer, V.L., Morr, M.L., Datta, H., Kurtzberg, D., & Schwartz, R.G. (2005) Neurophysiological indexes of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neuroscience, 17*(7), 1168-1180.

Sharma, A. & Dorman, M.F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106*(2), 1078-1083.

Sharma, A. & Dorman, M.F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107*(5 Pt 1), 2697-2703.

Shestakova, A., Huotilainen, M., Čeponienė, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology 114*(8), 1507-1512.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*(4), 456-466.

Strange, W. & Shafer, V.L. (2008). Speech perception in second language learners: The re-education of selective perception. Hansen-Edwards, J.G. & Zampini, M.L. (Eds.). *Phonology and second language acquisition.* Philadelphia, NJ: Johns Benjamins.

Sussman E, Ritter W, Vaughan Jr, HG. (1998). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport, 9,* 4167–70.

Sussman, E., Ritter, W., Vaughan Jr., H.G., (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology 36*, 22–34.

Sussman, E., Sheridan, K., Kreuzer, J., & Winkler, I. (2003). Representation of the standard: Stimulus context effects on the process generating the mismatch negativity component of event-related brain potentials.*Psychophysiology, 40*, 465-71.

Sussman, E., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *NeuroReport, 16*(13), 1519–1523.

Sussman, E., Kujula, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research, 190*(1-2), 128-140.

Sussman, E. (2007). A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology, 21*(3-4), 164-175.

Sussman, E. S., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research, 236*(1-2), 61-79.

Tiitinen, H., Sinkkonen, J., May, P., & Näätänen, R. (1994). The auditory transient 40-Hz response is insensitive to changes in stimulus features. *Neuroreport, 6*(1), 190-192.

Werker, J.F. & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustical Society of America, 75, 1866-1878.

Werker, J.F. & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, V., Ilmoniemi, R.J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*(5), 638-642.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M. Czigler, I., Csépe, V., Aaltonen, O. Raimo, I., Alho, K., Lang, H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Brain Research. Cognitive Brain Research, 7*(3), 357-369.

Winkler, I., & Cowan, N. (2005). From sensory to long-term memory: Evidence from auditory memory reactivation studies. *Experimental Psychology, 52*(1), 3-20.

Winkler, I., & Cowan, N. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch. *Journal of Cognitive Neuroscience*, *8*(5), 403.

Ylinen, S., Streinikov, K., Huotilainen, M., & Näätänen, R. (2009). Effects of prosodic familiarity on the automatic processing of words in the human brain. *International Journal of Psychophysiology, 73*(3), 362-368.

Zevin, J.D., Datta, H., Maurer, U., Rosania, K.A., & McCandliss, B.D. (2010). Native language experience influences the topography of the mismatch negativity to speech. *Frontiers in Human Neuroscience, 4,* 212.

Appendix Three: Vowel Study Tables

Table V1. Participants. Age, sex and handedness information for the four participant groups.

Participant group		Age(range, SD)	N (Sex)	Handedness
English	Long ISI	28.4 (20-42, 6.6)	15(8M, 7F)	1 LH,14 RH
	Short ISI	29.8 (22-41, 5.6)	16(7M,9F)	1 LH, 15 RH
Mandarin	Long ISI	29.1 (23-40, 5.3)	16(9M,7F)	all RH
	Short ISI	25.9 (21-36, 4.5)	16(8M, 8F)	1 ambidextrous, 15 RH

Appendix Three: Vowel Study Tables

Table V2. The event-related potential paradigm (One standard and two deviant conditions. Only the trials in bold were included in the analyses)

Standard	Deviant
gupa	Gypa Gipa

Sample structure
gu3pa ***gu3pagu3pa*** gu2pa gu3pa ***gu3pagi3pa*** gu3pa ***gu3pa***gu1pa gu3pa ***gy3pa*** gu3pa ***gu3pa*** gu2pa gu3pa ***gy3pa***gu3pa ***gu3pa***gu1pa gu3pa ***gi3pa*** gu3pa gu1pa gu3pa ***gu3pagy3pa***

Appendix Three: Vowel Study Tables

Table V3. Mixed effects logistic regression on vowel discrimination.

Fixed effect	Beta Estimate	S.E.	Z value	Pr
(Intercept)	3.4060	0.384	8.873	< 2e-16 ***
Stimulus	0.9909	0.335	-2.96	0.0003**
Language	0.767	0.547	1.40	0.16
ISI	-0.383	0.407	-0.94	0.347
Stimulus* language	-1.736	0.467	-3.716	0.0002***
Lang*ISI (English long)	-1.169	0.529	-2.21	0.03*

Significance codes: '***' for $p < 0.001$; '**' for $p < 0.01$; '*' for $p < 0.05$

Appendix Three: Vowel Study Tables

Table V4. Mixed effects logistic regression on vowel identification.

	Fixed effect	Beta Estimate	S.E.	z value	Pr
Main model	(Intercept)	4.74	0.42	11.2	< 2e-16 ***
	/gipa/	-2.56	0.41	-6.185	6.19e-10 ***
	/gypa/	-2.80	0.33	-8.37	< 2e-16 ***
	Language	-1.60	0.27	-5.84	5.22e-09***
	ISI	0.35	0.27	1.28	0.20
Sub-model for /gypa/-/gipa/ only	(Intercept)	1.88	0.25	7.36	11.8-13***
	/gypa/	0.75	0.40	1.87	0.06
	Language	-1.03	0.31	-3.35	0.0008***
	ISI	0.29	0.28	1.04	0.30
	Stimulus*Language	-1.75	0.53	-3.3	0.0009***

Significance codes: ‘***’ for $p < 0.001$; ‘**’ for $p < 0.01$; ‘*’ for $p < 0.05$

Appendix Three: Vowel Study Tables

Table V5. ANOVA results comparing standard /gupa/ with deviant /gipa/ between 100 and 180 ms (Stim = stimulus conditions; Man = Mandarin, Eng = English, Lang = language, T = time).

Group	Effect	MMN(100-180 ms)				
		DF	F	P	ε(G-G)	η_{ρ}^{2}
All	ISI	1,59	15.5	0.0002		0.21
	Lang	1,59	0.24	n.s.		
	ISI*Lang	1,59	0.152	n.s.		
	Stim	1,59	7.407	0.009		0.11
	Stim*ISI	1,59	0.024	n.s.		
	Stim *Lang	1,59	2.087	n.s.		
	Stim *ISI*Lang	1,59	0.001	n.s.		
	T*ISI	3,177	20.785	0.000	0.37	0.25
	T*Lang	3,177	0.413	n.s.		
	T*ISI*Lang	3,177	0.308	n.s.		
	Stim *T	3,177	11.250	0.000	0.76	0.16
	Stim *T*ISI	3,177	3.182	0.038	0.76	0.05
	Stim *T*Lang	3,177	0.940	n.s.		
	Stim*T*ISI*Lang	3,177	0.315	n.s.		
ManShort	Stim	1,15	4.871	0.043		0.245
	Stim *T	3,45	4.923	0.015	0.65	0.247
ManLong	Stim	1,15	2.783	n.s.		
	Stim *T	3,45	4.629	0.02	0.61	0.236
EngShort	Stim	1,15	0.547	n.s.		
	Stim *T	3,45	1.931	n.s.		
EngLong	Stim	1,14	0.542	n.s.		
	Stim *T	3,42	4.99	0.02	0.51	0.263

Appendix Three: Vowel Study Tables

Table V6. ANOVA results comparing standard /gupa/ with deviant /gypa/ between 160 and 240 ms (Stim = stimulus conditions; Man=Mandarin, Eng = English, Lang = language, T = time).

Group	Effect	MMN(160-240 ms)				
		DF	F	P	ε(G-G)	η_ρ^2
All	ISI	1,59	14.2	0.0004		0.193
	Lang	1,59	0.016	n.s.		
	ISI*Lang	1,59	0.62	n.s.		
	Stim	1,59	80.719	0.000		0.58
	Stim*ISI	1,59	0.012	n.s.		
	Stim *Lang	1,59	9.994	0.002		0.145
	Stim *ISI*Lang	1,59	0.055	n.s.		
	T*ISI	3,177	30.490	0.000	0.49	0.34
	T*Lang	3,177	3.976	0.03	0.49	0.06
	T*ISI*Lang	3,177	2.509	n.s.	0.49	
	Stim *T	3,177	5.609	0.01	0.49	0.09
	Stim *T*ISI	3,177	1.284	n.s.	0.49	
	Stim *T*Lang	3,177	1.554	n.s.	0.49	
	Stim*T*ISI*Lang	3,177	3.117	0.06(n.s.)	0.49	0.05
ManShort	Stim	1,15	38.928	0.000		0.72
	Stim *T	3,45	1.151	n.s.	0.43	
ManLong	Stim	1,15	34.510	0.000		0.70
	Stim *T	3,45	3.409	0.056(n.s)	0.56	0.185
EngShort	Stim	1,15	21.316	0.000		0.59
	Stim *T	3,45	1.583	n.s.	0.45	
EngLong	Stim	1,14	4.676	0.05.		0.25
	Stim *T	3,42	8.303	0.004	0.51	0.37

Appendix Three: Vowel Study Tables

Table V7. ANOVA results comparing standard /gupa/ with deviant /gipa/ between 380 and 460 ms (Stim = stimulus conditions; Man = Mandarin, Eng = English, Lang = language, T = time).

Group	Effect	LN (380-460 ms)				
		DF	F	P	ε(G-G)	η_ρ^2
All	ISI	1,59	3.01	0.08		
	Lang	1,59	0.23	n.s.		
	ISI*Lang	1,59	0.378	n.s.		
	Stim	1,59	59.8	0.000		0.50
	Stim*ISI	1,59	1.793	n.s.		
	Stim *Lang	1,59	0.117	n.s.		
	Stim *ISI*Lang	1,59	0.673	n.s.		
	T*ISI	3,177	7.696	0.002	0.48	0.12
	T*Lang	3,177	3.810	0.04	0.48	0.06
	T*ISI*Lang	3,177	1.711	n.s.		
	Stim *T	3,177	5.098	0.01	0.52	0.08
	Stim *T*ISI	3,177	0.309	n.s.		
	Stim *T*Lang	3,177	0.795	n.s.		
	Stim*T*ISI*Lang	3,177	1.504	n.s.		
ManShort	Stim	1,15	19.5	0.000		0.56
	Stim *T	3,45	3.280	n.s.		
ManLong	Stim	1,15	7.660	0.01		0.34
	Stim *T	3,45	0.307	n.s.	0.55	
EngShort	Stim	1,15	16.5	0.001		0.52
	Stim *T	3,45	1.440	n.s.	0.32	
EngLong	Stim	1,14	33.3	0.0000		0.70
	Stim *T	3,42	2.370	n.s.		

Appendix Three: Vowel Study Tables

Table V8. ANOVA results comparing standard /gupa/ with deviant /gypa/ between 380 and 460 ms (Stim = stimulus conditions; Man = Mandarin, Eng = English, Lang = language, T = time).

Group	Effect	LN (380-460 ms)				
		DF	F	P	ε(G-G)	η_ρ^2
All	ISI	1,59	3.95	0.052		
	Lang	1,59	0.14	n.s.		
	ISI*Lang	1,59	0.41	n.s.		
	Stim	1,59	9.787	0.003		0.14
	Stim*ISI	1,59	1.103	n.s.		
	Stim *Lang	1,59	0.000	n.s.		
	Stim *ISI*Lang	1,59	0.329	n.s.		
	T*ISI	3,177	2.379	n.s.		
	T*Lang	3,177	3.766	0.04	0.47	0.06
	T*ISI*Lang	3,177	0.435	n.s.		
	Stim *T	3,177	4.141	0.02	0.64	0.07
	Stim *T*ISI	3,177	5.059	0.008	0.64	0.08
	Stim *T*Lang	3,177	0.322	n.s.		
	Stim*T*ISI*Lang	3,177	0.271	n.s.		
ManShort	Stim	1,15	1.650	n.s		
	Stim *T	3,45	1.500	n.s.		
ManLong	Stim	1,15	2.450	n.s.		
	Stim *T	3,45	4.023	0.03	0.69	0.21
EngShort	Stim	1,15	0.790	n.s.		
	Stim *T	3,45	2.160	n.s.		
EngLong	Stim	1,14	6.520	0.02		0.32
	Stim *T	3,42	2.763	n.s.		

Appendix Vowel Study

Table V9. Amplitudes of the subtraction waves (deviant minus standard) from the composite FzCz in the MMN time window.

		Gipa MMN				Gypa MMN			
		100-120	120-140	140-160	160-180	140-160	160-180	180-200	200-220
Eng Long	Mean	-0.41**	-0.21*	0.08	0.25	0.272*	-0.017	-0.371*	-0.579**
	SD	0.61	0.45	0.51	0.72	0.527	0.637	0.733	0.722
Eng Short	Mean	-0.24	-0.01	0.05	-0.20	-0.106	-0.419**	-0.54***	-0.41***
	SD	0.60	0.68	0.62	0.67	0.666	0.607	0.570	0.385
Mand Long	Mean	-0.51**	-0.28	-0.14	-0.15	-0.37**	-0.80***	-0.98***	-0.83***
	SD	0.57	0.77	0.79	0.76	0.600	0.566	0.664	0.652
Mand Short	Mean	-0.39*	-0.24	-0.14*	-0.39**	-0.396**	-0.72***	-0.91***	-0.86***
	SD	0.67	0.62	0.53	0.55	0.419	0.519	0.645	0.608

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Appendix Three: Vowel Study Tables

Table V10. Comparing the MMN for the two deviant types. ANOVA results using subtraction waves examining the latency and amplitude effect within 100-240 ms time window (stim = /gipa/ versus/gypa/ stimulus conditions; Lang = language)

Effect		MMN Latency			MMN Amplitude		
	DF	F	p	η_{ρ}^{2}	F	p	η_{ρ}^{2}
ISI	1，59	0.895	n.s.		0.09	n.s.	
Lang	1，59	0.398	n.s.		4.82	0.03	0.08
ISI*Lang	1，59	1.531	n.s.		0.03	n.s.	
Stim	1，59	112.7	0.000	0.65	13.6	<0.001	0.187
Stim *ISI	1，59	2.751	n.s.		0.001	n.s.	
Stim *Lang	1，59	1.223	n.s.		4.64	0.036	0.07
Stim *ISI*Lang	1，59	0.599	n.s.		0.007	n.s.	

Appendix Three: Vowel Study Tables

Table V11. Amplitudes of the subtraction waveforms (deviant minus standard) from the composite FzCz in the LN time window.

		Gipa LN				Gypa LN			
		380-400	400-420	420-440	440-460	380-400	400-420	420-440	440-460
Eng Long	Mean	-0.79***	-0.88***	-0.76***	-0.57***	-0.49**	-0.53**	-0.39*	-0.30*
	SD	0.61	0.65	0.57	0.48	0.687	0.670	0.685	0.673
Eng Short	Mean	-0.42**	-0.53***	-0.46***	-0.33**	-0.028	-0.190	-0.187	-0.130
	SD	0.537	0.483	0.457	0.485	0.562	0.742	0.684	0.488
Mand Long	Mean	-0.64**	-0.55**	-0.56**	-0.57**	-0.46*	-0.40*	-0.264	-0.189
	SD	0.899	0.871	0.818	0.917	0.756	0.864	0.922	0.893
Mand Short	Mean	-0.60***	-0.60***	-0.50***	-0.31*	-0.151	-0.305	-0.344	-0.163
	SD	0.465	0.480	0.541	0.607	0.691	0.803	0.871	0.820

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Appendix Three: Vowel Study Tables

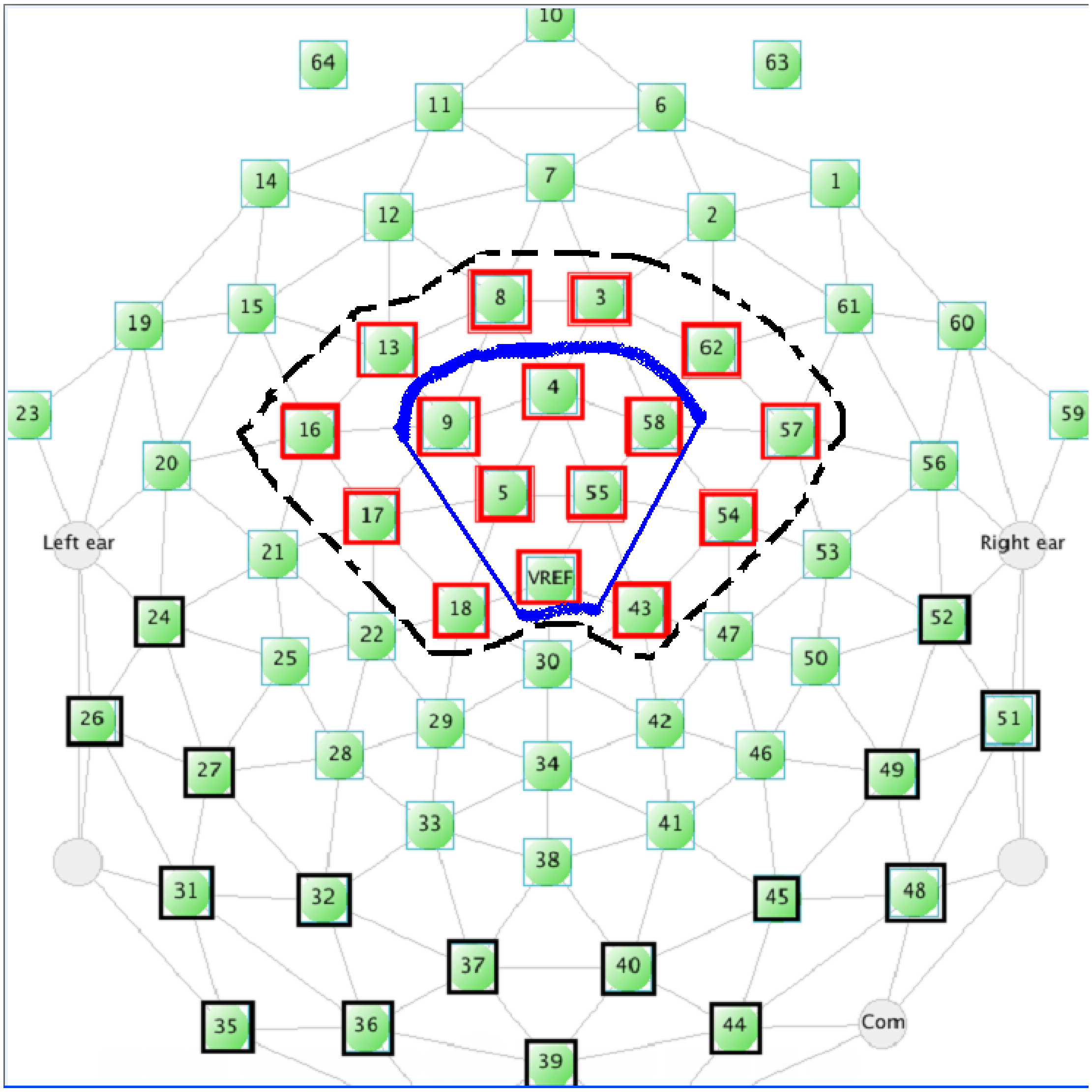
Table V12. Comparing the LN for the two deviant types. ANOVA results using subtraction waves examining the latency and amplitude effect within 380-460 ms time window (stim = /gipa/ versus/gypa/ stimulus conditions; Lang = language)

Effect		LN Latency			LN Amplitude		
	DF	F	p	η_ρ^2	F	P	η_ρ^2
ISI	1,59	4.58	0.03.	0.07	1.72	n.s.	
Lang	1,59	0.13	n.s.		0.056	n.s.	
ISI*Lang	1,59	0.23	n.s.		1.11	n.s.	
Stim	1,59	1.04	n.s.		13.3	<0.001	0.184
Stim *ISI	1,59	2.60	n.s.		0.03	n.s.	
Stim *Lang	1,59	0.42	n.s.		0.188	n.s.	
Stim *ISI*Lang	1,59	0.74	n.s.		0.001	n.s.	

Appendix Three: Vowel Study Tables

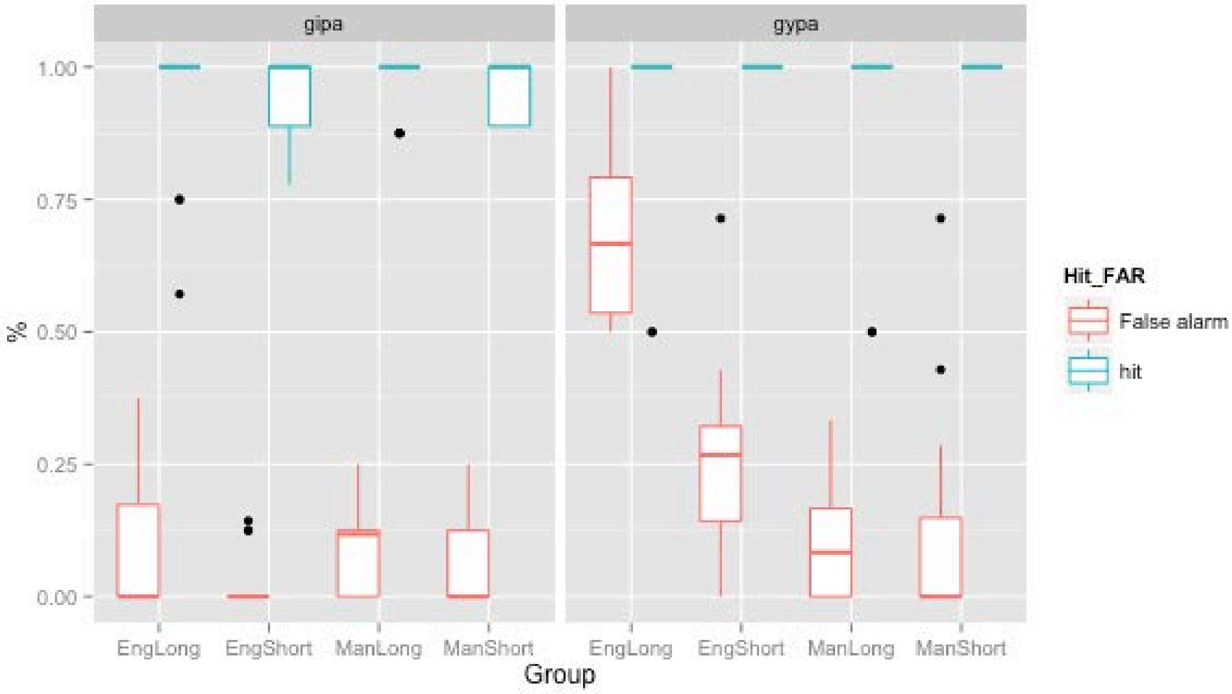
Table V13. Identification accuracy by token.

Group		gi3pa313c	gi3pa336c	gu3pa326	gu3pa341	gy3pa342a	gy3pa349f
Overall	Mean	0.9	0.65	0.92	0.87	0.65	0.71
	SD	0.20	0.23	0.21	0.21	0.33	0.33
EngLong	Mean	0.8	0.56	0.88	0.8	0.4	0.6
	SD	0.33	0.3	0.28	0.25	0.29	0.35
ManLong	Mean	0.97	0.77	1	0.96	0.91	0.9
	SD	0.07	0.14	0	0.06	0.15	0.18
EngShort	Mean	0.88	0.54	0.86	0.82	0.48	0.56
	SD	0.14	0.19	0.2	0.25	0.23	0.37
ManShort	Mean	0.93	0.73	0.95	0.9	0.8	0.78
	SD	0.13	0.22	0.22	0.22	0.31	0.3



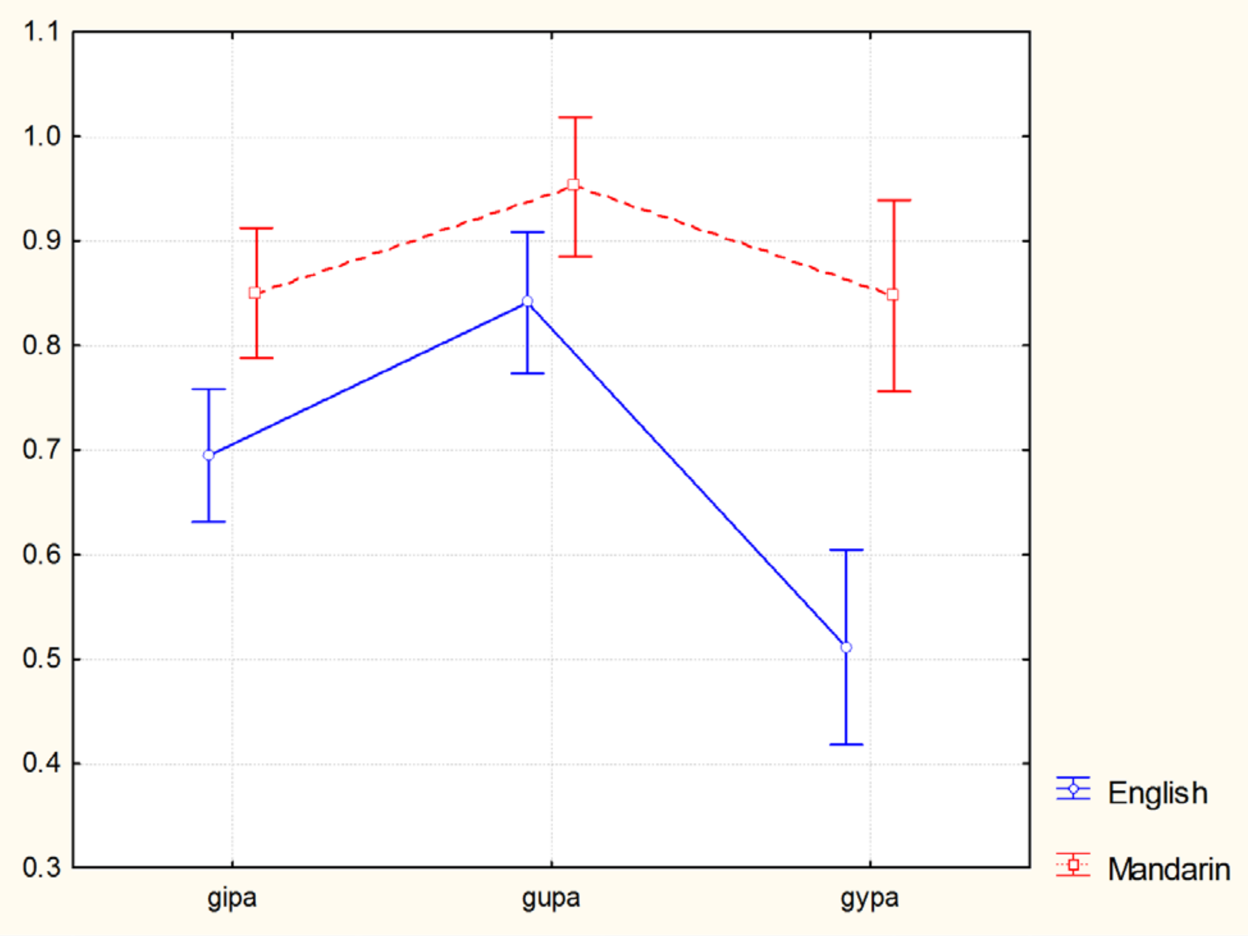
Appendix Four: Vowel Study Figures

Figure V1. The six frontocentral sites that are within the blue line enclosed area were used to build the composite FzCz site. These sites are Fz/4, 5, 9, 55, 58 and Cz/65. For comparison, the electrodes sites within the dashed black line were used to build the composite Fz site in paper 1.



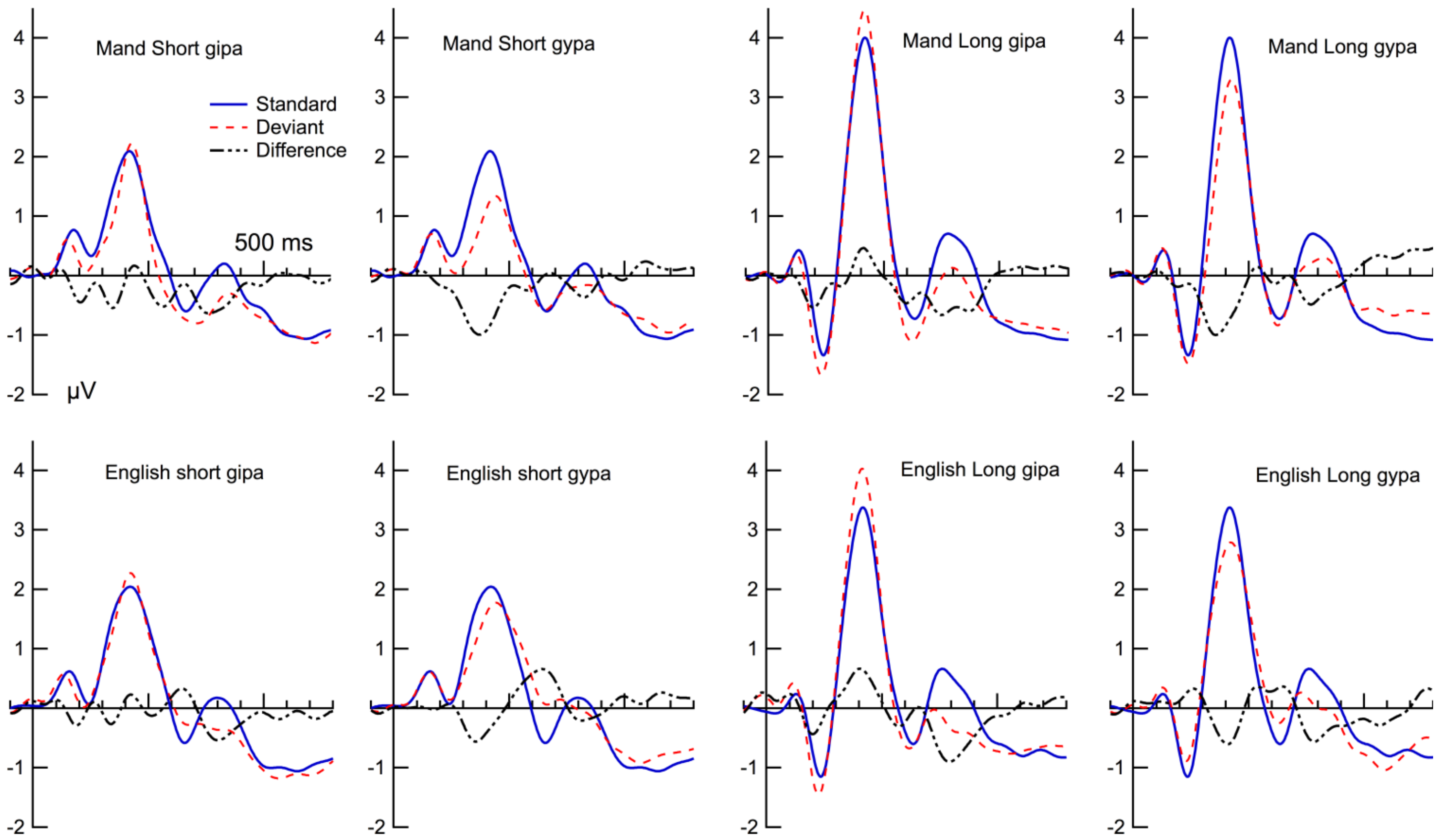
Appendix Four: Vowel Study Figures

Figure V2. Hit rate and false alarm rate for vowel discrimination for the two language groups under two ISI conditions.



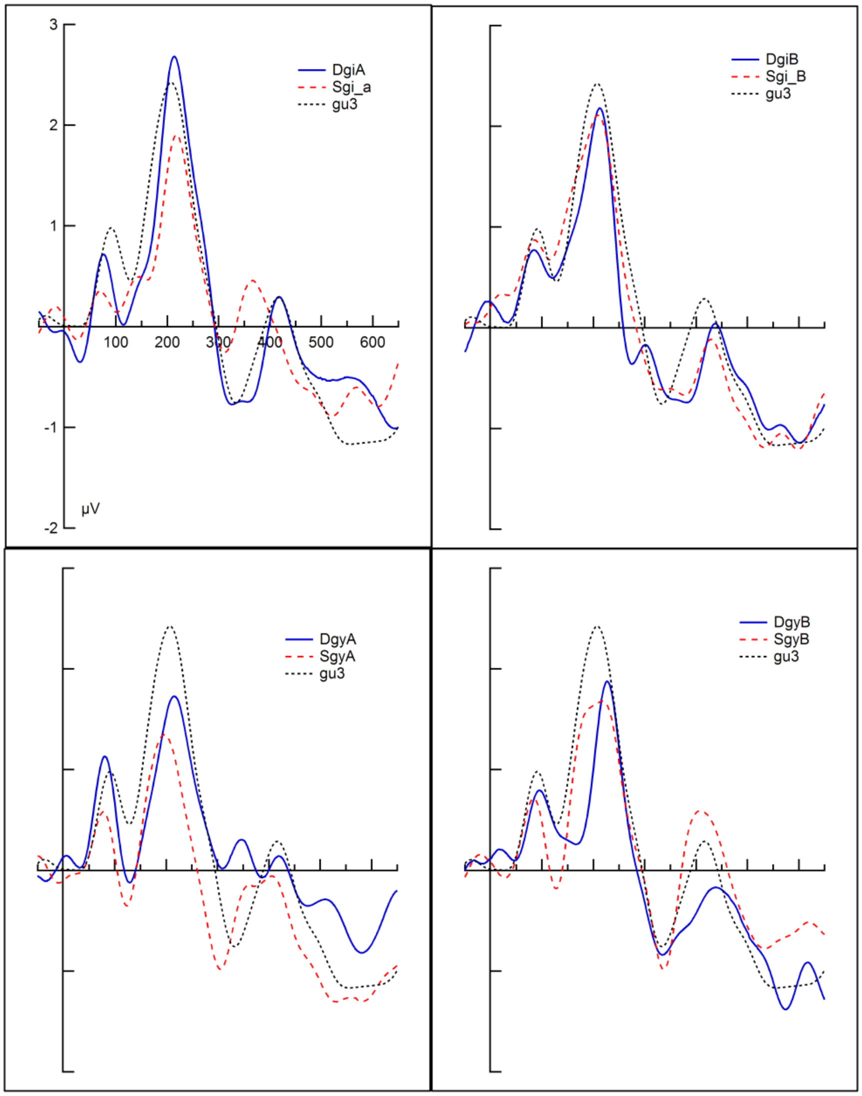
Appendix Four: Vowel Study Figures

Figure V3. Accuracy for vowel identification. Chance level is at 33.3%.



Appendix Four: Vowel Study Figures

Figure V4. The ERP responses at the composite FzCz site. The blue solid trace is the standard /gu3pa/, and the red dashed trace is the vowel deviant (either /gi3pa/ or /gy3pa/). The dotted black waveform is the difference between the standard and the deviant.



Appendix Four: Vowel Study Figures

Figure V5. Token data from 15 Mandarin listeners under the short ISI condition. The blue waves are the responses under the deviant condition, and the black dotted waves are from the standard /gupa/ condition. The red dotted waves are from the stimulus repetition alone condition, in which each token was played 100 times in a separate block at the end of the ERP experiment, with the same ISI (SOA = 900 ms) as in the oddball paradigm The upper panel figures are waveforms from the two /gipa/ tokens, and the lower panel figures are waveforms from the two /gypa/ tokens.

Appendix Five: Additional Summary and Discussion

The current studies were designed to examine how different factors such as degrees of sensory memory trace decay, long-term language experience, and acoustic distinctiveness of stimulus contrasts modulate lexical tone processing, and how the neural correlates of lexical tone processing differ from that of vowel processing. We had hypothesized that long-term language experience is necessary for speech discriminative processing under conditions when the degree of sensory memory trace decay is large, and lack of long-term language experience will lead to less robust neural responses as indexed by smaller MMN and different LN peak amplitude.

The results of the tone study were consistent with our hypotheses in that MMN responses were observed in the English groups when ISI was short, but not present in the English groups when ISI was long ISI. The LN amplitudes in the Mandarin long ISI groups were the same as or larger than those in the short ISI groups for both deviant tone types. For the English listeners, a larger LN was observed in the long ISI groups than in the short for T3/T1 contrast, but for T3/T2 contrast, no LN in either ISI conditions was present.

The results for the vowel study revealed that the/y/ -/u/ vowel contrast elicited smaller MMN for both of the English ISI groups than for the Mandarin groups, as predicted. The MMN for the /i/ - /u/ contrast was absent in the English short ISI condition, and the LNs for the /y/- /u/ contrast were absent in the short ISI condition for both language groups.

The results compared across the lexical tone and vowel studies, suggest that the neural indices of discrimination that are engaged in processing linguistically relevant F0 contour have a different pattern from the neural responses in vowel perception. Robust neural indices of discrimination (MMN and LN) for linguistically relevant tones were present in both Mandarin and English listeners when the ISI was short, and the auditory memory trace is strong, although the English listeners showed smaller amplitude responses than the Mandarin listeners. With the longer interval (ISI of 2675 ms), the neural responses as indexed by MMN and LN were either stronger or as strong as in the short ISI condition for tone and vowel /gipa/-/gupa/ contrasts in the Mandarin listeners. In contrast, the English listeners showed a difference between tone and vowel contrasts under the long ISI conditions. No MMN responses for tone contrasts were obtained, and the LN was observed only for the “easy” (T3/T1) contrast when the ISI was long. For nonnative vowel contrast (/gypa/-/gipa/), both MMN and LN were present under the long ISI conditions, although the MMN responses were smaller than those found for the Mandarin listeners.

Under the long ISI condition, reduced MMN amplitude and increased LN amplitude in the English listeners were observed for the more distinctive tone contrast (T3/T1)but in terms of vowel processing, this effect was subtle in the English listeners. The MMN amplitude decrease from the short to long ISI condition only approached significance for the /gypa/ deviant condition[1]. Larger LNs for the long ISI groups than for the short ISI groups were present for the /gypa/ deviant condition only.

The ERP discriminative response suggested that there is sufficient information to perceive the contrast difference even at the long ISI, but the behavioral data suggested that this information was not used effectively. The different pattern of findings for the behavioral and ERP measures has been observed in other studies (e.g., Shafer, Morr, Datta, Kurtzberg, &Schwartz, 2005). This disparity can be explained as resulting from measurement of different time points in processing. The neural measures tap into early stages of discriminative responses at the cortical and subcortical level, while the behavioral measures reveal an endpoint of

---

[1] The ANOVAs using subtraction waves, with ISI, language, four time intervals as factors showed that the ISI main effect was approaching significance (p = 0.07). This result was not reported in the vowel study section of the dissertation since only peak values were used to compare ISI effect and deviant type effect.

processing that requires attention and decision making among many other variables. The English listeners discriminated both tone and vowel contrasts at above chance levels under the short and long ISI conditions for both easy and hard deviant types. The absence of a clear MMN or LN response for the T3/T2 contrast in the English listeners suggests that attention to the stimulus in the behavioral task was necessary to resolve the difference. Hisagi and colleagues showed that attention led to an increase in MMN amplitude to a non-native vowel duration contrast (Hisagi, Shafer, Strange, & Sussman, 2010). Thus, we hypothesize that attention to the tone contrast would allow non-native English listeners to show discrimination of T2/T3 at the level indexed by MMN and/or LN.

The absence of LN to the nonnative linguistic pitch T3/T2 contrast and enhanced LN responses to the nonnative vowel /u/ –/y/ contrast may suggest that there are different mechanisms underlying lexical tone and vowel perception. In both chapters, we proposed two explanations for LN modulation. One possibility is that LN represents the recruitment of additional resource, and the other is that LN is by nature a later MMN that reflects discrimination based on different acoustic properties that occur at a different time window. For example, behavioral research suggested that for native tone language speakers discriminating linguistic tones is primarily dependent on the contours of the tone, while for nonnative listeners, it depends on the onset, or offset, or the average of F0 contour (Gandour & Harshman, 1978). The F0 contours of some tone contrast do not diverge until a later time window. In this dissertation, both T2 and T3 began low, but the significant F0 contour difference between the two tones emerged within 100 ms from the syllable onset. The later MMN peak latency in the T3/T2 condition than in the T3/T1 condition reflected this acoustic difference. If LN solely reflects that the perception of acoustic difference occurs later in time, then it would also be expected in the short ISI condition. The absence of LN in the short ISI condition for the easy contrast, together with the later MMN latency for the T3/T2 contrast, makes it less likely that LN is just simply a late MMN. The presence of LN appears to be associated with significant sensory memory trace decay. When the ISI is lengthened beyond 2.5 s, the depreciation of the echoic sensory memory imposes a greater cognitive demand, and recruitment of long-term memory or other compensatory strategies will have to be employed to achieve discrimination. The finding that this increase in LN occurred in a passive task suggests that some attentional resources were recruited, despite the instruction to ignore the stimuli. The study by Shafer and colleagues found evidence that typically developing children were shifting some attentional resources to processing vowel contrasts in a passive task (Shafer, Ponton, Datta, Morr & Schwartz, 2007). Thus, it seems reasonable to suggest that adults may also do this. Furthermore, it is possible that the general relevance of speech makes it more difficult for listeners to completely ignore the information without a demanding task to direct attention away from the stimuli (e.g., Hisagi et al., 2010).

In summary, pitch representation that allows for Mandarin lexical tone discrimination in native speakers of Mandarin is less susceptible to degradation at the neural level indexed by MMN than pitch representation in native English speakers. For non-tone language speakers, insufficient pitch information is available to offset sensory memory decay for Mandarin lexical tones. The findings for vowel contrasts differing in F1 and F2 formant frequencies also suggest that non-native experience is insufficient to support discrimination of the /u/ - /y/ contrast as indexed by poor behavioral discrimination accuracy and enhanced LN response.

The finding that nonnative tone discrimination worsens more rapidly than nonnative vowel discrimination for non-tonal language speakers is intriguing. Language-specific vowel processing is established before six months of age (e.g., Kuhl, Williams, Lacerda, Stevens, &

Lindblom, 1992; Dehaene-Lambertz & Dehaene, 1994; Shafer, Yu & Datta, 2011), while English-listening infants appear to lose sensibility in perceiving lexical tone difference later at around 9 months of age (Mattock & Burnham, 2006). Using formant frequency to distinguish vowels is a fundamental skill for both English and Mandarin listeners, despite the fact that F0 difference is not a reliable cue for English listeners to extract phoneme identity. For this reason, English listeners may be able to perceive vowel contrast with higher resolution if other factors are equal. Behaviorally, Yoruba (a tone language) listeners show more accurate discriminative responses to Thai tones than English listeners (Gandour & Harshman, 1978). In Kaan et al. study, although the ERP data were hard to interpret, behaviorally, the Mandarin listeners showed more sensitivity to Thai tones than the English listeners (Kaan, Barkley, Bao, & Wayland, 2008). Under this explanation, future research is required to examine whether non-native listeners whose native language includes phonemic tone show more robust neural representations (e.g., MMN and LN), and thus, less of an effect of ISI.

The use of longer ISIs, as demonstrated in this dissertation, may be applied to studies of language development, to clinical populations and would provide a better picture about the nature of phonological representations. So far, research examining sensory memory differences during development or in clinical populations (e.g., a patient with schizophrenia, a patient with chronic alcohol abuse) has only used pure auditory tones as stimuli. Using more complex and linguistically relevant stimuli is likely to provide a more sensitive measure. Further research should look into examining how language-specific sensory memory durability interacts with language proficiency.

Appendix Six: Bibliography for Additional Discussion and Summary

Dehaene-Lambertz, G., &Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*(6487), 292-295.

Gandour, J., & Harsham, R. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *The Journal of the Acoustical Society of America, 62,* 693-707.

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research, 11*(1360), 89-105.

Kaan, E., Barkley, C. M., 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.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., &Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science (New York, N.Y.), 255*(5044), 606-608.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

Shafer, V. L., Morr, M. L., Datta, H., Kurtzberg, D., & Schwartz, R. G. (2005). Neurophysiological indexes of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neuroscience, 17*(7), 1168-1180.

Shafer, V. L., Ponton, C., Datta, H., Morr, M. L., & Schwartz, R. G. (2007). Neurophysiological indices of attention to speech in children with specific language impairment. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 118*(6), 1230-1243.

Shafer, V.L., Yu, Y.H., & Datta, H. (2011). The Development of English Vowel Perception in Monolingual and Bilingual Infants: Neurophysiological Correlates. *Journal of Phonetics, 39(4):*527-545.

Appendix Seven Additional Tables

Table A1. Duration and intensity of the experimental stimuli

	Duration			Intensity		
	Overall	gV	Pa	Overall	gV	pa
gi3pa313c	314	114	200	71.4	70.8	71.8
gi3pa336c	336	113	223	67.9	72.1	66
gu1pa312a	291	115	175	71.7	69.3	73
gu1pa343b	343	124	199	72.1	71	72.6
gu2pa312c	312	115	197	70.8	72.5	70
gu2pa351	351	119	232	72.5	73.6	72.1
gu3pa320c	320	132	188	68.8	69.3	68.4
gu3pa326	326	139	187	70.9	70.6	71.1
gu3pa341	346	134	212	70.8	71.6	70.4
gy3pa342a	342	129	213	66.8	62.6	68.8
gy3pa349f	355	157	198	68.1	66.9	68.9

Appendix Seven Additional Tables

Table A2. Peak amplitude at nine sites between 100 and 350 ms for Tone 1

			MMN T1						
	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	-0.42	-0.55	-0.47	-0.54	-0.67	-0.65	0.49	0.38	0.53
EEL(SD)	*0.50*	*0.49*	*1.18*	*0.57*	*0.61*	*0.62*	*0.90*	*0.60*	*0.91*
EES	-0.63	-0.73	-0.67	-0.62	-0.75	-0.75	0.69	0.24	0.94
EES(SD)	*0.62*	*0.50*	*0.65*	*0.50*	*0.61*	*0.48*	*0.47*	*0.59*	*0.85*
MML	-0.89	-1.32	-0.69	-1.17	-1.32	-0.99	0.97	0.64	0.75
MML(SD)	*1.05*	*0.79*	*0.89*	*0.83*	*1.03*	*0.78*	*1.26*	*1.42*	*1.31*
MMS	-0.76	-1.12	-0.69	-0.75	-1.03	-0.56	0.70	0.45	1.20
MMS(SD)	*0.84*	*0.60*	*0.60*	*0.45*	*0.74*	*0.43*	*0.74*	*0.83*	*0.58*

Appendix Seven Additional Tables

Table A3. Peak amplitude at nine sites between 100 and 350 ms for Tone 2

	MMN T2								
	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	-0.61	-0.77	-0.52	-0.66	-0.91	-0.59	0.80	0.44	0.62
EE (SD)	*0.47*	*0.47*	*0.62*	*0.54*	*0.52*	*0.35*	*0.53*	*0.48*	*0.71*
EES	-0.81	-1.17	-0.90	-0.89	-1.05	-0.88	0.69	0.57	0.91
EES(SD)	*0.81*	*0.67*	*0.55*	*0.66*	*0.48*	*0.47*	*0.56*	*0.93*	*0.60*
MML	-0.79	-1.40	-1.07	-0.93	-1.34	-1.14	0.68	1.12	1.01
MML(SD)	*1.08*	*0.73*	*1.06*	*0.80*	*0.65*	*0.57*	*1.04*	*0.70*	*0.98*
MMS	-1.05	-1.19	-1.03	-0.87	-1.02	-0.88	1.31	0.89	1.27
MMS(SD)	*0.76*	*0.54*	*0.79*	*0.48*	*0.58*	*0.44*	*1.05*	*0.94*	*0.84*

Appendix Seven Additional Tables

Table A4. Peak latency at nine sites between 100 and 350 ms for Tone 1

MMN T1

	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	221	216	207	208	228	213	240	211	219
EEL(SD)	78	62	77	54	70	67	64	65	57
EES	213	209	231	193	200	199	206	198	213
EES(SD)	68	64	61	53	50	48	56	64	54
MML	213	214	229	231	226	224	241	220	223
MML(SD)	64	48	61	55	38	40	59	67	69
MMS	223	229	223	229	231	238	226	210	248
MMS(SD)	52	53	66	62	71	61	61	61	61

Appendix Seven Additional Tables

Table A5. Peak latency at nine sites between 100 and 350 ms for Tone 2

MMN T2

	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	232	213	228	209	221	212	228	223	209
EEL(SD)	66	55	62	55	56	51	61	47	59
EES	233	231	229	251	251	253	245	220	225
EES(SD)	59	50	43	56	52	41	52	50	44
MML	240	236	235	244	238	250	235	231	249
MML(SD)	61	47	56	53	49	45	55	62	56
MMS	255	250	256	250	258	256	253	256	258
MMS(SD)	54	53	59	53	53	53	46	52	48

Appendix Seven Additional Tables

Table A6. Peak amplitude at nine sites between 350 and 550 ms for Tone 1

			LN		T1				
	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	-0.69	-0.77	-0.85	-0.63	-0.96	-0.78	0.67	0.65	0.76
EEL(SD)	*0.90*	*1.00*	*1.41*	*0.63*	*0.76*	*0.78*	*1.34*	*1.05*	*1.14*
EES	-0.84	-0.74	-0.96	-0.61	-0.74	-0.55	0.89	0.56	0.80
EES(SD)	*0.78*	*0.36*	*0.93*	*0.54*	*0.47*	*0.38*	*0.69*	*0.79*	*0.83*
MML	-1.19	-1.43	-0.39	-1.15	-1.27	-0.81	1.16	0.59	1.15
MML(SD)	*1.20*	*0.75*	*1.21*	*0.83*	*0.78*	*0.62*	*0.98*	*1.30*	*1.07*
MMS	-0.72	-1.02	-0.44	-0.66	-1.15	-0.71	0.80	0.28	1.07
MMS(SD)	*1.03*	*0.66*	*0.70*	*0.51*	*0.94*	*0.56*	*0.82*	*0.83*	*0.68*

Appendix Seven Additional Tables

Table A7. Peak amplitude at nine sites between 350 and 550 ms for Tone 2

	LN T2								
	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	-0.53	-0.46	-0.30	-0.37	-0.66	-0.50	0.67	0.39	0.39
EEL(SD)	*0.67*	*0.41*	*0.95*	*0.39*	*0.33*	*0.23*	*0.73*	*0.59*	*0.80*
EES	-1.02	-0.68	-0.93	-0.68	-0.63	-0.56	0.47	0.38	0.39
EES(SD)	*0.26*	*0.86*	*0.75*	*0.54*	*0.54*	*0.49*	*0.53*	*0.76*	*0.54*
MML	-1.11	-0.98	-0.75	-0.92	-0.92	-0.85	0.33	0.86	1.03
MML(SD)	*0.98*	*0.81*	*1.31*	*0.71*	*0.78*	*0.84*	*1.12*	*1.05*	*1.03*
MMS	-0.83	-0.94	-0.90	-0.54	-0.83	-0.66	1.43	0.62	1.23
MMS(SD)	*0.61*	*0.68*	*1.08*	*0.47*	*0.63*	*0.66*	*1.41*	*1.28*	*1.10*

Appendix Seven Additional Tables

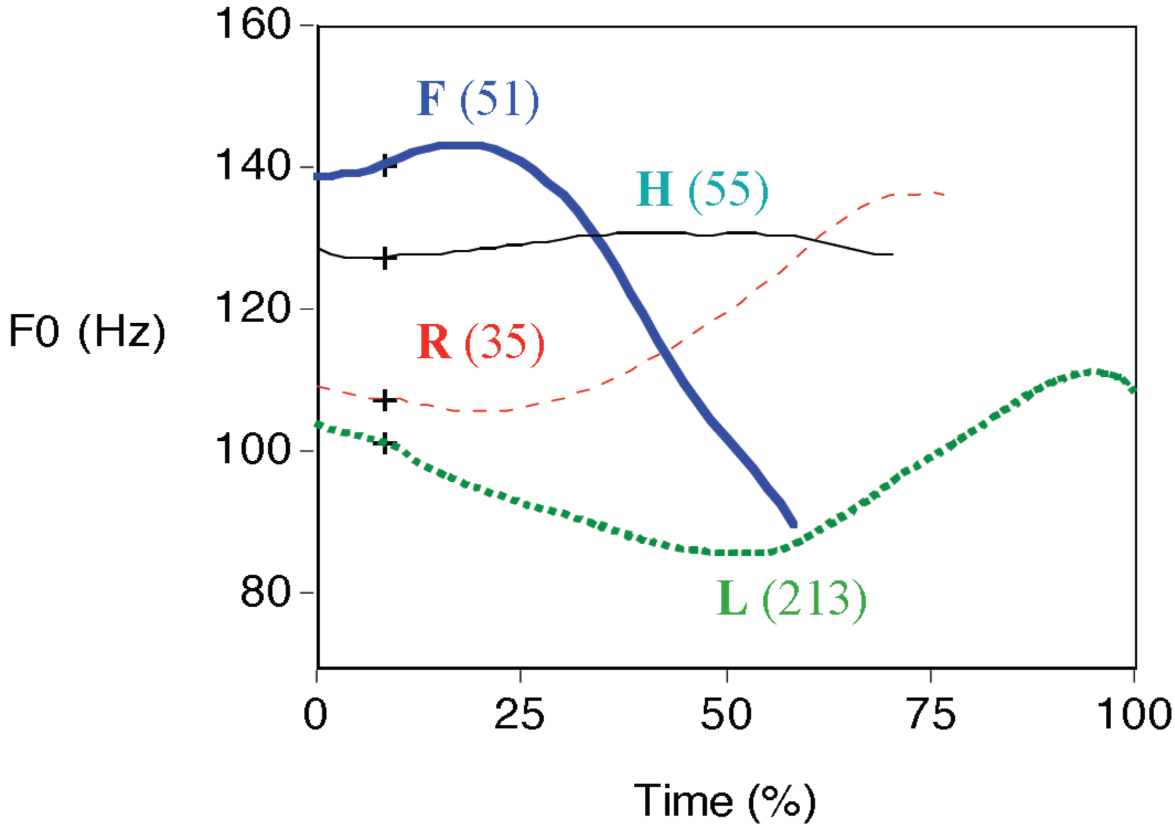
Table A8. Peak latency at nine sites between 350 and 550 ms for Tone 1

			LN	T1					
	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	445	452	456	427	443	433	432	463	431
EEL(SD)	48	55	57	40	52	45	61	53	48
EES	450	459	465	461	440	469	460	460	454
EES(SD)	58	54	48	50	55	41	38	62	35
MML	459	449	460	446	439	439	465	444	450
MML(SD)	44	42	45	47	31	35	57	59	47
MMS	474	449	468	455	463	461	479	463	464
MMS(SD)	57	53	59	60	38	29	46	60	53

Appendix Seven Additional Tables

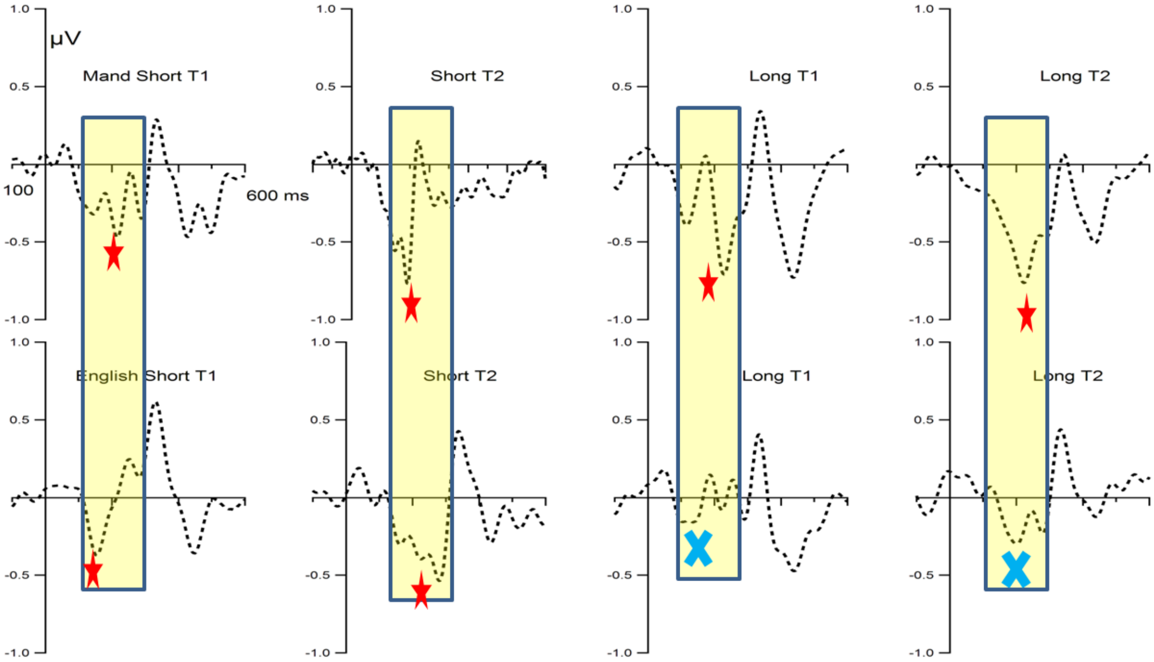
Table A9. Peak latency at nine sites between 350 and 550 ms for Tone 2

			LN	T2					
	F3	Fz	F4	C3	Cz	C4	LM	Oz	RM
EEL	431	413	459	428	433	427	452	444	445
EEL(SD)	49	45	70	43	52	58	54	42	39
EES	461	481	478	463	458	440	468	465	469
EES(SD)	62	49	58	61	60	48	49	56	55
MML	443	431	434	419	430	438	459	429	434
MML(SD)	52	50	75	42	47	53	61	68	49
MMS	459	470	435	459	456	449	463	438	459
MMS(SD)	63	56	66	70	56	64	65	59	55



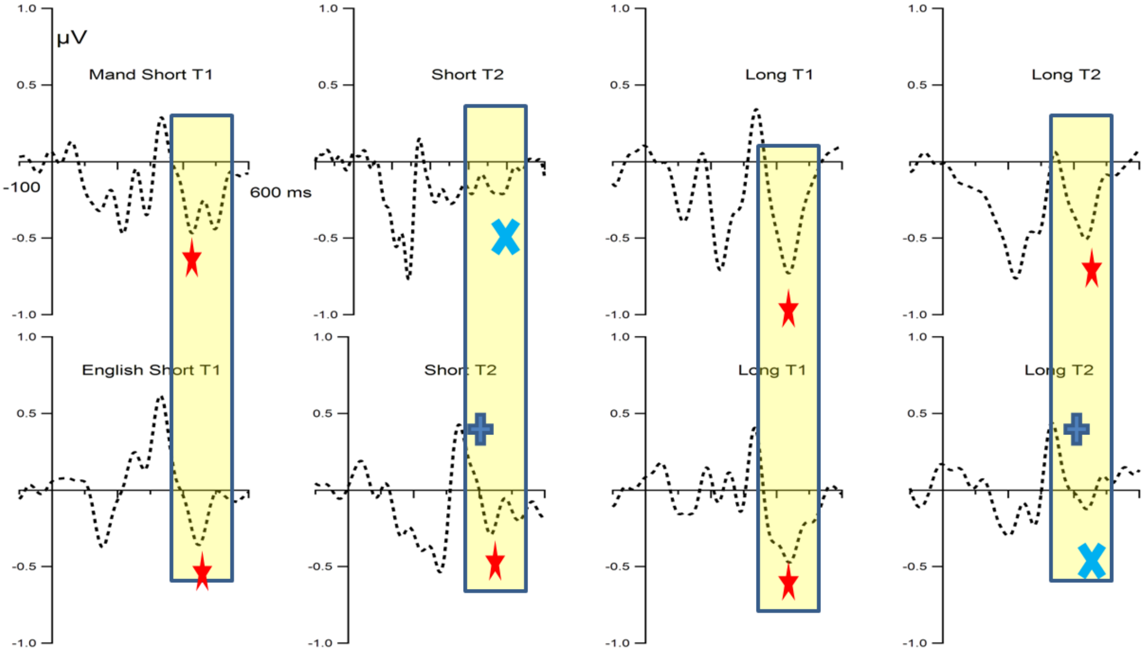
Appendix Eight Additional Figures

Figure A1. Mandarin tone produced in isolation. Modified from Xu (1997).



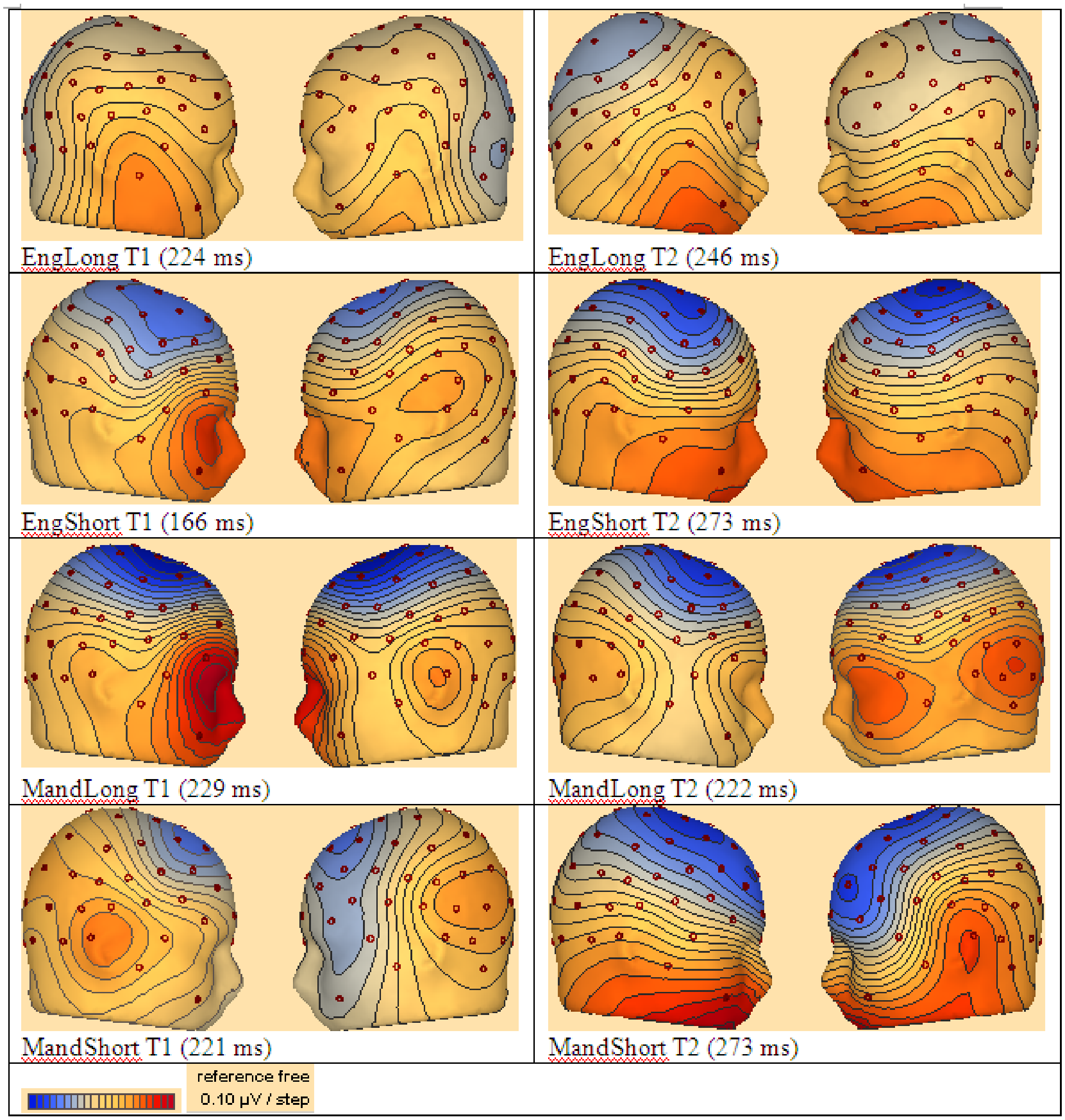
Appendix Eight Additional Figures

Figure A2. Subtraction waves at the Composite Fz for the MMN time region from 100 to 350 ms (highlighted). The stars indicate presence of MMN, and the crosses indicate absence of MMN.



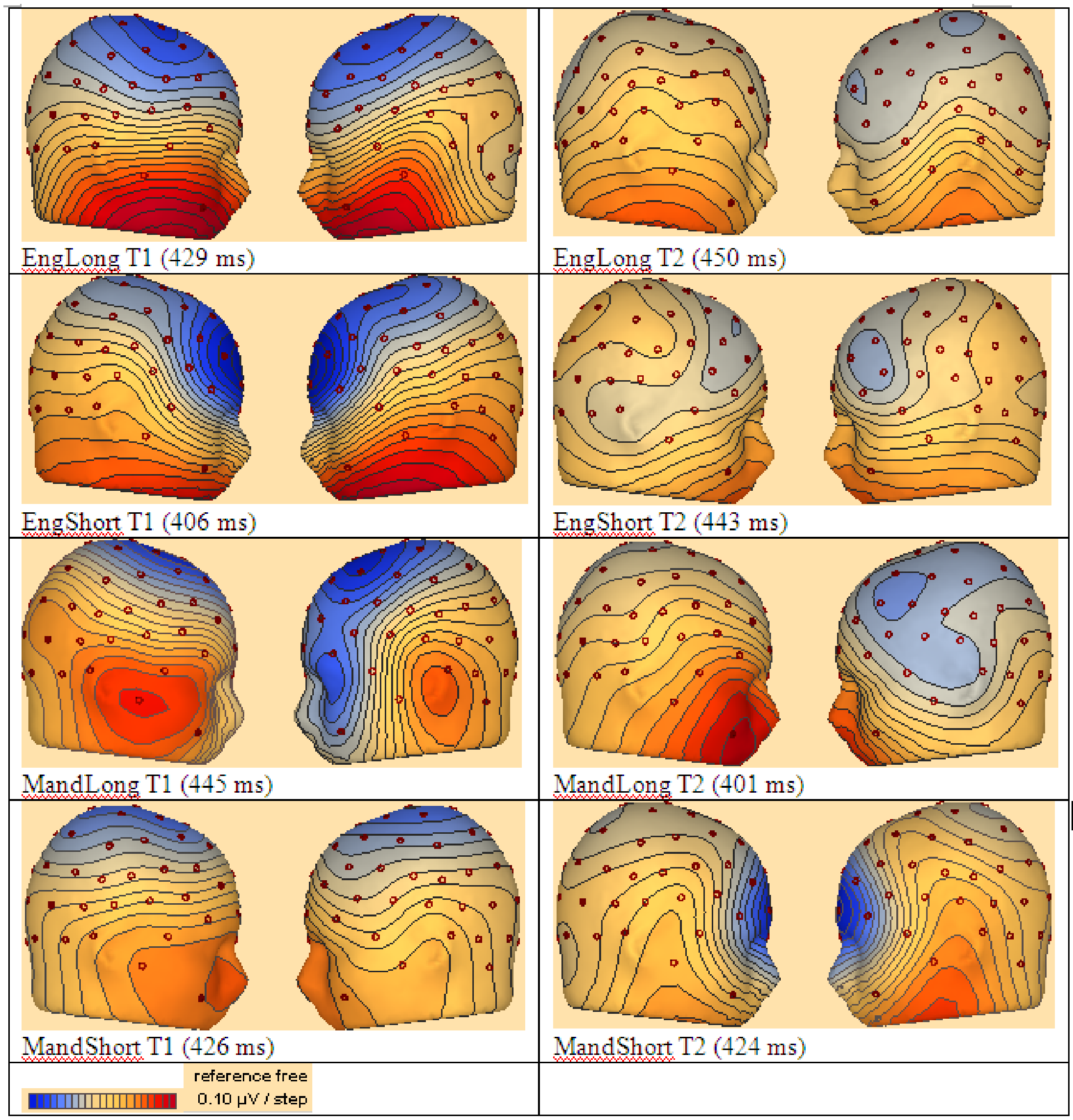
Appendix Eight Additional Figures

Figure A3. Subtraction waves at the Composite Fz for the LN time region from 350 to 550 ms (highlighted). The stars indicate presence of MMN, and the crosses indicate absence of MMN.


EngLong T1 (224 ms)
EngLong T2 (246 ms)
EngShort T1 (166 ms)
EngShort T2 (273 ms)
MandLong T1 (229 ms)
MandLong T2 (222 ms)
MandShort T1 (221 ms)
MandShort T2 (273 ms)

Appendix Eight Additional Figures

Figure A4. Topographic maps of MMN for the lexical tone conditions



EngLong T1 (429 ms)

EngLong T2 (450 ms)

EngShort T1 (406 ms)

EngShort T2 (443 ms)

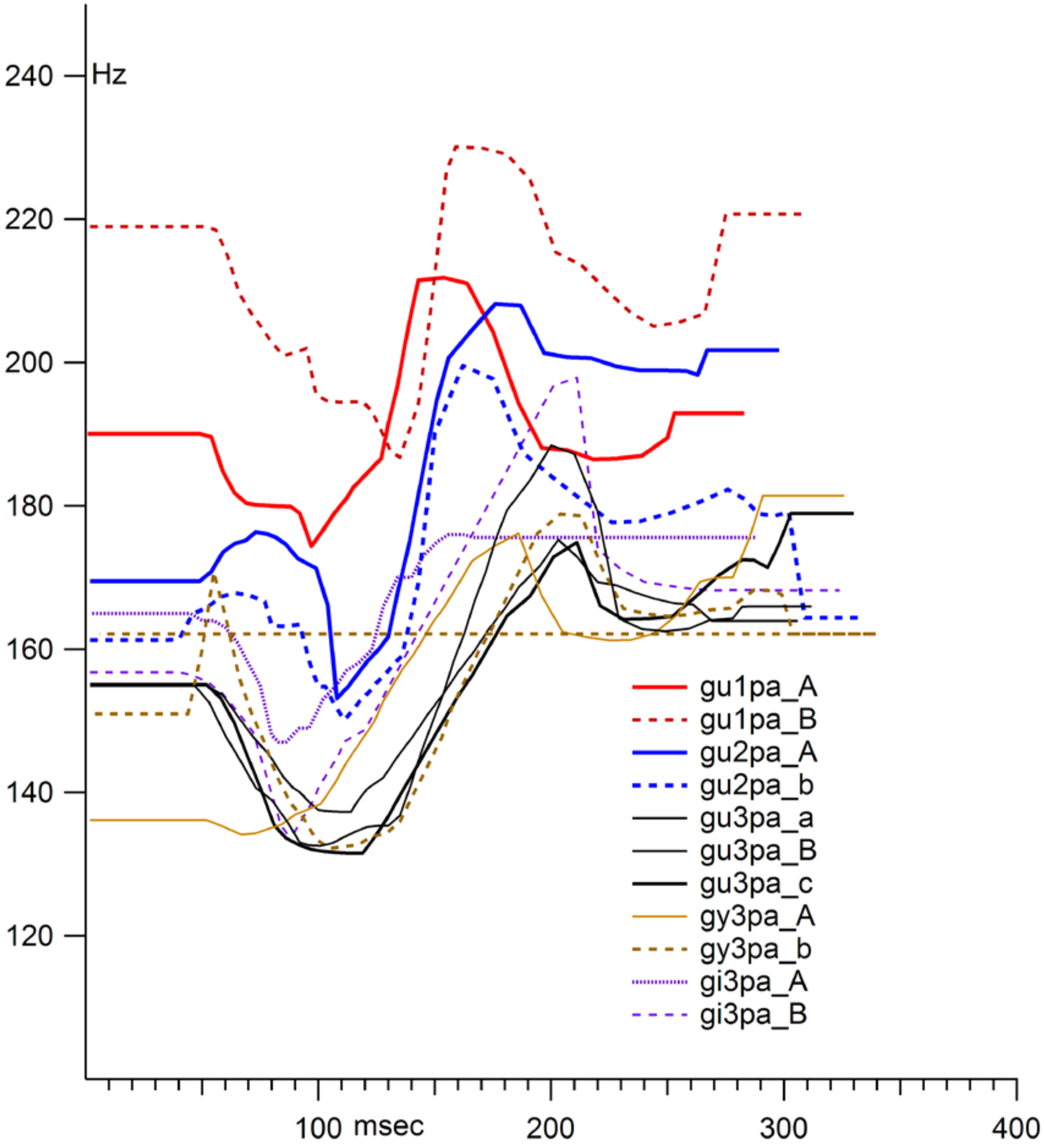
MandLong T1 (445 ms)

MandLong T2 (401 ms)

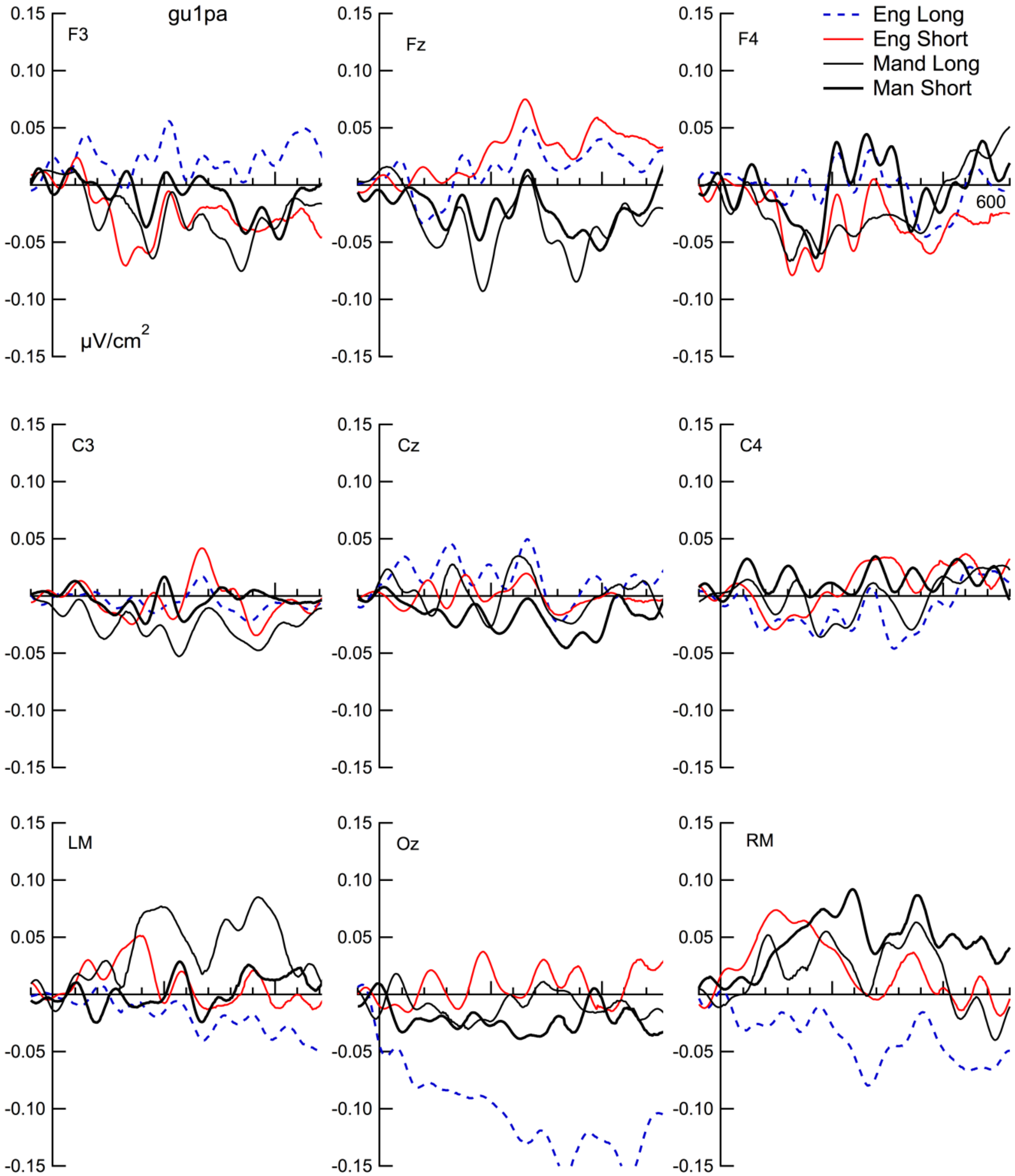
MandShort T1 (426 ms)

MandShort T2 (424 ms)

Appendix Eight Additional Figures
Figure A5. Topographic maps of LN for the lexical tone conditions

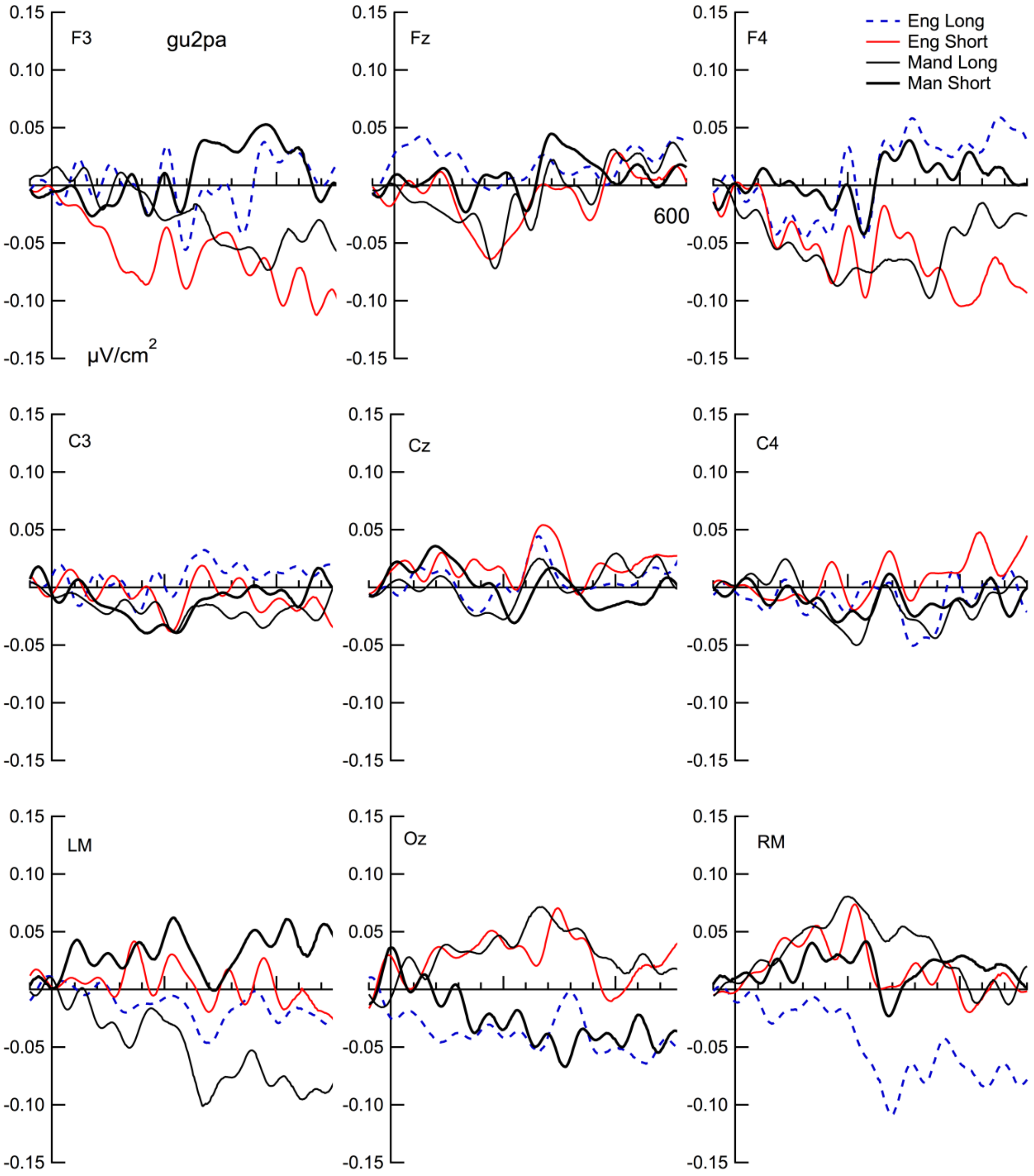


Appendix Eight Additional Figures
Figure A6. Fundamental frequency contours for all the stimuli



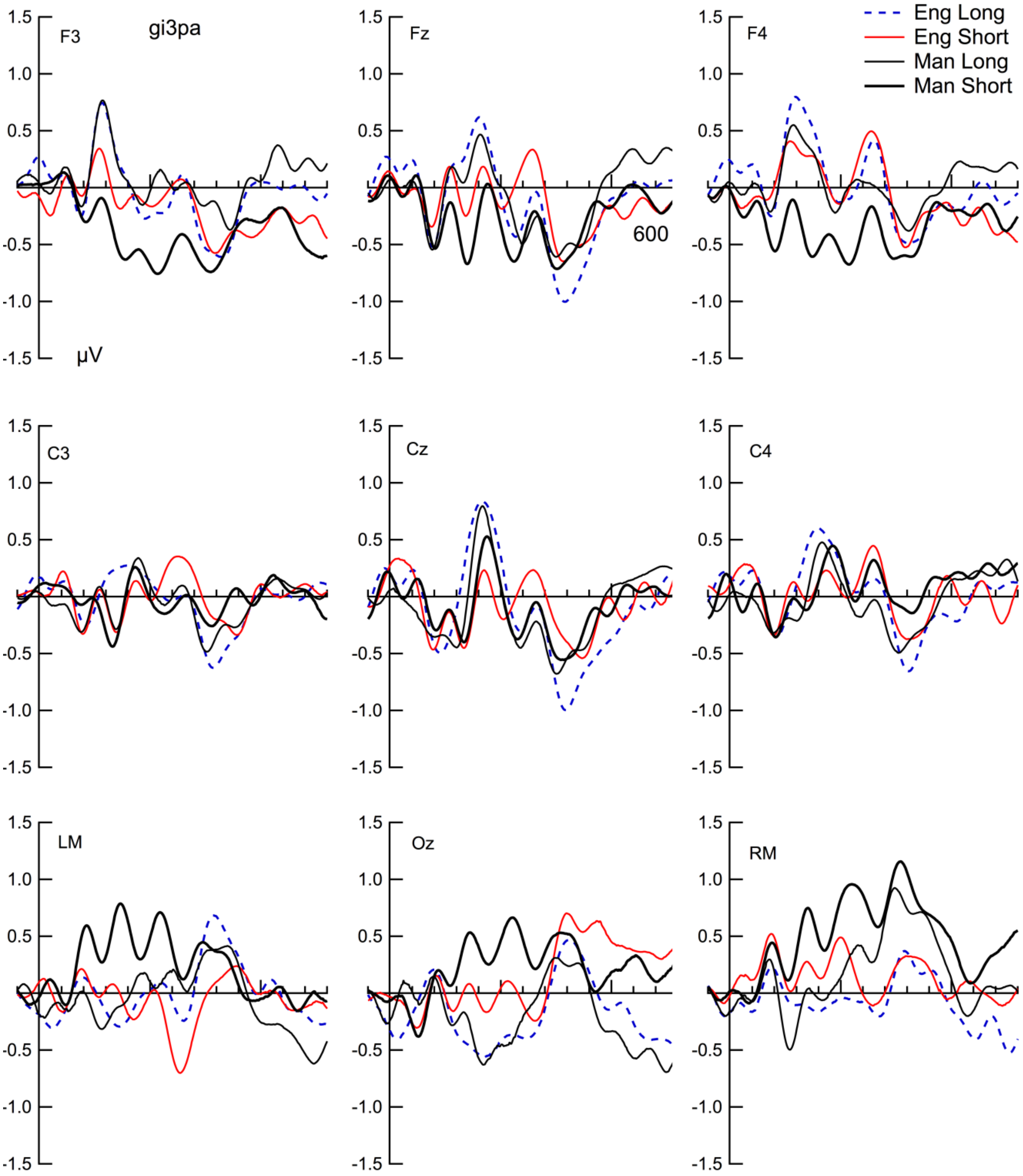
Appendix Eight Additional Figures

Figure A7. CSD subtractions waves for gu1pa deviant condition at nine sites for all groups



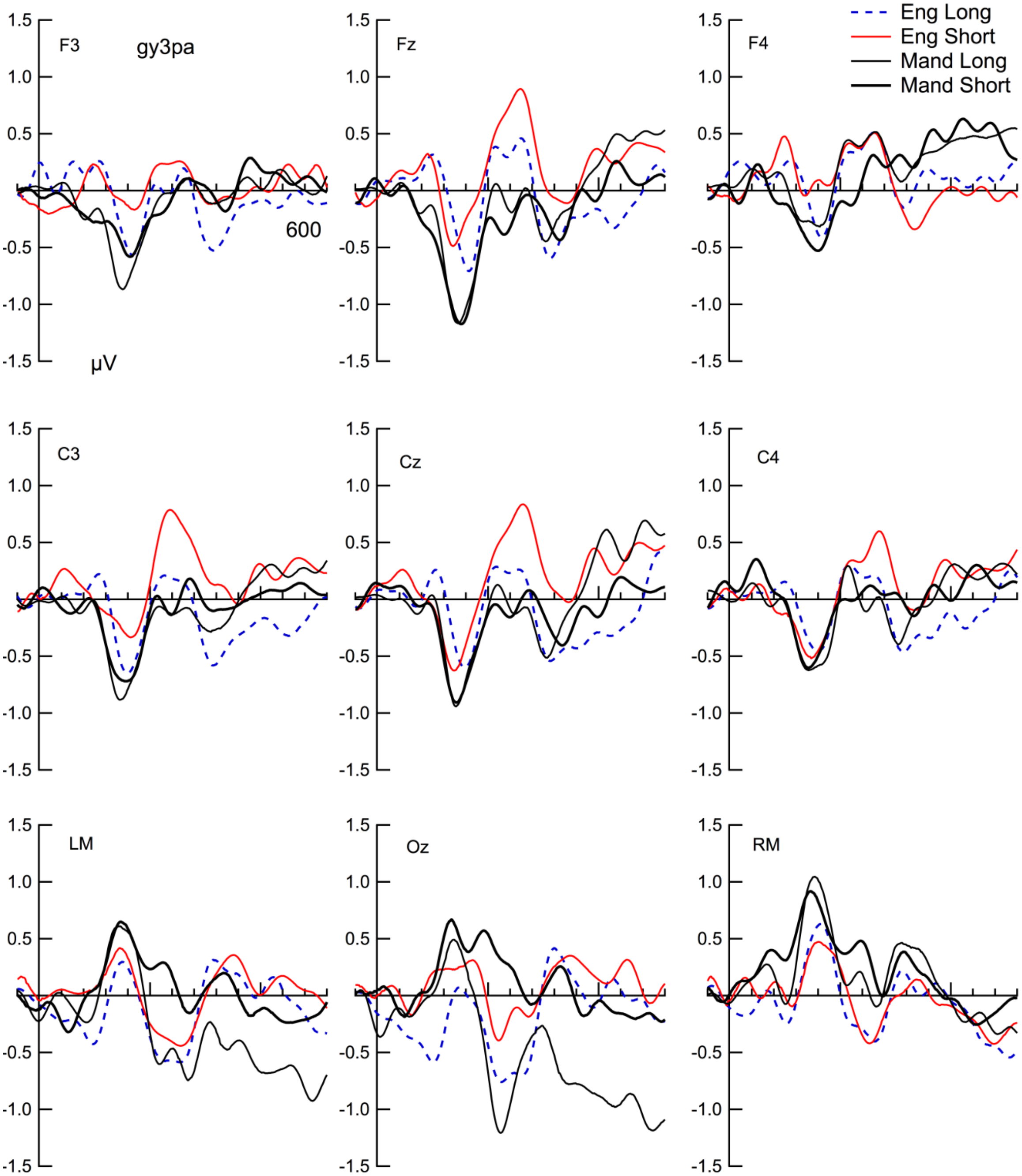
Appendix Eight Additional Figures

Figure A8. CSD subtractions waves for gu2pa deviant condition at nine sites for all groups



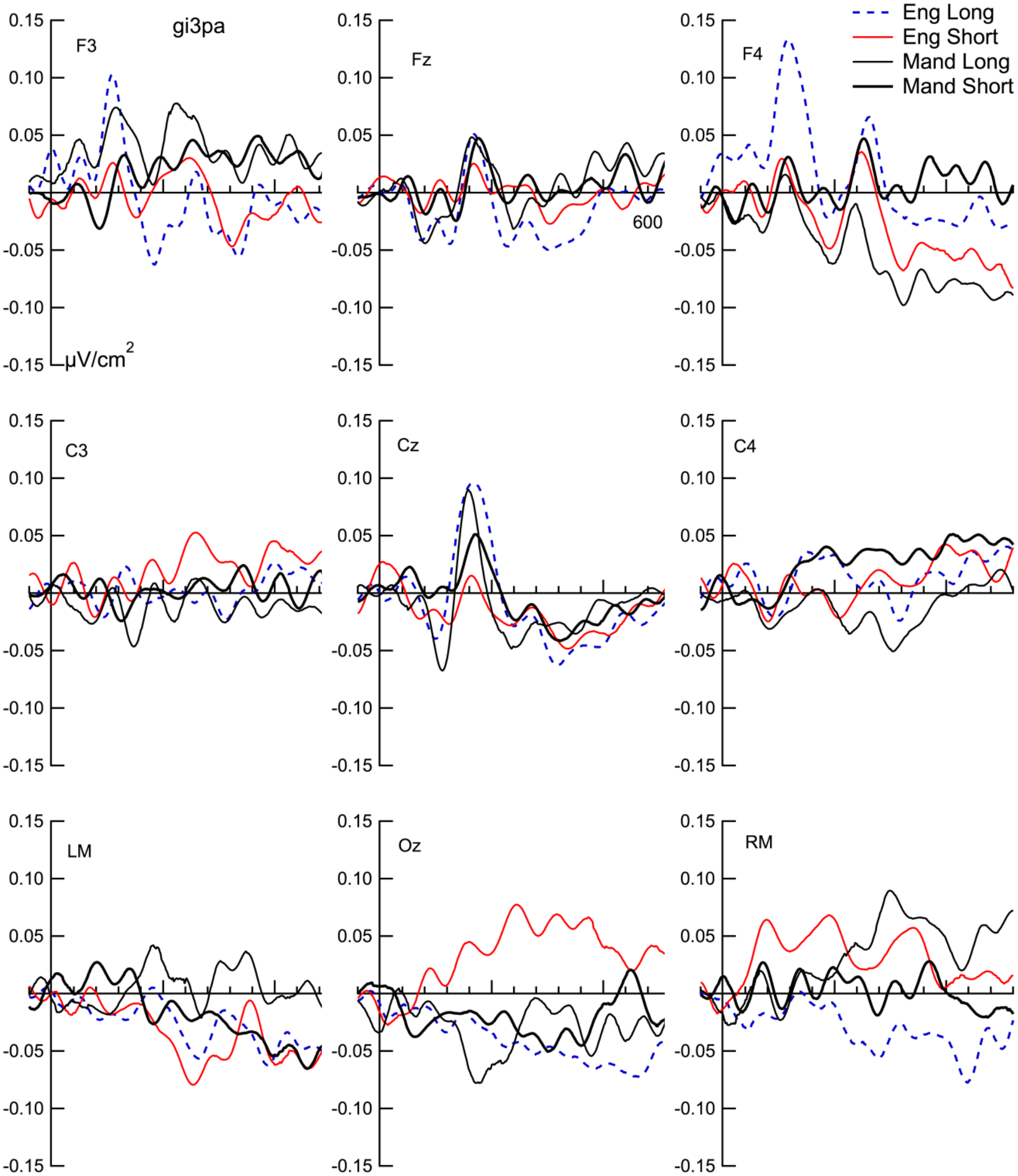
Appendix Eight Additional Figures

Figure A9. Voltage subtractions waves for gi3pa deviant condition at nine sites for all groups.



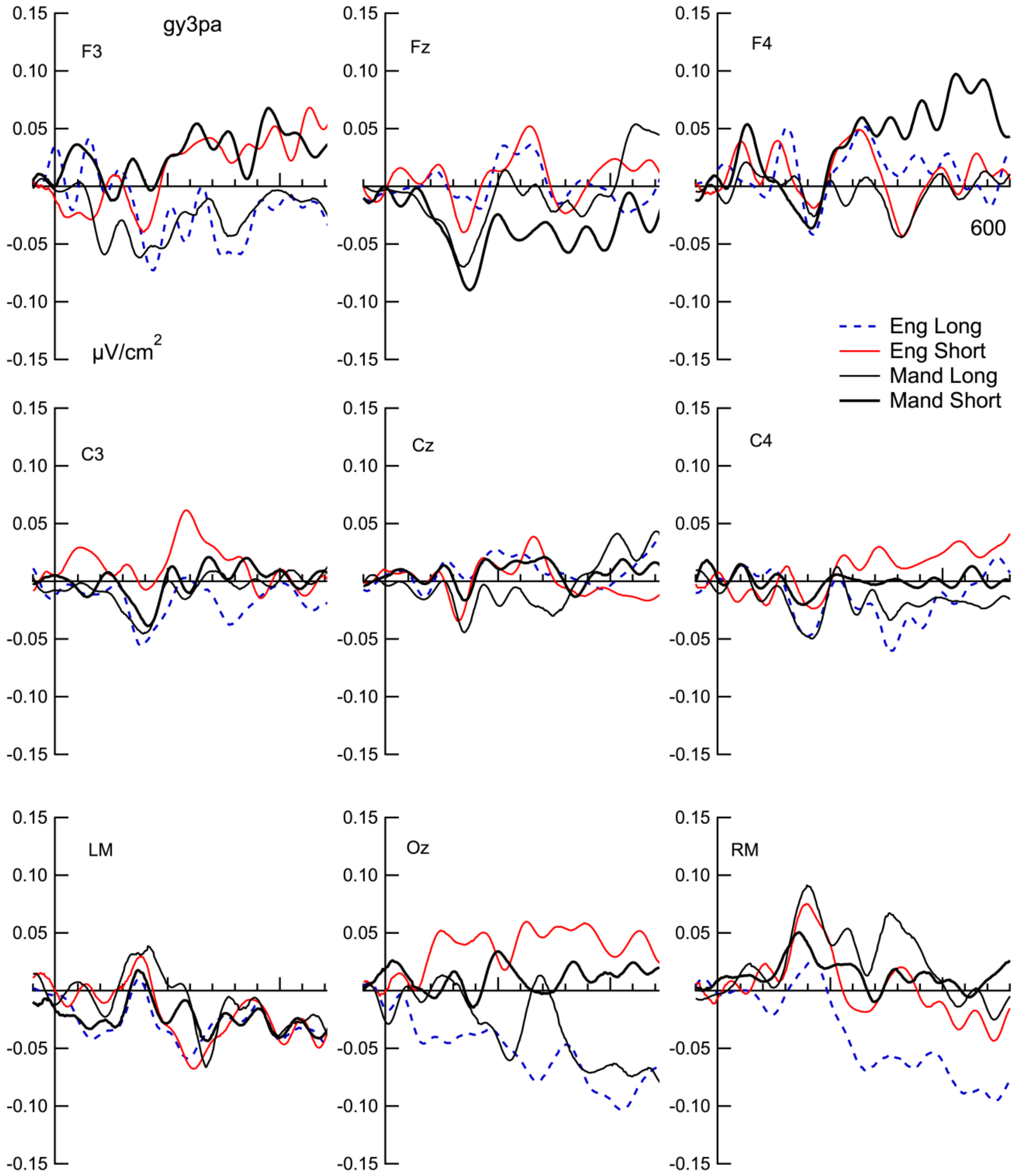
Appendix Eight Additional Figures

Figure A10. Voltage subtractions waves for gy3pa deviant condition at nine sites for all groups.



Appendix Eight Additional Figures

Figure A11. CSD subtractions waves for gi3pa deviant condition at nine sites for all groups



Appendix Eight Additional Figures

Figure A12. CSD subtractions waves for gy3pa deviant condition at nine sites for all groups

Bibliography

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *The Journal of the Acoustical Society of America, 101*(2), 1090-1105.

Addis, L., Friederici, A. D., Kotz, S. A., Sabisch, B., Barry, J., Richter, N., Ludwig AA, Rübsamen R, Albert FW, Pääbo S, Newbury DF, Monaco, A. P. (2010). A locus for an auditory processing deficit and language impairment in an extended pedigree maps to 12p13.31-q14.3. *Genes, Brain, and Behavior, 9*(6), 545-561.

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing, 16(1),* 38-51.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005, September). Lexical tone perception in musicians and non-musicians. In *INTERSPEECH* (pp. 397-400).

Barry, J. G., Hardiman, M. J., Line, E., White, K. B., Yasin, I., & Bishop, D. V. (2008). Duration of auditory sensory memory in parents of children with SLI: A mismatch negativity study. *Brain and Language, 104*(1), 75-88.

Barry, J.G., Hardiman, M.J., & Bishop, D.V. (2009). Mismatch response to polysyllabic nonwords: a neurophysiological signature of language learning capacity. PLoS One. 4(7):e6270. doi: 10.1371/journal.pone.0006270.

Bishop, D.V., Hardiman, M.J., & Barry, J.G. (2010). Lower-frequency event-related desynchronization: a signature of late mismatch responses to sounds, which is reduced or absent in children with specific language impairment. *Journal of Neuroscience, 17;30(46):*15578-84.

Böttcher-Gandor C, & Ullsperger P. (1992). Mismatch negativity in event-related potentials o auditory stimuli as a function of varying interstimulus interval. Psychophysiology 1992;29:546–50

Burnham, D. K. 1986 Developmental loss of speech perception: exposure to and experience with a first language. *Applied Psycholinguistics,7,* 207–240.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing.*ICSLP96 Proceedings4*, 2514-2517.

Čeponiene, R., Cheour, M., & Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology, 108*(4), 345-354.

Čeponiene, R., Service, E., Kurjenluoma, S., Cheour, M., & Näätänen, R. (1999). Children's performance on pseudoword repetition depends on auditory trace quality: Evidence from event-related potentials. *Developmental Psychology, 35*(3), 709-720.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007a). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25*(3-4), 195-210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148-156.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear and Hearing, 30*(5), 552-558.

Cohen, M. S. (2008). Handedness questionnaire. http://www.brainmapping.org/shared/Edinburgh.phpAdapted from Oldfield, R.C. *The assessment and analysis of handedness: the Edinburgh inventory*. Neuropsychologia. 9(1):97-113. 1971.

Cowan, N. (Ed.). (2005). *Working memory capacity*. Hove, East Sussex, UK: Psychology Press.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96*(2), 341-370.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163-191.

Cowan, N. (1995). Attention and memory: An integrated framework. Oxford, England: Clarendon Press.

Cowan, N. (2009). Sensory and Immediate Memory. In W.P. Banks (Ed.), *Encyclopedia of Consciousness*. (Vol.2, pp. 327-339). Oxford: Elsevier.

Crowder, R. G. (1982). Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 8*(2), 153-162.

Datta, H., Shafer, V. L., Morr, M. L., Kurtzberg, D., & Schwartz, R. G. (2010). Electrophysiological indices of discrimination of long-duration, phonetically similar vowels in children with typical and atypical language development. *Journal of Speech, Language, and Hearing Research, 53*(3), 757-777.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8*(4), 919-924.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*(4), 635-647.

Dien, J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: a model of converging pathways. *Biological Psychology, 80(1),*10-22. doi: 10.1016/j.biopsycho.2008.04.013.

Gandour, J., &Harshman, R. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *The Journal of the Acoustical Society of America, 62*, 693-707.

Gandour, J. (2006). Brain mapping of Chinese speech prosody. In: Handbook of East Asian psycholinguistics, Vol. 1 (Li P, Tan L, Bates E, Tzeng O, eds), pp 308 –319. Cambridge, UK: Cambridge University Press. [Chinese].

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage 23,* 344–357.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., Hutchins, G. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12,* 207–222.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., Lurito, J. (2002). Neural circuitry underlying perception of duration depends on language experience. *Brain & Language 83,* 268–290.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336-360.

Gomes, H., Sussman, E. S., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G., Jr. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Developmental Psychology, 35*(1), 294-302.

Grau, C., Polo, M.D., Yago, E., Gual, A., Escera, C. (2001). Auditory sensory memory as indicated by mismatch negativity in chronic alcoholism. *Clinical Neurophysiology, 112,* 728-731

Grossheinrich, N., Kademann, S., Bruder, J., Bartling, J., & Von Suchodoletz, W. (2010). Auditory sensory memory and language abilities in former late talkers: A mismatch negativity study. *Psychophysiology, 47*(5), 822-830.

Gumenyuk, V., Korzyukov, O., Escera, C., Hämäläinen, M., Huotilainen, M., Häyrinen, T., Oksanen, H., Näätänen, R., von Wendt, L, & Alho, K. (2005). Electrophysiological evidence of enhanced distractibility in ADHD children. *Neuroscience Letters, 374(3),* 212-7.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liegeois, C., Chauvel, P., & Musolino, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalography and Clinical Neurophysiology, 94*(3), 191-220.

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,* 291–453.

Hamburger, M., & Slowiaczek, L. (1996). Phonological priming reflects lexical competition. Psychonomic Bulletin & Review, 3, 520–525.

Hansen, J. C. &Hillyard, S. A. (1984). Effects of Stimulation Rate and Attribute Cuing on Event-Related Potentials During Selective Auditory Attention. *Psychophysiology, 21*,394–405.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Hedger, S.C., Heald, S.L., & Nusbaum, H.C. (2013, in press). Absolute Pitch May Not Be So Absolute. *Psychological Science.*

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews. *Neuroscience, 8(5),* 393-402.

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research, 11,* 1360:89-105.

Howie, J.M. (1976). Acoustical studies of Mandarin vowels and tones. New York, NY: Cambridge University Press.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685-701.

Hsieh, L., Gandour, J., Wong, D., Hutchins, G. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain & Language*, *76*, 227–252.

Johnson, P.R. (1977). Dichotically-stimulated ear differences in musicians and non-musicians. *Cortex, 13,* 385–389.

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,* 113-22.

Kaan, E., Barkley, C. M., 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.

Klouda, GV, Robin, DA, Graff-Radford, NR, &Cooper, WE (1988) The role of callosal connections in speech prosody. *Brain and Language, 35,*154–171.

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language, 76*(3), 332-339.

Korzyukov O, Winkler I, Gumenyuk V,& Alho K.(2003). Processing abstract auditory features in the human auditory cortex. *NeuroImage, 20*:2245–58.

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

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America, 125*(2), 1138-1152.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36,* 141-157.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word,* 384-422.

Mäntysalo , S., & Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology, 24*(3), 183-195.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

Mazziotta, J.C., Phelps, M.E., Carson, R.E., &Kuhl, D.E. (1982). Tomographic mapping of human cerebral metabolism, *Neurology, 32,* 921–937.

Miyawaki, K., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18,* 331-340.

Morais, J., Peretz, I. & Gudanski, M. (1982). Ear asymmetry for chord recognition in musicians and nonmusicians. *Neuropsychologia, 20,* 351–354.

Muller, D., Widmann, A. & Schröger, E. (2005) Auditory streaming affects the processing of successive deviant and standard sounds. *Psychophysiology,42*(6), 668-676.

Näätänen, R, Paavilainen P, Alho, K, Reinikainen, K, &Sams, M. (1987). Interstimulus interval and the mismatch negativity. In: Evoked Potentials III, The Third International Evoked Potentials Symposium (Eds. C. Barber, & T. Blum) 1987, pp. 392–397. Butter-worths, Boston.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 118*(12), 2544-2590.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: FMRI and ERP results. *NeuroImage, 15*(1), 167-174.

Ortiz-Mantilla, S., Choudhury, N., Alvarez, B., Benasich, A.A. (2010). Involuntary switching of attention mediates differences in event-related responses to complex tones between early and late Spanish-English bilinguals. *Brain Research, 1362*, 78-92.

Paavilainen, P., Jaramillo, M., & Näätänen, R. (1998). Binaural information can converge in abstract memory traces. *Psychophysiology, 35*:483–7.

Paavilainen, P., Saarinen, J., Tervaniemi, M., & Näätänen, R. (1995). Mismatch negativity to changes in abstract sound features during dichotic listening. *Journal of Psychophysiology, 9*, 243–9.

Paavilainen, P., Valppu, S., &Näätänen, R. (2001). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1 _ 1 _ 2 but 1 _ 1 _ 1 _ 3. *Neuroscience Letters,* ***301,***179–182.

Parra, J., Kalitzin, S.N., & da Silva, F.H. (2004). Magnetoencephalography: an investigational tool or a routine clinical technique? *Epilepsy & Behavior, 5*(3), 277-285.

Pekkonen, E., Jousmaki, V., Kononen, M., Reinikainen, K., Partanen, J. (1994). Auditory sensory memory impairment in Alzheimer's disease: an event-related potential study. *Neuroreport, 5*, 2537-2540

Pekkonen, E., Rinne, T., Reinikainen, K., Kujala, T., Alho, K., Näätänen, R. (1996). Aging effects on auditory processing: an event-related potential study. *Experimental Aging Research, 22,* 171-184.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253-260.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *The Journal of the Acoustical Society of America, 97*(2), 1286-1296.

Pulvermüller, F., & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimaging, 20,*159–72.

Ren, G. Q., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience, 162*(1), 87-95.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract stimulus attributes in human brain. *NeuroReport, 3,*1149–51.

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence. *Journal of Cognitive Neuroscience, 5*(3), 363-370.

Sams, M., Paavilainen, P., Alho, K., Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogram Clinical Neurophysiology, 62,* 437–48.

Sergent, J. (1982). The Cerebral Balance of Power: Confrontation or Cooperation? *Journal of Experimental Psychology: Human Perception and Performance, 8,* 253-272.

Schröger, E. & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport, 9(15),* 3355-8.

Shafer, V. L., Schwartz, R. G., &Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Brain Research. Cognitive Brain Research, 18*(3), 242-254.

Shafer, V.L , Morr, M.L., Datta H. Kurtzberg, D & Schwartz, R.G. (2005). Neurophysiological indices of speech processing deficits in children with specific language impairment. *Journal of Cognitive Neurosciences 17,* 1168-1180.

Shafer, V. L., Ponton, C., Datta, H., Morr, M. L., & Schwartz, R. G. (2007). Neurophysiological indices of attention to speech in children with specific language impairment. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 118*(6), 1230-1243.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106*(2), 1078-83.

Sharma, A., & Dorman, M. F. (2000). Neurophsiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107,* 2697-703.

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology, 88*(1), 64-71.

Shestakova, A., Huotilainen, M., Čeponiene, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 114*(8), 1507-1512.

Shtyrov, Y., Pihko, E., & Pulvermüller, F. (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *NeuroImage, 27,* 37–47.

Shtyrov, Y., Kujala, T. & Pulvermüller, F. (2010). Interactions between language and attention systems: early automatic lexical processing? *Journal of Cognitive Neuroscience, 22(7),* 1465-1478.

Shtyrov, Y., Kimppa, L., Pulvermüller, F., & Kujala, T. (2011). Event-related potentials reflecting the frequency of unattended spoken words: a neuronal index of connection strength in lexical memory circuits? *NeuroImage, 55*(2), 658-668.

Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topography, 3(1):*137-41.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience.* (pp. 3-45). Timonium, MD: York Press.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen-Edwards, & M. L. Zampini (Eds.), *Phonology and second* language acquisition. (pp. 153-192). Amsterdam and Philadelphia: Johns Benjamins.

Sussman, E., Ritter, W., & Vaughan Jr, H.G. (1998). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport, 9,* 4167–70.

Sussman, E., Ritter, W., & Vaughan Jr., H.G., (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology 36,* 22–34.

Sussman, E., Sheridan, K., Kreuzer, J., & Winkler, I. (2003). Representation of the standard: Stimulus context effects on the process generating the mismatch negativity component of event-related brain potentials.*Psychophysiology, 40,* 465-71.

Sussman, E., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *NeuroReport, 16*(13), 1519–1523.

Sussman, E. (2007). A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology, 21*(3-4), 164-175.

Sussman, E. S., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research, 236*(1-2), 61-79.

Szymanski ,M.D, Yund, E.W., & Woods, D.L.(1999). Phonemes, intensity and attention: Differential effects on the mismatch negativity (MMN). Journal of the Acoustical Society of America, 106, 3492–505.

Tervaniemi, M, Maury, S, & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport,5,*844–6.

The Electric Geodesics, I. *EEG system net station* (4.1.2 ed.)

Tiitinen, H., May, P., Reinikainen, K. &Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372,* 90-92.

Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *Neuroimage, 28 (2),* 417-428.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. Papers in linguistics. *International Journal of Human Communication, 13(2),* 201-277.

van Wijngaarden, S.J., Steeneken, H.J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native talkers. *Journal of the Acoustical Society of America, 112(6):*3004-13.

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

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.

Wayland, R., & Guion, S. (2004). Training native English and native Chinese speakers to perceive Thai tones. *Language Learning, 54(4),* 681–712.

Wayland, R. & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36,* 250-267.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37*(2), 278-286.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866-1878.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*(1), 25-47.

Winkler, I., & Cowan, N. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch. *Journal of Cognitive Neuroscience, 8*(5), 403.

Winkler, I., Reinikainen, K., & Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics, 53*(4), 443-449.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen O, Raimo I, Alho K, Lang H, Iivonen A, Näätänen R.(1999a). Pre- attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357–69.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski A, Czigler, V., Ilmoniemi, R.J., Näätänen, R. (1999b). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36,* 638–42.

Winkler I, Schröger E, & Cowan N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience,13(1):*59-71.

Winkler, I., & Cowan, N. (2005). From sensory to long-term memory: Evidence from auditory memory reactivation studies. *Experimental Psychology, 52*(1), 3-20.

Wong, P.C.M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletins, 59(2),* 83-95.

Wong, P.C.M., Parsons L.M., Martinez, M., & Diehl, R.L. (2004). The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *Journal of Neuroscience, 24(41):*9153–9160.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience, 10(4),* 420-422.

Xu, Y. (1997). Contextual tonal variations in Mandarin*. Journal of Phonetics, 25*, 61-83.

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

Zatorre, R.J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex 11*:946–953.

Zatorre, R.J., Belin, P., & Penhune V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences,6(1):*37-46.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. Philosophical Transactions of the Royal Society of London. *Series B, Biological Sciences, 363*(1493), 1087-1104.