10-10-2013 12:00 AM

Taking Tone into Account: Cognitive Neuroscientific Investigations of Mandarin Chinese Spoken Word Processing

Jeffrey G. Malins
The University of Western Ontario

Supervisor
Dr. Marc Joanisse
The University of Western Ontario

Graduate Program in Neuroscience
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Recommended Citation

Malins, Jeffrey G., "Taking Tone into Account: Cognitive Neuroscientific Investigations of Mandarin Chinese Spoken Word Processing" (2013). *Electronic Thesis and Dissertation Repository*. 1723.
https://ir.lib.uwo.ca/etd/1723

TAKING TONE INTO ACCOUNT: COGNITIVE NEUROSCIENTIFIC INVESTIGATIONS OF MANDARIN CHINESE SPOKEN WORD PROCESSING

(Thesis format: Integrated Article)

by

Jeffrey G. Malins

Graduate Program in Neuroscience

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

To date, theories of how humans recognize spoken words have yet to account for tonal languages such as Mandarin Chinese. One reason for this is that we know relatively little about how native speakers of tonal languages process spoken words in the brain. This dissertation addresses this problem by examining Mandarin spoken word processing in both adult native speakers and typically developing children. In adults, functional magnetic resonance imaging (fMRI) was used to assess the extent to which the brain regions involved in processing tonal information are distinct from those involved in vowel processing (Chapter 2), while event related potentials (ERPs) were used to investigate responses to different types of phonological competition in Mandarin (Chapter 3). In the fMRI experiment, subjects performed a passive listening task, in which they heard trains of Mandarin syllables consisting of a repeated standard followed by a deviant stimulus differing from the standard in either tone or vowels. Analyses revealed that the regions involved in processing tonal versus vowel deviants were not entirely overlapping. In the ERP experiment, subjects were presented with pictures of items and subsequently heard words that either matched or mismatched the pictures. Mismatches differed from expectations in different components of the Mandarin syllable, and analyses focused on ERP components associated with various stages of spoken word processing. A key finding from this study was that different cognitive processes underlie tonal versus phonemic (vowel) access. Chapter 4 extended these findings by studying the development of Mandarin spoken word processing. In this experiment, a group of typically developing children completed the same ERP picture-word matching task as a group of adults. It was observed that the children differed from the adults primarily in their responses to rhyme mismatches. On the basis of relevant findings, Chapter 5 puts forward several recommendations as to how current theories of spoken word recognition could be modified to account for tonal languages. This proposed theory is then used to explain the observed data from the adults and typically developing children. The dissertation then closes with a brief consideration of future research directions.

Keywords

Spoken Word Recognition, Lexical Tone, Mandarin Chinese, Phonological Processing, Tonemes, Event Related Potentials, Functional Magnetic Resonance Imaging, Development.

Co-Authorship Statement

The chapters of this dissertation are manuscripts that have been prepared for submission to scientific journals. The presented data are based on a series of collaborative research projects; however all manuscripts have been primarily written by Jeffrey G. Malins. Chapter 2 has been written in preparation for publication, authored by Jeffrey G. Malins and Marc F. Joanisse. Chapter 3 has been published in *Neuropsychologia* (volume 50, pages 2032-2043), authored by Jeffrey G. Malins and Marc F. Joanisse. Chapter 4 has also been written in preparation for publication, authored by Jeffrey G. Malins, Danqi Gao, Ran Tao, James R. Booth, Hua Shu, Marc F. Joanisse, Li Liu, and Amy S. Desroches. Finally, certain passages within Chapters 1 and 5 have appeared in conference proceedings for the Third International Symposium on Tonal Aspects of Languages, held in Nanjing, China in June 2012. These proceedings were authored by Jeffrey G. Malins and Marc F. Joanisse.

Acknowledgments

First and foremost, to my supervisor Marc Joanisse, thank you for taking a chance on me six years ago. I came into the lab without even a first-year psychology course, really confused about what I wanted to study, and I had a lot of catching up to do along the way. I really appreciate your patience, and your trust that I would figure things out as I went. I’m so grateful that you let me chart my own course and develop a research plan that suited my interests, and that I could take ownership of. I’m pleased that it seems to have taken shape, and is something that will carry on in the lab after I leave. I am also very thankful for your guidance and advice throughout the Master’s and PhD programs. I feel that I am emerging a much more confident and competent scientist as a result of my time in the lab. I look forward to continuing to collaborate with you in the future.

To my advisory committee: Ken McRae, Bruce Morton, and Stefan Köhler, I very much appreciate your feedback on all aspects of this body of work. This dissertation is much stronger as a result of your contributions.

To my fellow and former labmates: Amy Desroches, Kathy Maxwell, Janet Aucoin, Lisa Archibald, Suzanne Welcome, Laura Westmaas-Sweet, Emily Nichols, and Diedre Desouza, I consider all of you not only as colleagues but as close friends. You helped me grow both academically and personally, and I am so thankful for your guidance and support. I have many wonderful memories of our time together. And to Erin Robertson and Randy Newman: even though I didn’t get a chance to interact with you very much, I have still benefitted very much from your feedback on my work and from the protocols you put in place before I arrived.

To my collaborators at Beijing Normal University: Lily Liu, Danqi Gao, and Ran Tao, thank you so much for giving me an opportunity to work with you and for showing me such hospitality when I visited Beijing. I have already gained tremendously as a result of our collaboration, and I know that it will continue to be very prosperous in the coming years.

I would like to give a most special thanks to Yijun Gao. You gave invaluable assistance with stimulus preparation, subject recruitment, and data collection. I really enjoyed working with you, and found your enthusiasm for science infectious. Thank you also for sharing your love of Chinese culture with me.

I am also indebted to Li Xin, whom I am honoured to call a close friend. I know how busy you were during your last year at Western, and marvel at how you found time each week to help me practice speaking Mandarin. You are a most patient teacher!

Finally, to my close friends Heather, Meg, Melanie, Renée, and Diedre (you deserve a second mention here!), thanks so much for your encouragement along the way, and for helping me live a balanced life. To my parents, Robert and Linda, I love sharing my accomplishments with you, and am so thankful for all of your support. Thank you for giving me the freedom to make my own choices. To the rest of my extended family, and most especially to my late grandmother Isabel, thank you for reminding me of what is most important in life, and for inspiring me to reach my full potential. To my sister Pam, words cannot express my gratitude for all that you've done these past six years. I feel like we've both grown immeasurably during our time here, and I'm so excited to see where life will take you. And to my partner Charles, this dissertation is just as much your accomplishment as it is mine. I couldn't have done this without you by my side.

Table of Contents

List of Tables

List of Figures

List of Appendices

1 Introduction

We currently lack a rich understanding of how spoken words are comprehended in many of the world's languages. The theories and models we currently possess for how humans recognize spoken words are based on Indo-European languages such as English and Dutch (McClelland & Elman, 1986; Marslen-Wilson, 1987; Luce & Pisoni, 1998; Norris & McQueen, 2008). Consequently, these theories fail to capture spoken word processing in Mandarin Chinese – one of the most widely spoken languages in the world – for two key reasons. The first is that most Indo-European languages lack the feature of lexical tone, which is a central component of tonal languages such as Mandarin. The second is that Mandarin is very different from languages such as English and Dutch in segmental structure. In order for these theories to account for Mandarin, it is imperative they address these two aspects of Mandarin spoken word processing. My research seeks to contribute to these theories by using cognitive neuroscientific techniques to investigate how healthy native Mandarin speakers represent and process spoken words in the brain.

By conducting this type of work, I aim to offer recommendations on how to improve current theories and models so they can account for Mandarin spoken word processing. This is important not only from a theoretical standpoint, but from an applied one as well. For example, as shown in this dissertation, an adequate characterization of how Mandarin spoken words are recognized in healthy adults can be used as a platform for studying the development of spoken word processing in Mandarin-speaking children. This approach in turn has the potential to address questions about phonological representations in tonal language speakers with reading impairment (Ho, Law, & Ng, 2000; Ho, Chan, Tsang, & Lee, 2002; Ho, Chan, Lee, Tsang, & Luan, 2004; Siok, Perfetti, Jin, & Tan, 2004; Lee, Hung, & Tzeng, 2006; Cheung et al., 2009; Liu, Shu, & Yang, 2009; Newman, Tardif, Huang, & Shu, 2011).

1.1 The Structure of Mandarin Syllables

1.1.1 Lexical Tone

Tone refers to modulation in the pitch of a speaker's voice during articulation of a spoken syllable. In Mandarin, there are four tones, each with distinct pitch profiles. Tone 1 is characterized by a high and level pitch; tone 2 begins in the mid-range of a speaker's modal register, and rises in pitch during articulation of a syllable; tone 3 starts in the mid-range, dips to a low level and rises again; tone 4 begins at a high register and sharply falls. Mandarin tones are denoted as contour tones, as pitch dynamically changes over the course of articulation. It should be noted that these contour tones are unlike the register tones in languages such as Cantonese, which sometimes only differ in the pitch height of a speaker's voice, and not in the shape of pitch profiles. For example, Cantonese speakers make a distinction between low-level versus mid-level versus high-level tones.

Tonal contrasts in Mandarin are lexical in nature. This means that words sharing the same segmental content, or phonemes, carry different meanings depending on the tone in which they are articulated. For example, in Mandarin the syllable *tang* means 'soup' when pronounced in a high and level tone (tone 1), whereas it means 'candy' when pronounced in a rising tone (tone 2). As a result, listeners must track how pitch contour changes over time as speech unfolds in order to understand spoken words.

1.1.2 Segmental Structure

In addition to lexical tone, there is another important difference between Mandarin and languages such as English: segmental structure. In English, clusters of consonants are permitted at the beginning of syllables; for example, the word ‘string’ begins with the cluster /str/. In Mandarin, only a single onset consonant is permitted. Furthermore, rime units (i.e., the final vowel and consonant in a syllable) are more restrictive: English again permits consonant clusters at the end of syllables, while Mandarin only permits two nasal consonants, /n/ and /ŋ/. Otherwise, Mandarin syllables do not terminate in any consonant; rather, they are open syllables that end in vowel clusters, comprised of one, two, or three vowels. These vowels are thought to be tone-bearing units because tonal information is carried over vowels in syllables (Howie, 1974). This results from differences in how consonants versus vowels are articulated: consonants are brief, transient segments that involve obstruction of the vocal tract in some fashion, and it is difficult to fluctuate the pitch of consonants as a result. This has implications for the studies in this dissertation that compare tonal versus segmental processing; often, segmental processing is taken as synonymous with vowel processing. The reason for this is that because tones are carried over vowels, a comparison of the relative processing of tonal versus segmental information is thought to be most reasonable if tones are compared to vowels (e.g., Cutler & Chen, 1997).

1.2 Current Theories of Spoken Word Recognition

Spoken word recognition can be considered as taking acoustic information and mapping it onto word knowledge in the brain in order to uncover the meaning of a spoken form (Samuel & Sumner, 2012). This activation of word knowledge, or lexical access (Klatt, 1979), is considered complete once activation of a word exceeds a certain threshold for recognition; how quickly it does this depends on a few factors. First, the amount of activation an item requires to cross the threshold is determined by its resting state, which is affected by word frequency (Dahan, Magnuson, & Tanenhaus, 2001). Second, a word form might receive partial activation based on contextual cues. This partial activation means that subsequent bottom-up processing of the word form will be facilitated, as it will take less additional activation to cross the threshold for recognition (McClelland & Elman, 1986). Third, activation of lexical items is greatly influenced by competition with other lexical items that share phonological overlap with a word form. For example, if we hear the phoneme sequence /kæn/ at the beginning of a word, this sequence could spread activation to a number of words, such as *candy*, *candle*, *canister*, *canopy* and so on (Marslen-Wilson, 1987). To disambiguate among these competitors, a listener must use incoming auditory information to narrow down the competitor set to a single word.

The dominant theories of how humans recognize spoken words include Cohort (Marslen-Wilson, 1987), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994; Norris & McQueen, 2008), and the neigborhood activation model (NAM; Luce & Pisoni, 1998). These models differ with respect to the types of competitors they predict will be active for any given spoken word, as well as whether they allow for top-down flow of information to influence word recognition. In both Cohort and Shortlist, word recognition proceeds in an entirely bottom-up feedforward manner. That is, information flows solely from acoustic-phonetic information towards meaning – contextual cues do not influence lexical access. In contrast, TRACE allows for both bottom-up and top-down processing. In the TRACE architecture, acoustic information first activates phonemic feature detectors, which then activate the phonemes that contain these features. These phonemes then activate the words that contain them. Connections between the feature level and the phoneme level are excitatory, as are connections between the phoneme level and the word level. In addition, within levels, connections between units are mutually inhibitory. Importantly, this architecture also incorporates feedback connections from words to phonemes, and from phonemes to features. These feedback connections allow for top-down flow of information.

An important consequence of this architecture is that Cohort and TRACE differ in the competitor set that they predict is active for any given spoken word. Because TRACE allows for feedback from the word to phoneme level, expecting to hear a certain word form can result in partial activation of the phonemes that comprise the word form, and these can in turn spread activation to other words that also contain these phonemes. For example, expecting to hear the word *bat* can result in activation of the phoneme units /b/ /æ/ and /t/, which can then spread activation to words that contain some of these phonemes, such as *bag* and *cat*. This model therefore predicts that not only words that share onset will be partially activated when expecting to hear a given word form, but also words that share word-final phonemes, or rhymes. A strictly bottom-up model like Cohort does not predict activation of rhyme competitors.

The neighborhood activation model (Luce & Pisoni, 1998), like TRACE, does predict activation of rhyme competitors, but it does so for a different reason. In TRACE – and in Cohort as well – the temporal ordering of words is considered when calculating phonological similarity among words. Cohort implements the most extreme version of this principle because word-initial overlap is considered to be the only type of phonological similarity that affects lexical competition (Marlsen-Wilson & Zwitserlood, 1989). In TRACE, even though rhyme competition is predicted, it is thought to be weaker than competition based on shared onsets, because by the time overlapping rhyme information arrives in the acoustic signal, activation of rhyming word forms has been greatly attenuated as a result of mutual inhibition amongst lexical competitors (Allopenna, Magnuson, & Tanenhaus, 1998). Theories such as Cohort and TRACE therefore posit that spoken words are processed incrementally, in that phonological similarity is tracked over time (Magnuson, Tanenhaus, Aslin, & Dahan, 2003). In contrast, NAM defines competitors, or neighbors, as differing from a word in one phoneme in any position. Consequently, NAM predicts that the words *bag*, *hat*, and *bit* are all competitors for the word *bat*, and importantly, should not differ from one another in their competitive influence on recognition of the word *bat*. This type of processing, in which a type of global similarity between spoken words is computed without taking temporal information into account, is considered holistic as opposed to incremental.

In studies of English spoken word recognition, a number of investigations using time-sensitive measures have supported theories of incremental processing (e.g., Allopenna et al., 1998; McMurray, Tanenhaus, & Aslin, 2002; Desroches, Robertson, & Joanisse, 2006; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Desroches, Newman, & Joanisse, 2009). In addition, several studies have reported evidence for lexical competition among rhyming words that is weaker in nature than onset-based competition (Allopenna et al., 1998; Desroches et al., 2006; Desroches et al., 2009). These observations are most consistent with the TRACE model of speech perception. However, studies of spoken word processing in Mandarin have been less conclusive.

1.3 Spoken Word Recognition in Mandarin

Recently, I conducted a study using eyetracking to investigate spoken word processing in Mandarin (Malins & Joanisse, 2010). The advantage of eyetracking is that it is an online language measure that offers continuous insight into the processes in operation during the unfolding of a spoken word. In this study, adult native Mandarin speakers were presented with arrays of pictures of items, and heard a spoken word that matched one of the pictures. They were asked to indicate via button press the position of the picture that matched the spoken word. Within the array, in addition to the target word, there was a picture of a phonological competitor that shared some sort of relationship with the target. For example, in some trials in which the target was *hua1*, I also included a picture of a cohort competitor differing in word-final vowels (*hui1*), while in other trials I included a rhyme competitor differing in onset consonant (*gua1*). During the unfolding of spoken words, I monitored subjects' looks to both targets and competitors. The extent to which subjects looked at the pictures was taken as an index of the level of activation for different words (Tanenhaus, Magnuson, Dahan, & Chambers, 2000). I found that the time course over which subjects looked to targets was affected by the presence of cohort competitors, but not by the presence of rhyme competitors. This suggests spoken words in Mandarin are processed incrementally, as in TRACE, but the lack of observed rhyme effects is more consistent with Cohort. However, even though this particular study failed to observe rhyme effects in Mandarin, they have been documented in an experiment using disyllabic words (Liu, Shu, & Wei, 2006). As a result, it is currently unclear whether or not rhyme competitors participate in lexical competition in Mandarin, and therefore whether or not TRACE is the most adequate starting point for a theory of spoken word recognition in Mandarin.

Furthermore, this uncertainty is exacerbated by recent evidence suggesting that spoken words are processed holistically in Mandarin. Zhao, Guo, Zhou, and Shu (2011) reported evidence that onset and rime-based similarity among words are treated similarly in Mandarin, offering support for theories such as NAM. They argued that the differences in segmental structure between Mandarin and English syllables result in Mandarin speakers processing onsets versus rimes in a different fashion than English speakers. This discrepancy with our earlier eyetracking results implies that further investigation is warranted before any strong claims can be made regarding whether Mandarin spoken word processing is incremental or holistic.

In addition, it is unclear how best to incorporate tone into theories of spoken word recognition. To address this, researchers initially investigated whether tonal information is used to constrain lexical access at all, or whether it is used only after a set of lexical candidates have been activated (Cutler & Chen, 1997). Early studies using behavioral tasks such as same-different judgments suggested that tonal information is accessed later than segmental information and therefore plays a weaker (or nonexistent) role in lexical access compared to segmental information (Repp & Lin, 1990; Taft & Chen, 1992; Cutler & Chen, 1997; Ye & Connine, 1999 – Experiment 1). However, recent research using more time-sensitive measures, including my own work, has suggested that tonal information plays a role equivalent to that of segmental information in constraining word recognition (Schirmer et al., 2005; Malins & Joanisse, 2010; Zhao et al., 2011). Therefore, I argue that because tone has been shown to participate in lexical access, it must be incorporated into theories of spoken word recognition.

1.4 Research Questions

To develop a theory of Mandarin spoken word processing, we need to answer several questions. First, how are tones represented in the brain? Given that tones are carried on vowels in Mandarin, this can be asked another way: to what extent are the cortical representations for tonal information distinct from the representations for the vowels on which they are carried? If tones are represented separately from vowels at some level, then an adequate theory of Mandarin spoken word processing would need to incorporate units that are solely responsible for handling tonal information. Second, how do tones participate in lexical access? Are the cognitive processes underlying tonal access the same as those underlying access to phonemes such as vowels, or different? Finally, it is important to consider a few other theoretical issues because these can help to adjudicate among current theories of word recognition to determine which is most appropriate for Mandarin. Namely, is Mandarin spoken word processing incremental or holistic? Does top-down processing influence word recognition in Mandarin? And furthermore, does this change over the course of development?

1.4.1 To What Extent are Tones Represented Separately from the Vowels on Which They are Carried?

An informed theory of Mandarin spoken word recognition foremost needs to address the issue of representation of tonal information. It is my view that this is best guided by an examination of how tones are represented in the brains of native Mandarin speakers. Given that tones are carried on vowels, a logical starting place for investigations of tonal representation is a consideration of how vowels and other phonemic units are represented in the brain.

One type of methodology that is particularly appropriate for localizing phonemic processing in the brain is functional magnetic resonance imaging (fMRI). fMRI measures differences in the homogeneity of local magnetic fields, which are thought to arise due to fluctuations in the level of blood oxygenation in the brain (Goebel, 2007). Using fMRI, one can monitor changes in this blood-oxygenation level dependent (BOLD) signal while subjects perform cognitive tasks. An increase in BOLD signal can be interpreted as an increased amount of neural activity in a particular brain region, resulting in a greater amount of blood flow, and therefore a greater amount of oxygenation in the blood vessels supplying the tissue. One can then correlate changes in the BOLD signal with the time points during which a subject is engaged in a cognitive task to test if there is a relationship between activity in a particular brain region and the task of interest.

Studies using fMRI to investigate the perception of speech stimuli have suggested that areas in the left temporal lobe are differentially sensitive to variation in the acoustic signal (Peelle, Johnsrude, & Davis, 2010). Specifically, it has been proposed that while primary auditory cortex is very sensitive to acoustic variation, as one moves downstream from this region, cortical areas become more tolerant of variation and respond in a more categorical fashion to speech stimuli (Liebenthal et al., 2010). For example, Joanisse, Zevin, & McCandliss (2007) conducted an fMRI experiment in which subjects were presented with trains of spoken syllables, with the first three syllables being identical. After this, a deviant stimulus was presented that differed from the repeated standard in its initial phoneme. Importantly, this initial phoneme either fell within the bounds of a phonemic category, or crossed it. The authors found that a region in left posterior superior temporal sulcus (STS) showed differential activation for between versus within-category deviants, suggesting that this region is involved in representing phonemic categories. While there is disagreement about the underlying computations that contribute to this category invariance and the brain regions associated with them (Myers, Blumstein, Walsh, & Eliassen, 2009), there has been a considerable amount of evidence in the field pointing to abstract representations of phonemes that allow listeners to successfully recognize consonants and vowels in spoken words amidst multiple sources of variability (DeWitt & Rauschecker, 2012).

These prior investigations of phonemic processing offered a foundation for studying tone processing in native Mandarin speakers. In the fMRI study reported in Chapter 2, we were interested in the extent to which tones are processed separately from the vowel phonemes on which they are carried. In other words, we wished to assess whether tone processing is entirely accomplished using pre-existing structures for phonemic processing, or whether additional structures are also involved in the representation of tonal categories. This research question is of importance from a theoretical standpoint, as it addresses the issue of whether or not a theory of Mandarin spoken word recognition should include phonological processing units devoted solely to tone.

To date, a number of neuroimaging studies have focused on tone processing in the brain (Gandour et al., 2000; Klein et al., 2001; Hsieh et al., 2001; Gandour et al., 2003; Li et al., 2003; Liang and van Heuven, 2004; Xu et al., 2006; Li et al., 2010). Importantly, most of these studies have used selective attention tasks (Hsieh et al., 2001; Gandour et al., 2003; Li et al., 2010), which have required subjects to attend to a particular component of the Mandarin syllable (namely, onsets, rimes, or tones) to decide whether they are the same or different. Because these tasks involve segmenting the syllable in order to extract particular components, they drive activation in prefrontal areas of the brain involved in attention and working memory (Burton, Small, & Blumstein, 2000). Consequently it remains an open question whether regions of brain areas involved in more automatic processing of auditory stimuli might also differ for tonal versus vowel processing (Joanisse, Zevin, & McCandliss, 2007). Therefore, the fMRI study reported in Chapter 2 did not involve active processing; rather subjects viewed a silent film while passively listening to Mandarin syllables. This experiment employed the same short-interval habituation paradigm reported in Joanisse et al. (2007), except deviant stimuli differed from repeated standards in either tone or vowels. This passive listening task was designed to drive activation of auditory cortical areas associated with automatic processing of auditory stimuli, and therefore allowed for isolation of subregions associated with tonal and vowel representations within these areas.

1.4.2 Do the Same Cognitive Processes Underlie Tonal versus Phonemic Access?

Next, I examined the cognitive processes involved in tonal versus phonemic access. To study this, I elected to employ an active processing task that required subjects to process spoken words in order to make a behavioral judgment. In contrast to the fMRI study, which focused on automatic perceptual processing of tones versus vowels, this task instead focused on how subjects use tonal versus phonemic information to constrain word recognition. As a result, rather than use fMRI, I recorded event related potentials (ERPs). ERPs are advantageous in this case because they have a higher temporal resolution than fMRI. Furthermore, certain components of the ERP waveform have been reliably associated with distinct stages of the word recognition process (Newman, Connolly, & Forbes, 2012). For example, the phonological mapping negativity (PMN; Connolly & Phillips, 1994; Newman & Connolly, 2009) is an earlier-going negativity associated with pre-lexical phoneme mapping, or the merging of top-down phonological expectations with bottom-up cues. The PMN is thought to reflect processes engaged prior to lexical access, as there are no apparent differences in PMN response between words and non-words (Newman & Connolly, 2009). In contrast, the N400 is a later-going negativity associated with whole-word processing (Kutas & Hillyard, 1984). Because these two components are thought to be distinct (Newman, Connolly, Service, & McIvor, 2003), this methodology therefore allowed me to gain insight into the cognitive processes underlying access to tonal versus phonemic information.

ERPs have been previously used to investigate spoken word processing in Mandarin (Zhao et al., 2011). However, the task employed in their study may not have been ideal for investigating the processes involved in word recognition. In their study, Zhao et al. (2011) asked subjects whether or not two pictures belonged to a semantic category. During presentation of these pictures, subjects heard an intervening auditory word that was phonologically related to one of the pictures, but this word was not critical to the task in any way. Therefore, subjects did not actively process auditory word forms in order to complete a task. As a result, Zhao et al. (2011) failed to observe any differences between tonal and phonemic processing, but this might simply be because the task they employed was not sensitive enough to detect them.

In the study reported in Chapter 3, I improved upon the design of Zhao et al. (2011) by employing a picture-word matching task. In this task, pictures of items cue subjects to generate expectations of incoming auditory stimuli, which are then confirmed or violated by subsequent auditory input (Desroches et al., 2009). Mismatches differed in the phonological relationship between presented and expected words, which allowed me to investigate how listeners resolved different types of phonological competition. Specifically, I compared responses to mismatches that differed in tone from expectations (segmental mismatches; see: *tang2* ‘candy’; hear: *tang1* ‘soup’) to mismatches that differed from expectations in word-final vowels (cohort mismatches; see: *tang2* ‘candy’; hear: *tao2* ‘peach’).

1.4.3 Is Mandarin Spoken Word Processing Incremental or Holistic?

Third, I tested whether Mandarin spoken word processing is incremental or holistic. As mentioned previously, evidence for incremental processing in Indo-European languages often consists of a difference in the competitive effects of onset-based cohort competitors versus rhyme competitors. However these effects might reflect the idiosyncrasies of the segmental structure of those languages, as they tend to have more permissive syllable structures yielding a greater variety of segmental word forms. Mandarin represents an ideal contrast to this, as its segmental structure is more restrictive. This has led researchers such as Zhao and colleagues (2011) to assert that Mandarin spoken word processing is holistic as opposed to incremental. This assertion was made on the basis of ERP data showing that responses to onset versus rime-based competitors were equivalent in adult native Mandarin speakers. However, due to the methodological limitations outlined in the previous section, it is possible that onset and rime-based competitors do differ in their competitive effects, and Zhao et al. (2011) failed to detect these differences. To address this issue, I also included rhyme competitors (see: *tang2* ‘candy’; hear: *lang2* ‘wolf’) in the ERP experiment reported in Chapter 3. I then compared ERP responses to cohort versus rhyme mismatches to test if subjects showed temporally dissociable effects to these two mismatch types.

1.4.4 Does Top-Down Information Influence Lexical Competition in Mandarin?

Another issue that is important from a theoretical standpoint is whether there is an influence of top-down processing on Mandarin spoken word processing. As discussed previously, an observation of rhyme effects is often taken as support for models that allow for top-down flow of information to affect the set of activated competitors for a spoken word. In ERP studies, these rhyme effects often take the form of an attenuation of the N400 component for rhyming compared to non-rhyming forms (Praamstra, Meyer, & Levelt, 1994; Radeau, Besson, Fontenau, & Castro, 1998; Coch, Grossi, Coffey-Corina, Holcomb, & Neville, 2002; Desroches et al., 2009). The logic here is that because rhyming words differ in onset phonemes, mismatch is signaled early on during the unfolding of a spoken word. As a result of inhibitory connections at the whole-word level, rhyme competitors are quickly suppressed; however, because of feedback connections from the lexical to phoneme layer, the phonemes that comprise the rhyme remain partially active. Therefore, processing of word-final phonemes is facilitated for rhyming words. This leads to a reduction in the N400 component, which is thought to index the extent of lexical competition that results for a particular word (O’Rourke & Holcomb, 2002). In light of this prior work, I tested for this influence of top-down feedback by comparing responses to rhyme and unrelated mismatches in the ERP study reported in Chapter 3.

1.4.5 How Does Spoken Word Processing Develop in Mandarin-Speaking Children?

Last, it is worth considering how spoken word processing develops in children, as these types of investigations can offer important theoretical constraints. To date, the development of spoken word processing in Chinese has been relatively understudied. Researchers have investigated children's awareness of the different components of the Mandarin syllable using behavioral measures (Shu, Peng, & McBride-Chang, 2008), but to date no study has looked at online processing of spoken words in Mandarin-speaking children. In the study reported in Chapter 4, I wished to answer the following two questions. First, do Mandarin-speaking children process spoken words incrementally or holistically? Second, does top-down information influence lexical competition in children to the same extent that it does in adults? I therefore collected data from a group of typically developing children and a group of adults using the same ERP picture-word matching task reported in Chapter 3, and compared groups on their responses to the different mismatch types.

1.5 Putting it All Together

The studies reported in this dissertation satisfy two aims. The first is to better understand how spoken words are processed in healthy adult native speakers of tonal languages, so that current theories and models of speech perception can be informed by these findings and ultimately more adequately account for tonal languages. Chapters 2 and 3 address this aim by providing important information regarding both the time course and neurobiology of spoken word recognition in Mandarin. The second aim is to elucidate the kinds of changes in spoken word processing that accompany development in Mandarin-speaking children. The study reported in

Chapter 4 addresses this aim by comparing a group of children and a group of adults in their responses to different types of phonological competition.

In Chapter 5, I propose a theoretical framework outlining how Mandarin spoken words are processed in healthy adults, taking into account the findings from Chapters 2 and 3. Subsequently, I use this theory to explain the findings observed in the children studied in Chapter 4, by making specific reference to the aspects of this theory that could most readily account for the observed differences between children and adults. I then close with some suggestions for future research directions.

1.6 References

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*, *12*(4), 679-690.

Cheung, H., Chung, K. K., Wong, S. W., McBride‐Chang, C., Penney, T. B., & Ho, C. S. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry*, *50*(6), 726-733.

Coch, D., Grossi, G., Coffey–Corina, S., Holcomb, P. J., & Neville, H. J. (2002). A developmental investigation of ERP auditory rhyming effects. *Developmental Science*, *5*(4), 467-489.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256-266.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, *59*(2), 165-179.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317-367.

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, *100*(3), B32-B42.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893-1906.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*(8), E505-E514.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207-222.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., & Tong, Y. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*(4), 185-200.

Goebel, R. (2007). Localization of brain activity using functional magnetic resonance imaging. In C. Stippich (Ed.), *Clinical functional MRI: Presurgical functional neuroimaging* (pp. 9-52). New York, NY: Springer.

Ho, C. S. H., Chan, D. W. O., Lee, S. H., Tsang, S. M., & Luan, V. H. (2004). Cognitive profiling and preliminary subtyping in Chinese developmental dyslexia. *Cognition*, *91*(1), 43-75.

Ho, C. S. H., Chan, D. W. O., Tsang, S. M., & Lee, S. H. (2002). The cognitive profile and multiple-deficit hypothesis in Chinese developmental dyslexia. *Developmental Psychology*, 38(4), 543.

Ho, C. S. H., Law, T. P. S., & Ng, P. M. (2000). The phonological deficit hypothesis in Chinese developmental dyslexia. *Reading and Writing*, *13*(1-2), 57-79.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30* (3), 129-148.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, *76*(3), 227-252.

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*(9), 2084-2093.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(312), 1-26.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, *13*(4), 646-653.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163.

Lee, J. R., Hung, D. L., & Tzeng, O. J. L. (2006). Cross-linguistic analysis of developmental dyslexia – does phonology matter in learning to read Chinese? *Language and Linguistics*, 7, 573-594.

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., & Dzemidzic, M. (2010). Hemispheric asymmetries in phonological processing of tones vs. segmental units. *Neuroreport*, *21*(10), 690.

Li, X., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Lowe, M., & Tong, Y. (2003). Selective attention to lexical tones recruits left dorsal frontoparietal network. *Neuroreport*, *14*(17), 2263-2266.

Liang, J., & van Heuven, V. J. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language*, *91*(3), 282-293.

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, *20*(12), 2958-2970.

Liu, W., Shu, H., & Yang, Y. (2009). Speech perception deficits by Chinese children with phonological dyslexia. *Journal of Experimental Child Psychology*, *103*(3), 338-354.

Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context: Evidence from Chinese ERP analyses. *Brain and Language*, *96*(1), 37-48.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1-36.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133-156.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71-102.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33-B42.

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*(7), 895-903.

Newman, E. H., Tardif, T., Huang, J., & Shu, H. (2011). Phonemes matter: The role of phoneme-level awareness in emergent Chinese readers. *Journal of Experimental Child Psychology*, *108*(2), 242-259.

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114-121.

Newman, R. L., Connolly, J. F., Service, E. & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event‐related brain potentials. *Psychophysiology*, *40*(4), 640-647.

Newman, R.L., Forbes, K., & Connolly, J.F. (2012). Event-related potentials and magnetic fields associated with spoken word recognition. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *Cambridge handbook of psycholinguistics*. New York, NY: Cambridge University Press.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357-395.

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology*, *60*(2), 121-150.

Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, *4*: 51.

Radeau, M., Besson, M., Fonteneau, E., Castro, S.L. (1998). Semantic, repetition and rime priming between spoken words: Behavioural and electrophysiological evidence. *Biological Psychology*, *48*, 183–204.

Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech perception: A cross‐linguistic study. *The Journal of the Acoustical Society of America*, *87*, S46.

Samuel, A. G., & Sumner, M. (2012). Current directions in research in spoken word recognition. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *Cambridge handbook of psycholinguistics*. New York, NY: Cambridge University Press.

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

Shu, H., Peng, H., & McBride‐Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science*, *11*(1), 171-181.

Siok, W. T., Perfetti, C. A., Jin, Z., & Tan, L. H. (2004). Biological abnormality of impaired reading is constrained by culture. *Nature*, *431*(7004), 71-76.

Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. *Advances in Psychology*, *90*, 151-172.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557-580.

Wang, M. I. N., & Cheng, C. (2008). Subsyllabic unit preference in young Chinese children. *Applied Psycholinguistics*, *29*(02), 291-314.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., Li, X. & Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173-183.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and* Cognitive *Processes Special Issue: Processing East Asian Languages, 14*(5-6), 609-630.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761-1770.

2 Investigating the Neurobiology of Mandarin Spoken Word Processing in Adults Using fMRI

2.1 Introduction

Of all the world's languages, Mandarin Chinese has the highest number of native speakers. Yet in spite of this prevalence, there is not an adequate theory that characterizes how spoken words are recognized in Mandarin. One reason for this is that we have only a limited understanding of how tones are recognized in the brain, relative to other types of phonetic units. Tone, which refers to the fluctuation of a speaker's voice during articulation of a syllable, is a critical feature of tonal languages like Mandarin because syllables carry different meanings depending upon their tone; for example, in Mandarin, *tang* means 'soup' when pronounced in a high and level tone, while it means 'candy' when pronounced in a rising tone. As a result, a viable theory of how spoken words are recognized must include tone.

In recent years, researchers have begun to investigate how tones are processed in the brains of native speakers of tonal languages (Gandour et al., 2000; Klein, Zatorre, Milner, & Zhao, 2001; Hsieh, Gandour, Wong, & Hutchins, 2001; Gandour et al., 2003; Xu et al., 2006; Li et al., 2010). In several of these studies, researchers have compared recognition of tones to recognition of phonemes, or the consonants and vowels that make up spoken words (Gandour et al., 2003; Li et al., 2010). The rationale for this is that much is known about the neural areas involved in phoneme recognition (Joanisse & Gati, 2003; Blumstein, Myers, & Rissman, 2005; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Joanisse, Zevin, & McCandliss, 2007; Myers, Blumstein, Walsh, & Eliassen, 2009; Liebenthal et al., 2010), and likewise, current theories of speech processing characterize phoneme processing in great detail (McClelland & Elman, 1986;

Norris & McQueen, 2008). Therefore, this work provides a useful reference point when looking at tonal processing.

In this line of research, an overarching question is whether the brain regions involved in representing tones and phonemes are shared or separate. One reason to expect that they might be distinct is that tones are thought to be suprasegmental, as they are often realized over multiple phonemic units. As detailed below, current evidence converges on the idea that tonal and phonemic processing are functionally dissociable in regions involved in attentive processing. However, it remains an open question whether tones and phonemes are also processed via distinct representations in brain areas involved in automatic processing of auditory stimuli. Because it has been shown that vowels carry tonal information in Mandarin (Howie, 1974), in the current study we have restricted our analysis to tonal versus vowel processing, addressing the following research question: to what extent are the brain regions involved in representing tone separate from the regions involved in representing the vowels on which tones are carried? To do this, we employed a passive listening paradigm to drive activation of auditory cortical areas involved in automatic processing. The conclusions from this study can be used to help guide researchers in how best to include tonal representations in theories of spoken word processing, as they give information as to the level of processing at which tonal units should be incorporated.

2.1.1 Functional Dissociation of Tonal and Phonemic Processing

As mentioned previously, neuroimaging evidence to date supports the idea that tones and phonemes are functionally dissociable. Part of this evidence has come from studies of word production. Liang and van Heuven (2004) reported the case of PYF, a 38-year old native Mandarin speaker who suffered damage to her left hemisphere. A CT scan revealed atrophy in the left precentral gyrus, located near premotor cortex, as well as the pars triangularis, located within the inferior frontal gyrus (IFG). Behaviorally, PYF showed an interesting profile: her tonal production was impaired, while her phonemic production was spared. This was confirmed by having her record Mandarin syllables, and then having naïve listeners identify the tone, consonants, or vowels associated with these syllables. The authors found that naïve listeners identified the tone of the syllables uttered by PYF at a lower accuracy than they identified the tone of syllables uttered by a healthy control subject. However, there was no difference in the identification rate for consonants and vowels in the syllables recorded by PYF versus those recorded by the healthy control subject. This was taken as evidence for a dissociation in the cortical areas involved in tonal versus phonemic specification, as PYF's damage appeared to be selective to tones. This finding was further supported by Liu et al. (2006), who used fMRI adaptation to examine the brain areas involved in tonal versus phonemic encoding. Subjects were asked to read aloud lists of Mandarin syllables while in the scanner, some of which differed from repeated standards in either tone or vowels. The authors observed a number of differences in the brain regions involved in tonal versus vowel processing, including left temporal areas, bilateral prefrontal areas, and bilateral cingulate gyrus.

More pertinent to the present investigation, complementary evidence has also come from studies of perception using PET and fMRI. These studies have required native Mandarin speakers to selectively attend to the tones of the first and last syllable within a string in order to judge whether they were the same or different (Gandour et al., 2000; Klein et al., 2001; Hsieh et al., 2001; Gandour et al., 2003; Li et al., 2003; Li et al., 2010). These studies have all pointed to frontoparietal areas as being critically involved in tonal processing. Additionally, some of these studies have also included trials in which subjects were asked to judge whether onsets and/or rimes were the same across groups of syllables (Gandour et al., 2003; Li et al., 2010). This manipulation allowed the authors to directly compare tonal and phonemic processing within Mandarin speakers. It was found that tones and phonemes are processed in slightly different areas within frontoparietal circuits. For example, Gandour et al. (2003) observed greater activation in the left posterior middle frontal gyrus (MFG) for vowels compared to consonants and tones. Li et al. (2010) observed that compared to onsets and rimes, tones were processed to a greater extent in right frontoparietal areas such as IFG and the inferior parietal lobule. These findings were taken to suggest that the brain areas involved in tonal versus phonemic processing are not entirely overlapping.

However, these prior studies have a significant limitation. Because subjects were required to extract particular phonetic components from syllables in order to perform overt judgments, attentional processes were engaged in these tasks (Burton, Small, & Blumstein, 2000). Therefore these tasks might have emphasized differences in activity in frontoparietal areas, as these are considered to be part of the attention network (Posner & Peterson, 1990). As a result of this limitation, it is currently unclear whether tonal and phonemic processing might also differ in areas associated with automatic processing of auditory stimuli (Myers et al., 2009; Joanisse et al., 2007).

2.1.1.1 Effects of Native Language Experience

Another important factor affecting the neurobiology of tonal versus segmental processing is the role of native language experience. For example, using PET, Klein, Zatorre, Milner, and Zhao (2001) had Mandarin and English speakers judge whether the tones of pairs of Mandarin syllables were the same or different. They observed increased cerebral blood flow in left frontoparietal areas during this task for the Mandarin speakers compared to the English speakers, while the English speakers exhibited increased cerebral blood flow in right prefrontal areas.

Similarly, Hsieh, Gandour, Wang, and Hutchins (2001) had listeners perform a same-different judgment task for sub-syllabic components of Mandarin syllables using a design that was similar to the Gandour et al. (2003) study. They found evidence for increased activity in left frontal regions for the Mandarin speakers for all types of syllable extraction, but also observed that for tone trials, English speakers actually showed increased activity in the right IFG. This involvement of the right IFG in tone processing for speakers whose native language is non-tonal has also been shown in tone training studies. For example, after a period of training with Mandarin lexical tones, the right IFG emerged as an area involved in Mandarin tone processing in native English speakers (Wang, Sereno, Jongman, & Hirsch, 2003). This was observed in addition to an increased extent of involvement of the left IFG. Therefore, there is a considerable body of evidence that native language experience plays a role in shaping the network of brain areas engaged in processing of tones and phonemes. As a result, an important component of the present study will involve assessing the extent to which language experience modulates brain activity in relevant cortical regions.

2.1.2 Rationale for the Current Study

In the current study, we employed an fMRI auditory short-interval habituation paradigm to investigate the brain areas involved in tonal versus phonemic processing of Mandarin syllables. In this paradigm, subjects are presented with a repeated stimulus, following which they are presented with a deviant stimulus that differs from the repeated standard in some important feature (Zevin & McCandliss, 2005). This is somewhat akin to fMRI visual adaptation studies, which employ the following logic. First, the neural response to a repeated stimulus becomes progressively attenuated over each successive presentation. Second, the presentation of a deviant stimulus results in a ‘release’ from adaptation (Grill-Spector & Malach, 2001). The extent of release from adaptation, measured as the magnitude of the difference in response between the standard and deviant, can be taken as an index of the degree to which a particular region is selective for the critical feature that differs between the standard and the deviant. In auditory short-interval habituation however, the logic is slightly different. Responses to repeated auditory items do not seem to progressively attenuate in the way that they do for visual items (Joanisse, Zevin, & McCandliss, 2007). So rather than looking at the degree to which a given region is released from adaptation, it is preferable to instead assess the extent to which a given region is involved in detecting auditory change (Zevin, Yang, Skipper, & McCandliss, 2010). This refers to a comparison between predicted stimulus features and the current description of a listener's sensory state (Zevin et al., 2010).

In the present study, we presented subjects with trains of stimuli, the first three of which were repeated. Following this, we presented subjects with a deviant stimulus that differed from repeated standards in either tone, vowels, or all phonemes and tone. Importantly, because this task involved passive listening, we were able to drive activation of auditory association areas in the absence of attentional modulation. Based on prior work, we rationalized that potential differences between tonal and vowel processing would arise not only from acoustic differences between tones and vowels, but also from effects due to native language experience with these distinctions (Gandour et al., 2002; Gandour et al., 2003; Xu et al., 2006). To tease these apart, we also scanned a group of native English speakers who were unfamiliar with any tonal language. Because the English speakers were unfamiliar with Mandarin syllables, regions that differed between tones and vowels in this group were ascribed to acoustic differences in the stimuli. Conversely, brain areas that differed for tones and vowels in the Mandarin speakers that showed a different pattern in the English speakers were ascribed to native language experience.

Critically, because we expected native language experience to modulate brain activity in areas relevant to the short-interval habituation task, we hypothesized that we would observe an interaction between speaker group and deviant type in several cortical regions.

2.2 Methods

2.2.1 Subjects

Sixteen native Mandarin speakers (eight female; mean age 24) and 16 native English speakers (eight female; mean age 27) participated in this study. Subjects were selected with the assistance of a native Mandarin speaker, who engaged in a brief telephone conversation with interested individuals to verify they were fluent in the standard Beijing dialect. All Mandarin speakers were born in Mainland China, and had been living in North America for an average of 3.4 years. Native English speakers were selected only if they reported that they were not proficient in a second language, and had no prior exposure to any of the Chinese dialects or other tonal languages of Southeast Asia.

2.2.2 Stimuli and Procedures

Auditory stimuli consisted of sets of spoken Mandarin words, beginning with a standard item (e.g., ‘zhai2’), from which two types of deviants were derived: tonal (same phonemes but different tone; e.g., ‘zhai4’) and vowel (same onset and tone but different vowels; e.g., ‘zhou2’). In addition, we selected unrelated deviants for each standard, which were chosen from within this set of items (i.e., tonal and vowel deviants for other standards) in order to have a closed set of stimuli. These unrelated deviants shared zero percent phonemic overlap and differed in tone from standards. The full list of stimuli is presented in Table 2.1.

Table 2.1: Standards and Deviants in Each of the Word Type Conditions

Standard	*Deviant*		
	Tonal	*Vowel*	*Unrelated*
cun4 /ts^{h}uən4/	cun1 /ts^{h}uən1/	cai4 /ts^{h}ai4/	tou2 /t^{h}əu2/
dian1 /tian1/	dian4 /tian4/	dui1 /tuei/	cai4 /ts^{h}ai4/
guo1 /kuo1/	guo2 /kuo2/	gou1 /kəu1/	zhai4 /tʂai4/
tian2 /t^{h}ian2/	tian1 /t^{h}ian1/	tou2 /t^{h}əu2/	gou1 /kəu1/
zhai2 /tʂai2/	zhai4 /tʂai4/	zhou2 /tʂəu2/	dian4 /tian4/

As is shown in Table 2.2, we selected stimuli that had a primary character that occupied a fairly large portion (at least 60%) of the total syllabic frequency of the syllable. These primary characters were associated with nouns. A repeated-measures ANOVA showed that deviants in the three conditions (tonal, vowel, and unrelated) did not differ from each other in percent of the total syllabic frequency occupied by the dominant character, as given in the *Modern Chinese Frequency Dictionary* (1986) [$F(2,8) = 0.551$, $p = .56$], nor did they differ in logarithmic syllabic frequency [$F(2,8) = 0.652$, $p = .49$], syllable duration [$F(2,8) = 0.454$, $p = .59$], nor number of homophones [$F(2,8) = 0.304$, $p = .68$].

Table 2.2: Parameters Associated with the Linguistic Stimuli, as given in *Modern Chinese Frequency Dictionary* (1986)

Syllable	*Syllabic Frequency*	*Primary Character*	*% of Syllabic Frequency Occupied by Primary Character*	*Duration (ms)*	*Number of Homophones*
cai4	431	菜 'vegetables'	100	328	0
cun1	826	村 'village'	100	408	0
cun4	102	寸 'inch'	100	353	0
dian1	66	颠 'peak'	92.4	304	2
dian4	2660	电 'electricity'	77.4	306	11
dui1	251	堆 'pile'	100	303	0
gou1	296	沟 'ravine'	60.1	309	3
guo1	162	锅 'pot'	91.4	321	1
guo2	6593	国 'nation'	100	348	0
tian1	5709	天 'sky'	97.8	389	1
tian2	575	田 'field'	70.6	417	3
tou2	4672	头 'head'	92.5	400	1
zhai2	76	宅 'house'	100	365	0
zhai4	77	债 'debt'	68.8	326	1
zhou2	62	轴 'axis'	95.2	366	2

Auditory stimuli were recorded as monosyllables spoken in isolation, as produced by a male native speaker of Mandarin. These stimuli were digitized and recorded to computer disk at 16 bits quantization, 44,100 Hz sample rate, and were volume normalized to -10 dB of maximum signal level. Each item was spoken seven times, and tokens selected for experimental use were all within the middle five of these utterances. Next, we concatenated syllables together to form trains of stimuli, comprised of a standard repeated three times, followed by a deviant (either tonal, vowel, or unrelated). For example, for the standard ‘zhai2’, the following three trains were constructed: zhai2-zhai2-zhai2-zhai4 (tonal); zhai2-zhai2-zhai2-zhou2 (vowel); zhai2-zhai2-zhai2-dian4 (unrelated). Between each syllable in the train, 50 ms of silence were inserted. Care was taken to ensure that no standard-deviant pairs formed disyllabic words in Mandarin (e.g., zhai2-zhai4 is not a Mandarin word). In addition, tone 3 words were avoided, as it is known that due to tonal sandhi, two tone 3 words do not follow each other in natural speech without an accompanying tonal change (Shen, 1990).

Using an fMRI-compatible set of earphones, stimulus trains were presented to subjects in the silent interval between scans (see Figure 2.1). In each run, there were ten trials per condition (tonal, vowel, and unrelated), as well as ten trials in which we presented silence. Trials were presented in a pseudorandom order, and occurred every four volumes, such that there were 14 s between trials. Each subject participated in six runs over the course of the experiment; therefore, there were 60 trials per condition. During scanning, subjects viewed an animated film presented silently without subtitles. They were instructed to simply view the film and not to attend to the auditory stimuli.

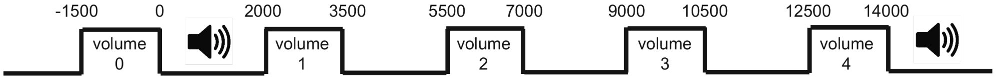


Figure 2.1: Schematic of the stimulus presentation paradigm

All testing protocols and materials were approved by the Health Science Research Ethics Board at The University of Western Ontario.

2.2.3 Imaging Acquisition and Analysis

Images were acquired using a 3T Siemens TIM Trio scanner at the Robart's Research Institute in London, Ontario, Canada. T2*-weighted functional scans were acquired in an axial orientation using multi-shot echo-planar imaging (EPI) with a 32 channel head coil with an iPAT parallel acquisition sequence (GRAPPA; acceleration factor = 2). We acquired 28 slices in an axial orientation (matrix size 64 x 64, 3 mm^3 voxels, flip angle = 90˚, FOV = 192, TE = 30 ms). Slices were parallel to the Sylvian fissure, and extended below it to the temporal pole and above it to cover most of the prefrontal and parietal cortex. This gave coverage of most of the brain except for superior portions of the parietal lobe and inferior portions of the cerebellum (see Figure 2.2 for a slice plan from a representative subject). We used sparse scanning such that sounds were played in the silent gaps between volume acquisitions (silent gaps were 2000 ms long, with a volume acquisition time of 1500 ms, yielding an effective TR of 3500 ms), and 160 volumes were collected for each of six runs. For each subject, we also acquired T1-weighted whole head anatomical scans (MPRAGE; GRAPPA acceleration factor = 2; matrix size 256 x 256, 1 mm^3 voxels, flip angle = 9˚, FOV = 256, TR = 2300 ms, TE = 2.98 ms).

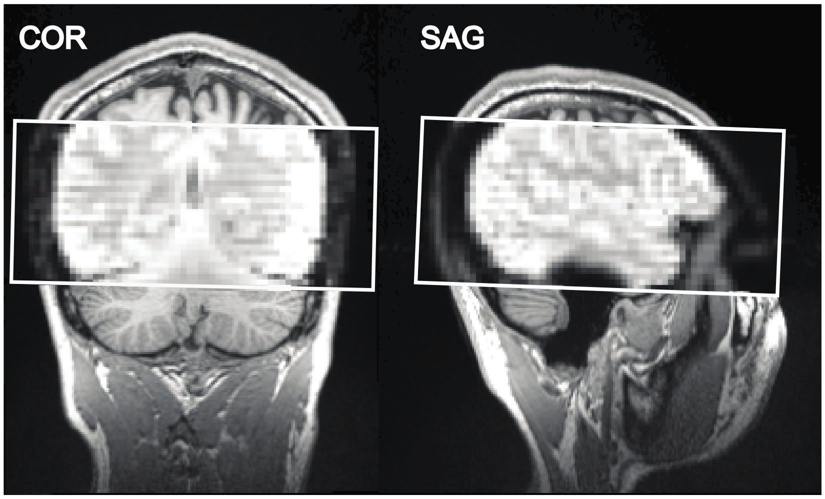


Figure 2.2 Slice plan of functional scans from a representative subject.

Functional data were analyzed using AFNI (Cox, 1996). Scans were first de-obliqued (AFNI *3dWarp*), and corrected for motion (AFNI *3dvolreg*) via a rigid-body transform, referenced to the first volume of the first run (coinciding with the acquisition of the anatomical scan). Next, the signal level of each run was normalized to percent signal change. Subject-wise scans were then analyzed using a general linear model (GLM; AFNI *3dDeconvolve*) with three predictors of interest (tonal trials, vowel trials, and unrelated trials) as well as one predictor of non-interest (a global estimate of motion, calculated by taking the root mean square of the six motion parameters). Stimulus times were defined as the onset of the fourth sound in each stimulus train, with a mean duration of 0.343 s. Volumes containing large signal spikes or dropouts (less than 1% of data points) were excluded from analyses by setting all predictor levels to zero.

Anatomical scans for each subject were co-registered to a standardized template (the N27 'Colin' brain) in Talairach coordinate space (Talairach & Tournoux, 1988). As it has been shown that Chinese and Caucasian groups differ with respect to morphometry (Tang et al., 2010; Chee et al., 2011), it was suspected that an affine transform to Talairach space might result in different levels of distortion across groups[1]. For this reason, each subject's brain was aligned to the N27 'Colin' brain using a diffeomorphic transformation, which is thought to preserve neuroanatomy to a greater extent than affine transforms when registering brains (Avants et al., 2011). Using Advanced Normalization Tools (ANTs, http://www.picsl.upenn.edu/ANTS/; Avants et al., 2011), we used greedy symmetric normalization to warp subjects' brains to the template (cross-correlation similarity metric; Gaussian regularizer with a sigma of 3; 100 x 50 x 30 x 10

---

[1] Indeed, when we used a 12-parameter affine transform to register each subject's brain to Talairach space using the N27 'Colin' brain as a template (AFNI @auto_tlrc), three of the twelve affine parameters differed between groups of speakers at $p < .01$.

iterations; step size of 0.25). Next, for each subject, statistical maps were first resampled to 1 mm^3, and then registered to Talairach space using the same warp as the anatomical scan. Following this, statistical maps were brain-masked using each subject's warped anatomical scan, and then spatially smoothed using a 2 mm FWHM Gaussian kernel.

Groupwise statistical maps were created using a two-way mixed design analysis of variance (AFNI *3dANOVA3*) with the between-subjects factor of group (Mandarin speakers, English speakers) and the within-subjects factor of deviant type (tonal, vowel). For any clusters that showed a significant interaction between group and deviant type, we investigated the nature of the interaction by comparing the tone and vowel conditions separately in each group using paired samples t-tests. All statistical maps were thresholded at an uncorrected alpha level of $p = .005$. To correct for multiple comparisons, we calculated a minimum cluster size using a Monte Carlo simulation (AFNI *AlphaSim*; 10,000 iterations) using a mask of all brain voxels within one volume of an EPI sequence from one of the subjects in native space. This simulation provided a corrected alpha level of $p = .05$ for a minimum cluster size of 95 mm^3. Brain regions were identified using the Talairach applet (University of Texas Health Science Centre San Antonio; http://www.talairach.org/applet). Reported Talairach coordinates indicate the peak of each cluster.

2.3 Results

Statistical analyses focused on main effects of native-speaker group (Mandarin, English) and deviant type (Tonal, Vowel), as well as the interaction of these two factors. Results from a 2 (Group) by 2 (Deviant Type) ANOVA are presented in Table 2.3. Five regions (right MTG, left insula, right IFG, right precentral gyrus, and right caudate) showed an interaction between group

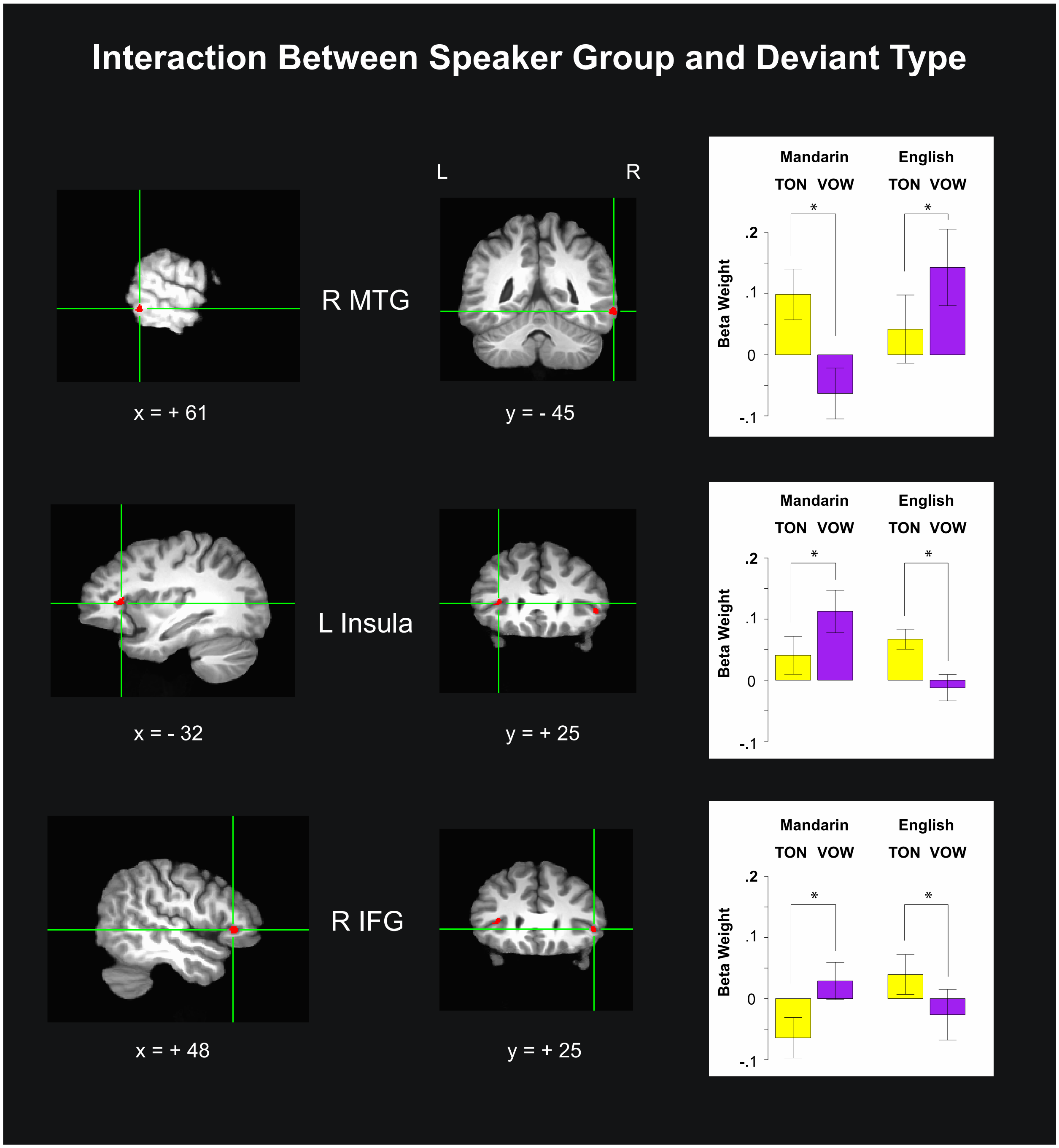
Table 2.3: Results for the Speaker Group by Deviant Type ANOVA

Main Effect or Interaction	*Region*		*Talairach coordinates*[a]				
	L/R	*Area*	*x*	*y*	*z*	*Size (mm^3)*	*BA*
Group x Deviant Type Interaction	R	Middle temporal gyrus	61	-46	-9	108	37
	L	Insula	-35	25	8	95	13
	R	Inferior frontal gyrus	48	25	2	95	45
	R	Precentral gyrus	53	0	16	108	6
	R	Caudate	16	5	5	127	–
Main Effect of Speaker Group MDN > ENG	L	Superior temporal gyrus	-47	2	1	138	22
	L	Inferior frontal gyrus	-43	24	14	100	45
	L	Middle frontal gyrus	-45	7	37	123	9
	R	Middle frontal gyrus	41	45	5	186	46
	R	Middle frontal gyrus	34	13	41	102	6
	–	Medial frontal gyrus	0	31	35	106	6
	R	Postcentral gyrus	37	-30	45	114	40
	L	Cingulate gyrus	-12	17	37	127	32
	L	Cingulate gyrus	-15	-17	39	97	24
	L	Precuneus	-19	-47	42	106	7
	R	Precuneus	13	-69	27	110	31
	L	Thalamus	-19	-33	2	138	–
	R	Thalamus	12	-18	19	109	–
	R	Corpus callosum	34	21	24	539	–
ENG > MDN	R	Superior temporal gyrus	59	-32	12	221	22
Main Effect of Deviant Type TON > VOW	L	Middle frontal gyrus	-29	53	5	110	10
	R	Middle frontal gyrus	51	28	20	129	46
	R	Middle frontal gyrus	44	5	39	96	6
	L	Medial frontal gyrus	-1	60	17	211	10
	L	Medial frontal gyrus	-3	49	29	134	9

Main Effect or Interaction	*Region*		*Talairach coordinates*[a]			*Size (mm³)*	*BA*
	L/R	*Area*	*x*	*y*	*z*		
VOW > TON	L	Inferior parietal lobule	-52	-39	47	98	40
	L	Cingulate gyrus	-19	26	33	143	32
	L	Posterior cingulate	-2	-27	19	366	23
	L	Parahippocampal gyrus	-25	-49	5	106	30
	R	Caudate	24	-40	12	213	–
	R	Caudate	4	15	9	171	–
	–	Corpus callosum	2	2	19	202	–

[a] Talairach coordinates indicate the peak of each cluster.

and word type (Table 2.3; Figure 2.3). For each cluster that showed an interaction, we analyzed beta weights separately for each group within these regions of interest, as illustrated in Figure 2.3. This analysis revealed that the area in right posterior MTG showed greater activation for tones compared to vowels in Mandarin speakers [*t*(15) = 2.870, *p* = .01], while the reverse pattern held for English speakers [*t*(15) = -3.379, *p* = .004]. Conversely, the other four areas in which we observed an interaction (left insula, right inferior frontal gyrus, right precentral gyrus, and right caudate) showed greater activation for vowels compared to tones in the Mandarin speakers [left insula: *t*(15) = -2.518, *p* = .02; right IFG: *t*(15) = -5.387, *p* < .001; right precentral gyrus: *t*(15) = -3.698, *p* = .002; right caudate: *t*(15) = -3.162, *p* = .01], and the reverse in the English speakers [left insula: *t*(15) = 3.496, *p* = .003; right IFG: *t*(15) = 2.206, *p* = .04; right precentral gyrus: *t*(15) = 3.002, *p* = .01; right caudate: *t*(15) = 3.380, *p* = .004].


Interaction Between Speaker Group and Deviant Type
L
R
R MTG
x = + 61
y = - 45
L Insula
x = - 32
y = + 25
R IFG
x = + 48
y = + 25
Mandarin
English
TON VOW
TON VOW
*
Beta Weight
.2
.1
0
-.1

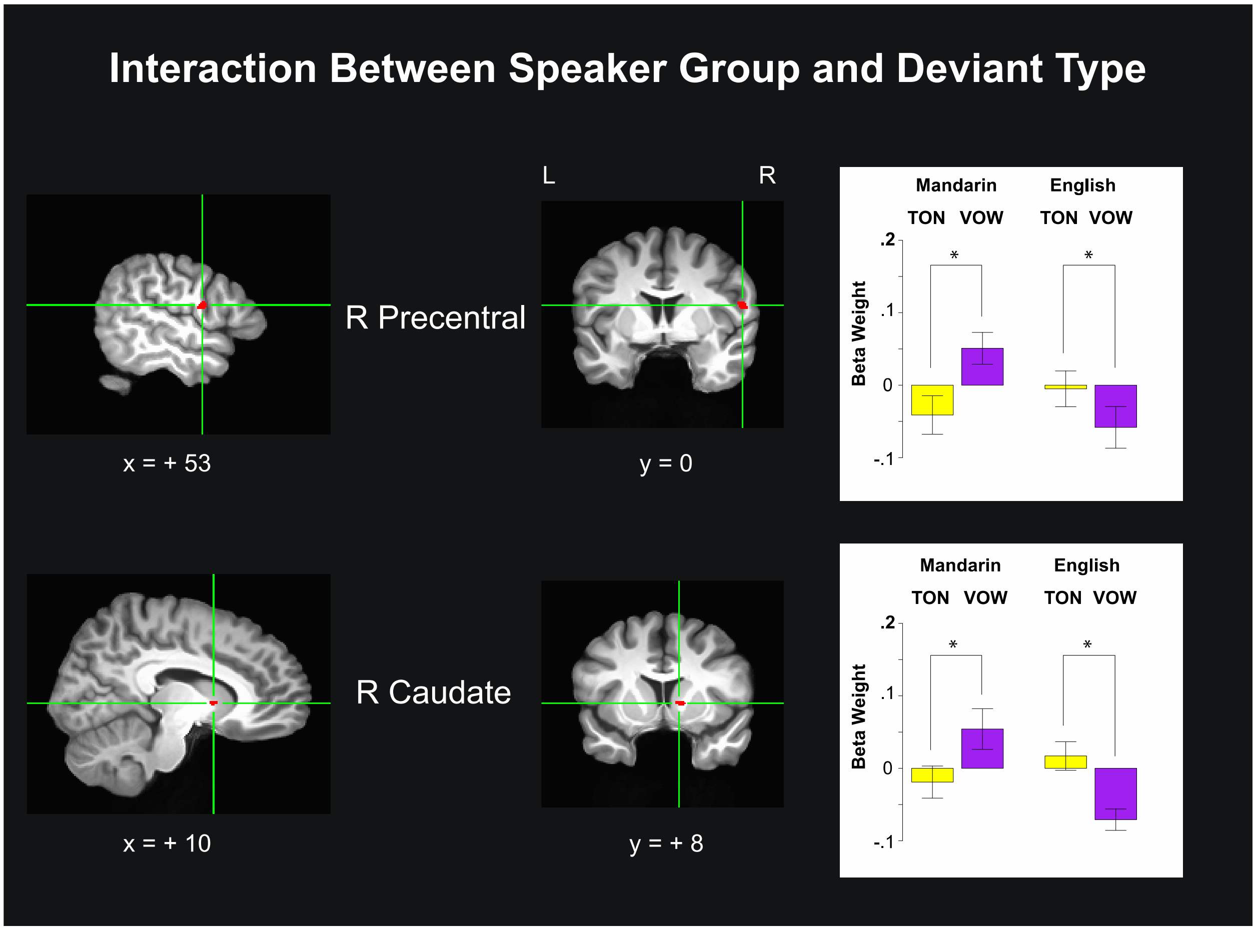


Figure 2.3: Clusters that showed a significant interaction between speaker group and deviant type. The voxelwise threshold was set to $F = 9.16$, $p < .005$ (uncorrected). Only clusters significant at $p < .05$ (corrected) are shown. Statistical maps are overlaid on an average brain, which was calculated by first using a diffeomorphic transform to warp each subject's brain to the N27 Talairach template, and then taking the average warped brain across all subjects.

Additionally, a number of regions showed main effects of speaker group (Table 2.3; Figure 2.4). Notably, greater activation was observed in Mandarin speakers compared to English speakers in a region of left anterior STG. Interestingly, greater activation was observed in English speakers compared to Mandarin speakers in a region of right posterior STG. Furthermore, Mandarin speakers showed greater activation than English speakers in a number of frontal regions, including left IFG and bilateral MFG. Mandarin speakers also activated the precuneus bilaterally to a greater extent than English speakers.

Last, several regions also showed main effects of deviant type (Table 2.3; Figure 2.5). Namely, tones showed greater activation than vowels in bilateral MFG as well as the inferior parietal lobule, while vowels showed greater activation than tones in areas such as left posterior cingulate gyrus and subcortical areas such as the right caudate.

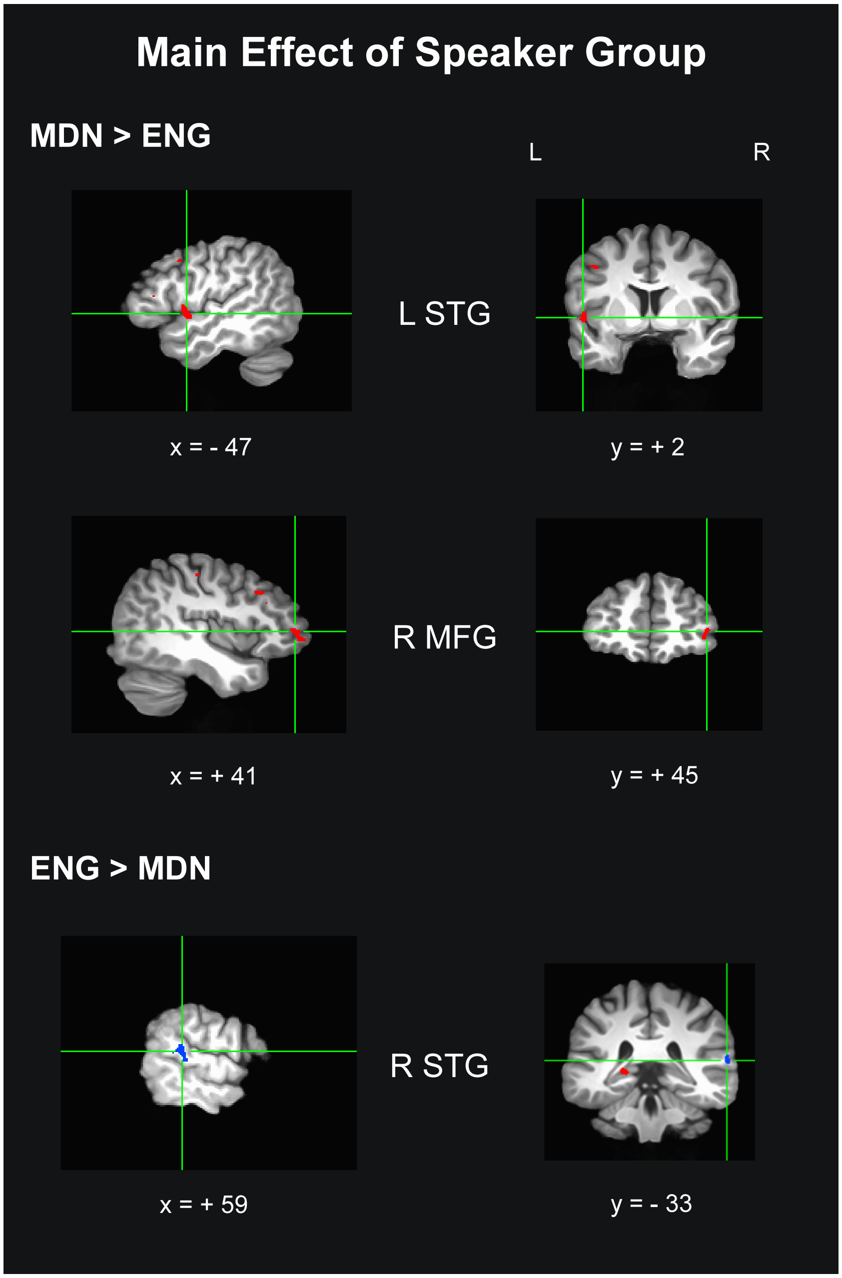


Figure 2.4: Clusters that showed a significant main effect of speaker group. The voxelwise threshold was set to $t = 3.03$, $p < .005$ (uncorrected). Only clusters significant at $p < .05$ (corrected) are shown. Statistical maps are overlaid on an average brain, which was calculated by first using a diffeomorphic transform to warp each subject's brain to the N27 Talairach template, and then taking the average warped brain across all subjects.

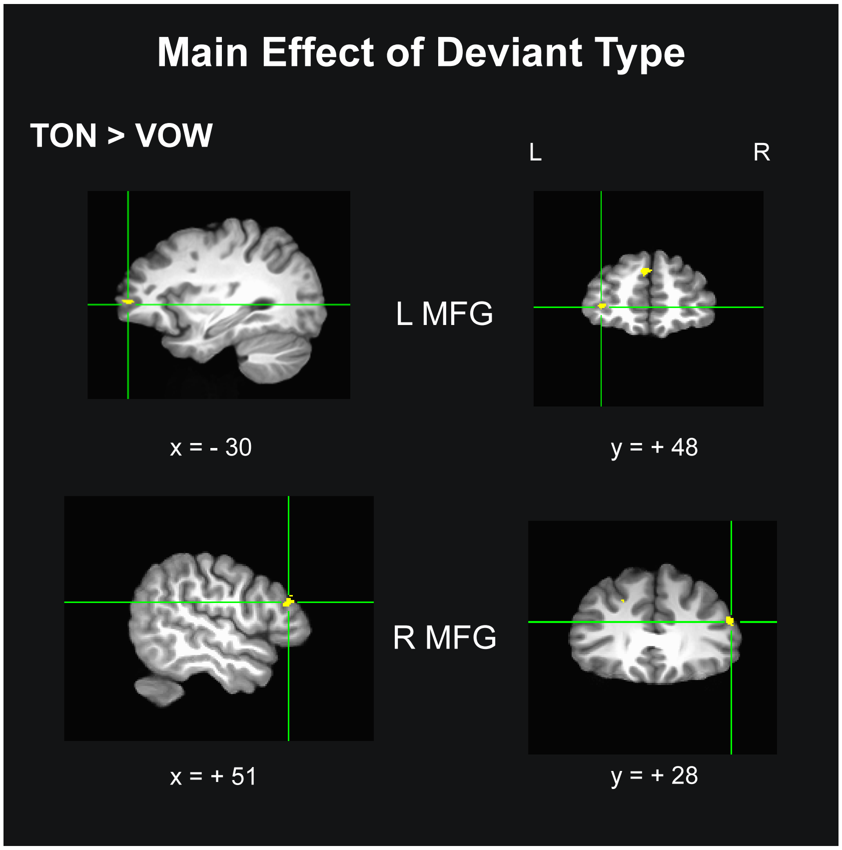


Figure 2.5: Clusters that showed a significant main effect of deviant type. The voxelwise threshold was set to $t = 3.03$, $p < .005$ (uncorrected). Only clusters significant at $p < .05$ (corrected) are shown. Statistical maps are overlaid on an average brain, which was calculated by first using a diffeomorphic transform to warp each subject's brain to the N27 Talairach template, and then taking the average warped brain across all subjects.

2.4 Discussion

The key motivation for this study was to assess the extent to which the cortical regions involved in representing Mandarin tones are separate from those involved in representing Mandarin vowels. This was done with the aim of informing current theories of spoken word processing as to how to incorporate tonal information in a way that captures how native speakers handle this type of information. Previous studies have identified differences in the brain regions involved in tonal versus vowel processing in native Mandarin speakers; however they have tended to use overt segmentation tasks that require attention and working memory (Gandour et al., 2000; Klein et al., 2001; Hsieh et al., 2001; Gandour et al., 2003; Li et al., 2003; Li et al., 2010). The novel contribution of this study is the use of a passive listening task that does not involve extraction of syllable components. As a result, we would argue that we were able to observe differences in tonal versus phonemic processing in the absence of attentional modulation (Myers et al., 2009; Joanisse et al., 2007). The paradigm itself involved short-interval habituation to trains of auditory stimuli, and we analyzed activation following presentation of a deviant stimulus that differed from the repeated standard in either tone or vowels. Importantly, the standard was common across conditions, and so differences in activation between the deviant type conditions can be attributed to the different types of phonological change associated with each deviant type (Zevin & McCandliss, 2005).

Critically, we compared how native Mandarin speakers processed tonal and vowel categories in their native language to how a group of English speakers unfamiliar with these categories processed the same acoustic information. This allowed us to disentangle effects of native language experience from effects due to acoustic differences in tonal versus vowel deviants. To this end, we performed statistical analyses to test for main effects of speaker group (Mandarin, English) and deviant type (Tones, Vowels), as well as the interaction of these two factors. Most importantly, we wished to identify regions that showed an interaction between speaker group and deviant type. The rationale for this was that we expected both groups of speakers to show differences in activity between tones and vowels, but expected these differences to be modulated by native language experience. Thus we wished to identify regions that differed according to deviant type in divergent ways across speaker groups.

Analyses uncovered several regions that showed an interaction between speaker group and deviant type. Namely, activity in the right posterior middle temporal gyrus was greater for tones than vowels in the Mandarin speakers, while the reverse pattern was observed in English speakers. While posterior MTG activity is typically left lateralized in studies of speech perception (Davis & Johnsrude, 2003), some studies have shown right hemisphere activation in this region for speech stimuli (Uppenkamp, Johnsrude, Norris, Marslen-Wilson, & Patterson, 2006), including studies of auditory change detection (Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003; Zevin & McCandliss, 2005; Joanisse et al., 2007; Zevin et al., 2010). This activity could reflect differences in lexical semantic processing between the different deviants in Mandarin (Hickok & Poeppel, 2007). However, a difference in activation between deviant types in the English speakers suggests that this region is responsible for aspects of acoustic processing as well, as the English speakers were not expected to associate the presented syllables with lexical semantic representations. We would thus argue that this area was partially involved in phonological change detection in the Mandarin speakers.

In addition to the right posterior MTG, other right hemisphere areas showed an interaction as well, including the right IFG and right precentral gyrus. In both these areas, activation was greater for tones than for vowels in the English speakers, while the reverse pattern was observed in Mandarin speakers. A greater amount of right hemisphere activation for tones compared to vowels in English speakers could be reflective of the role of the right hemisphere in extracting pitch information (Zatorre, Evans, Meyer, & Gjedde, 1992), as it has been observed that English speakers unfamiliar with tonal categories engage pitch processing mechanisms in the right hemisphere when processing lexical tones (Wong, Parsons, Martinez, & Diehl, 2004). Importantly, this difference between tones and vowels was actually the opposite of what was observed for the Mandarin speakers and again suggests this effect was driven by acoustic rather than phonological discrimination of presented sounds (Jacquemot et al., 2003).

The observed interaction in the left insula was also interesting, as left insular cortex plays a role in phonological word recognition (Bamiou, Musiek, & Luxon, 2003), and has been previously shown to be involved in detection of phonological as opposed to acoustic change (Jacquemot et al., 2003). The current results suggest that this area is involved in vowel processing to a greater extent than tonal processing in native Mandarin speakers. This supports work by Gandour and colleagues (2003), who found that the left anterior insula was involved in phonological processing of native speech categories in Mandarin, although they did not observe a difference between tonal and vowel deviants in this region.

Together, these findings suggest that the brain regions involved in tonal versus vowel processing are not entirely overlapping in native Mandarin speakers even during a passive listening task. A similar finding has been observed in an ERP study using the mismatch negativity (MMN). Luo et al. (2006) presented native Mandarin speakers with deviant stimuli that differed from repeated standards in either word-initial consonants or tones, and found that depending on the deviant type, the MMN was maximal over different electrodes. As the fMRI short-interval habituation task employed in the present study is a modification of the MMN paradigm optimized for fMRI

(Zevin & McCandliss, 2005), both the current study and the Luo et al. (2006) study likely engaged similar brain regions corresponding to cortical areas involved in formulating and updating a predictive model of the acoustic environment (Näätänen, Paavilainen, Rinne, & Alho, 2007). One key difference between studies however is that Luo et al. (2006) observed a difference in terms of lateralization of tonal versus phonemic deviants, in that MMN amplitude was greater in the right as compared to the left hemisphere for tones, while the reverse pattern was observed for consonants. Importantly, we did not observe this same pattern of lateralization. However, it has been previously shown that consonants and vowels differ in terms of degree of lateralization (Joanisse & Gati, 2003), with consonants activating a region of left STG to a greater extent than vowels. Therefore, had we used consonants rather than vowels, our results may have been in greater accordance with those of Luo et al. (2006).

The current results are also consistent with another recent MMN study in which native Mandarin speakers were presented with different kinds of tonal deviants (Xi, Zhang, Shu, Zhang, & Li, 2010). The authors, who were interested in categorical perception of tonal information in Mandarin, presented native Mandarin speakers with tonal deviants that fell within a category boundary, or crossed it. The amplitude of the MMN was larger for across-category deviants than within-category deviants in the left hemisphere. Given that the left hemisphere is associated with linguistic processing (Scott, Blank, Rosen, & Wise, 2000), this was taken as evidence that phonological representations associated with tonal categories are activated very early in auditory cortical processing in native Mandarin speakers (Xi et al., 2010). By contrast, native Korean speakers who had no knowledge of Mandarin tones did not show this effect.

The clusters that showed a main effect of speaker group gave further confirmation that phonological representations associated with tonal and vowel categories were activated during this task. As can be seen in Figure 2.4, English speakers recruited a greater extent of right posterior STG than did Mandarin speakers. In addition, Mandarin speakers showed greater activation compared to English speakers in an anterior portion of left STG. We argue that a lifetime of experience with phonetic distinctions in Mandarin resulted in less right posterior STG activation in these individuals compared to the English speakers. Thought of in another way, Mandarin speakers showed a greater leftward shift in activation of posterior STG. A similar type of asymmetry in this region has been observed previously in tonal language speakers when processing native tonal categories (Xu et al., 2006). In the Xu et al. (2006) study, Thai and Mandarin tones were superimposed onto Mandarin syllables, which were presented to native Thai and Mandarin speakers in an fMRI experiment. The authors observed a double dissociation in the left planum temporale (PT), in that speakers showed greater activation in the left PT when processing native tonal categories. The location of the PT is spatially very close to the cluster of posterior STG that differed between groups of speakers, and it is possible that the two groups in the present study differed in activation of right PT. As it has been argued that the PT is involved with matching incoming auditory information to learned templates among other tasks (Griffiths & Warren, 2002), Mandarin speakers could have shown a greater leftward shift in this region because they were making contact with phonological representations stored in the left hemisphere. Another possibility, which is not mutually exclusive, is that the English speakers recruited additional areas in the right hemisphere because they were processing non-native contrasts which might have been difficult for their perceptual systems to integrate. Similar findings have been observed by Guenther, Nieto-Castanon, Ghosh, and Tourville (2004), who presented native English speakers with prototypical and non-prototypical examples of the vowel /i/. They found increased activation in auditory association cortex when subjects listened to non-prototypical examples, especially in right Heschl's gyrus and right PT. Conversely, processing of prototypical exemplars resulted in more focal activation in these areas, similar to what was seen for the Mandarin speakers in the present study.

In addition to these differences in superior temporal areas, analyses revealed differential activation across speaker groups in other brain regions as well. Specifically, Mandarin speakers showed greater activation than English speakers in left IFG and bilateral MFG. This could be the result of a greater amount of attentional engagement for the Mandarin compared to the English speakers, as the stimuli were real words in Mandarin. It could be that Mandarin speakers were thus involved in more active discrimination of these categories compared to the English speakers even though the task involved passive listening. Furthermore, Mandarin speakers activated the precuneus bilaterally to a greater extent than English speakers. This could be the result of engagement of regions involved in mental imagery, as the Mandarin speakers were familiar with the spoken syllables and associated them with particular meanings, and the precuneus is argued to be involved in visual imagery among other functions (Cavanna & Trimble, 2006). Thus it would appear that even though this was a passive listening task, the native Mandarin speakers still activated lexical semantic features associated with these words.

In spite of this limitation, we observed differences in activity between tones and vowels in several brain regions involved with detecting change at the phonological level. Therefore, we are in a position to draw several conclusions that are useful for word recognition theories. First, as supported by prior studies, an adequate theory of Mandarin spoken word processing should include some units involved in representing tonal categories that are separate from those involved in processing vowel categories (Ye & Connine, 1999; Liang & van Heuven, 2004; Zhao, Guo, Zhou, & Shu, 2011). Second, just like vowels, tones are processed as categories even in the absence of attentional modulation (Wang, Gu, He, Chen, & Chen, 2012; Xi et al., 2010). This supports the idea that early cortical processing is selectively tuned to the acoustic dimensions underlying tonal contrasts (Chandrasekaran, Krishnan, & Gandour, 2007). These conclusions can help offer guidance to researchers attempting to incorporate tonal representations into theories of spoken word recognition by offering critical neurobiological constraints.

2.5 Conclusions

The current study provides evidence that the cortical areas involved in tonal processing are not entirely overlapping with those involved in vowel processing, even during a task that does not require conscious attentive processing. In a short-interval habituation task, native Mandarin speakers showed differences in activation between tones and vowels in several regions, including right posterior MTG, right IFG, and left insula. Importantly, English speakers unfamiliar with these phonetic distinctions did not show the same pattern of activation in these areas; therefore, differences in activity in the Mandarin speakers can be ascribed to detection of phonological as opposed to acoustic change. These results have important implications for theories of spoken word processing, as they suggest that while many units for tone and vowel processing are inevitably shared, an adequate theory should incorporate some processing units that are solely responsible for handling tonal information.

2.6 References

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage, 54(3), 2033-2044.

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing: literature review. *Brain Research Reviews*, *42*(2), 143-154.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353-1366.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience*, *12*(4), 679-690.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564-583.

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

Chee, M. W. L., Zheng, H., Goh, J. O. S., Park, D., & Sutton, B. P. (2011). Brain structure in young and old East Asians and Westerners: comparisons of structural volume and cortical thickness. *Journal of Cognitive Neuroscience*, *23*(5), 1065-1079.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162-173.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, *23*(8), 3423-3431.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207-222.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., & Li, X. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience*, *14*(7), 1076-1087.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., & Tong, Y. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*(4), 185-200.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*(7), 348-353.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*(1), 293-321.

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language and Hearing Research*, *47*(1), 46.

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

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30*(3), 129-148.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, *76*(3), 227-252.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, *23*(29), 9541-9546.

Joanisse, M. F., & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage*, *19*(1), 64-79.

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*(9), 2084-2093.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, *13*(4), 646-653.

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., & Dzemidzic, M. (2010). Hemispheric asymmetries in phonological processing of tones vs. segmental units. *Neuroreport*, *21*(10), 690.

Li, X., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Lowe, M., & Tong, Y. (2003). Selective attention to lexical tones recruits left dorsal frontoparietal network. *Neuroreport*, *14*(17), 2263-2266.

Liang, J., & van Heuven, V. J. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language*, *91*(3), 282-293.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*(10), 1621-1631.

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. C*erebral Cortex*, *20*(12), 2958-2970.

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., & Chen, C. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *Neuroimage*, *29*(2), 515-523.

Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, *103*(51), 19558-19563.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

Modern Chinese Frequency Dictionary. (1986). Beijing: Beijing Language Institute Publisher [in Chinese].

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*(7), 895-903.

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

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400-2406.

Shen, X. N. S. (1990). *The Prosody of Mandarin Chinese*. University of California Press.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.

Tang, Y., Hojatkashani, C., Dinov, I. D., Sun, B., Fan, L., Lin, X., ... & Toga, A. W. (2010). The construction of a Chinese MRI brain atlas: A morphometric comparison study between Chinese and Caucasian cohorts. *Neuroimage*, *51*(1), 33-41.

Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech–sound processing in human temporal cortex. *NeuroImage*, *31*(3), 1284-1296.

Wang, X. D., Gu, F., He, K., Chen, L. H., & Chen, L. (2012). Preattentive extraction of abstract auditory rules in speech sound stream: a mismatch negativity study using lexical tones. *PloS ONE*, *7*(1), e30027.

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

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

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223-231.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., Li, X. & Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173-183.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and* Cognitive *Processes Special Issue: Processing East Asian Languages, 14*(5-6), 609-630.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846-849.

Zevin, J. D., & McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions*, *1*(4).

Zevin, J. D., Yang, J., Skipper, J. I., & McCandliss, B. D. (2010). Domain general change detection accounts for “dishabituation” effects in temporal–parietal regions in functional magnetic resonance imaging studies of speech perception. *The Journal of Neuroscience*, *30*(3), 1110-1117.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761-1770.

3 Investigating the Time Course of Mandarin Spoken Word Processing in Adults Using ERPs[*]

3.1 Introduction

Even though there is an increasing interest in the psycholinguistics literature concerning processing Chinese languages (Li, Tan, Bates, & Tzeng, 2006), it remains the case that the predominant theories and models of speech perception have been developed primarily using data from Indo-European languages such as English and Dutch. These languages are different in nature from tonal languages such as Mandarin Chinese, and as a result, current theories might account for speech perception in only a subset of the world's speakers; namely, speakers of non-tonal languages. To address this problem, a number of studies have begun to focus on how and when lexical tone is processed in the brain, as tone is the key feature that distinguishes tonal languages such as Mandarin Chinese from non-tonal languages such as English (Gandour, Xu, Wong, Dzemidzic, Lowe, Li, & Tong, 2003; Klein, Zatorre, Milner, & Zhao, 2001; Luo, Ni, Li, Li, Zhang, Zeng, & Chen, 2006; Malins & Joanisse, 2010; Zhao, Guo, Zhou, & Shu, 2011). In addition, the structure of Mandarin syllables differs substantially from the structure of English syllables, and some studies have looked at how these differences in syllabic structure might also lead to differences in spoken word recognition between the two languages (Liu, Shu, & Wei, 2006; Wang & Cheng, 2008; Zhao et al., 2011). While these studies have outlined some of the ways by which theories of spoken word recognition might be updated to include Mandarin, there remain a number of outstanding issues in the field. The current study uses ERPs to address the following three issues: (1) whether the processes underlying tonal versus phonemic access differ;

---

[*] This chapter has been published with the following title: *Setting the Tone: An ERP Investigation of the Influences of Phonological Similarity on Spoken Word Recognition in Mandarin Chinese*.

(2) whether onsets and rimes play similar roles in Mandarin word recognition as they do in English; (3) whether, as some have suggested, Mandarin syllables are processed more holistically than their English counterparts.

3.1.1 Access to Tonal Information During Spoken Word Processing

Lexical tone refers to variation in the fundamental frequency of a speaker's voice that is used to differentiate phonemically identical words. For instance, Mandarin Chinese includes four lexical tones: high and level (tone 1), mid-rising (tone 2), low-dipping (tone 3), and falling (tone 4), and a syllable can carry different meanings depending on the tone that is used. For example, *ma1* means 'mother', yet *ma3* means 'horse' (numerals indicate the tone of the syllable). Thus to access the meaning of a spoken word, Mandarin listeners must process not only the phonemes that comprise the word (/m/ and /a/ in the previous example), but the tone of the syllable as well. Tonal information is denoted as suprasegmental, as tone is best described as being associated with an entire syllable rather than an individual phoneme, and its effect can span multiple phonemes. For example, in Mandarin a tonal contour can be realized in the same way over very different vowel clusters such as /iɑu/ versus /uei/. Tonal languages are particularly different from non-tonal languages in that a suprasegmental cue is being used pervasively to distinguish lexical items from each other. This is not to say that non-tonal languages do not make use of suprasegmental information; indeed, Dutch and Japanese are considered non-tonal yet do make considerable use of the suprasegmental features of stress (van Donselaar, Koster & Cutler, 2005) and pitch accent (Cutler & Otake, 1999). However, tonal languages like Mandarin represent an extreme case of the use of suprasegmental information to constrain word recognition on a syllable-by-syllable basis. Tonal information thus plays a key role in this group of languages and must be accounted for in a viable theory of spoken word recognition.

Critically, one of the questions of particular interest from a theoretical standpoint is whether the cognitive processes underlying tonal versus phonemic access are the same or different. A number of early studies using behavioral measures of word recognition such as accuracy and reaction time have suggested that tonal information is accessed later than phonemic information and is a weaker cue for word recognition (Repp & Lin, 1990; Taft & Chen, 1992; Cutler & Chen, 1997; Ye & Connine, 1999 – Experiment 1). However, later access to tone does not necessarily indicate different processes from those involved in phonemic access. As discussed below, recent studies using online language measures have suggested some equivalence among the two.

One type of online language measure that has been used extensively in recent years is associated with the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In a typical study, subjects are presented with pictures of items, and then hear a spoken word that corresponds to one of the pictures. As the auditory word unfolds, eyetracking is used to monitor eye movements to the visual display. Of interest is both the time course of eye movements to the target picture (e.g., a picture of a *cat*), but also looks to the other pictures that are phonological competitors to this target (e.g., a picture of a *cap*).

We have recently used this methodology to examine auditory word recognition in Mandarin (Malins & Joanisse & 2010). Subjects were presented with arrays of pictures that included target items along with segmental competitors, which shared all phonemes but differed in tone from targets (e.g., *shu3* ‘mouse’ vs. *shu1* ‘book’), as well as cohort competitors, which shared word-initial phonemes and tone with targets but differed in word-final vowels (e.g., *qiu2* ‘ball’ vs. *qian2* ‘money’). Acoustic analyses suggested the point of divergence of segmental competitors from targets (in tone) closely approximated the point of divergence of cohort competitors from targets (in phonemes), as tone is thought to be carried predominantly on the vowels of syllables

(Howie, 1974). Likewise the eye movement data showed that competitor effects were comparable between these segmental and cohort conditions, providing evidence that tones and phonemes are accessed simultaneously.

However, a lack of difference in processing time or levels of activation (as indexed by looks to items) is not necessarily an indicator that equivalent cognitive processes underlie these effects. Indeed, it could be the case that tones versus phonemes are processed in parallel but within slightly different neural circuits. A technique that offers additional insight into these underlying processes is the use of event related potentials (ERPs). Specific components of the ERP waveform have been reliably associated with particular cognitive events, each of which has its own characteristic latency and scalp distribution (Newman, Forbes, & Connolly, 2012). Therefore the use of ERPs is ideally suited to uncovering potential differences in processing that may underlie access to tonal versus phonemic information during spoken word recognition.

Schirmer, Tang, Penney, Gunter, and Chen (2005) used ERPs to investigate tonal versus phonemic processing in Cantonese, a Chinese language that has six tones. They introduced semantic anomalies by altering either the tone or phonemes (specifically, vowels) of the terminal word of sentences. The authors found that tonal and phonemic violations modulated the N400 (an ERP component associated with accessing word-level information such as meaning; Kutas & Hillyard, 1984) at a similar latency and amplitude. They concluded that tonal and phonemic information are accessed in parallel in Cantonese and play comparable roles in constraining activation of phonologically similar words. However, sentence processing involves components beyond those involved in single word identification, and so the employed task may have tapped processes beyond ones that are specific to single word recognition.

Most recently, Zhao and colleagues (2011) examined recognition of spoken monosyllables in Mandarin by employing a novel picture/spoken-word/picture task. In this task, Mandarin speakers were presented with a picture of an item in conjunction with its Chinese character before presentation of an auditory word. Following this, subjects were presented with a second picture, and asked whether or not the first and second pictures belonged to the same semantic category. ERPs recorded during passive processing of the auditory word showed that words that differed in tone from the name of the first picture modulated the N400 in the same fashion as words that differed in phonemes (specifically, rimes). Again, the authors reached the conclusion that tonal and phonemic information are accessed over a similar time course and play comparable roles in lexical access. However, Zhao and colleagues (2011) did not employ a task that required subjects to actively make decisions about the critical stimuli they were hearing, and so their task may not have been optimal for detecting potential differences underlying access to tonal versus phonemic information necessary for lexical processing. In addition, neither Zhao et al. (2011) or Schirmer et al. (2005) analyzed any components earlier than the N400, leaving open the question of whether tonal information also influences earlier-going prelexical processes, and whether these might differentiate how and when speakers access segmental versus non-segmental information.

3.1.2 Syllable Processing in Mandarin

In addition to asking whether equivalent processes underlie phonemic versus tonal access, we are also interested in whether the nature of Mandarin syllabic processing is fundamentally different from syllable processing in non-tonal languages such as English. Mandarin syllables are structured differently from English syllables in several ways. Mandarin permits only single onset consonants, whereas English syllables permit consonant clusters in onsets (e.g., /str/ in ‘string’).

Furthermore, Mandarin syllables do not terminate with consonants (with the exception of the nasals /n/ and /ŋ/), while English syllables are again much more permissive in this respect, allowing many single consonants as well as many clusters in this position. As a result of these distinctions, Mandarin has both a smaller set of possible syllables, and a narrower set of possible syllable structures. So, for instance, Mandarin includes a greater proportion of open syllables (i.e., ones with no final consonant) than does English (Wang & Cheng, 2008). This difference in syllable structure variety has led to the hypothesis that the components of the Mandarin syllable (onsets versus rimes) might be weighted differently during Mandarin versus English syllable processing (Wang & Cheng, 2008; Zhao et al., 2011), and that syllables thus play different roles in the word recognition process (Liu, Shu, & Wei, 2006). Furthermore, it has been suggested that whole syllables might merit a special status in Mandarin and be treated in a more holistic fashion than their English counterparts (Zhou & Marslen-Wilson, 1994; Zhao et al., 2011).

To test the relative weighting of parts of the syllable (onsets versus rimes), one can examine competitive effects for words that overlap in onset with target words (often called cohort competitors or mismatches) and words that overlap in rime with expected words (rhyme competitors or mismatches). In English, Allopenna et al. (1998) and Desroches, Joanisse, and Robertson (2006) have used eyetracking to show that when hearing a word such as *beaker* and searching an array for a corresponding picture, listeners looked at pictures depicting cohort competitors (such as *beetle*) more often than pictures of phonologically unrelated items early on during word recognition. In contrast, listeners looked at rhyme competitors (such as *speaker*) later on during the unfolding of spoken words.

Likewise, a more recent study (Desroches, Newman, & Joanisse, 2009), used event related potentials (ERPs) to investigate the temporal dynamics of auditory word recognition in English.

Subjects were shown pictures of common objects, followed by words that either matched or mismatched them. The picture-word matching task was used in a way that generated expectations of incoming auditory stimuli, which were either confirmed or violated by the subsequent auditory input. Mismatches differed in the phonological information they shared with expected words: they either shared word-initial overlap (cohort mismatches; e.g., see *cake*, hear *cage*), word-final overlap (rhyme mismatches; e.g., see *cake*, hear *rake*), or shared no phonemes at all (unrelated mismatches: see *cake*, hear *hose*). For cohort mismatches, the authors observed a greater negativity compared to matching words that was delayed and more prolonged compared to the mismatch effect for unrelated words; conversely, for rhyme mismatches, a greater negativity compared to match was observed much earlier on during the unfolding of spoken words. Together, these findings suggest that mismatches in onsets versus rhymes show temporally distinct effects in English.

In tonal languages however, the story has been less clear. Our previous eyetracking study in Mandarin (Malins & Joanisse, 2010) included a rhyme condition, but failed to show competitive effects of rhyming words. However, using ERPs, Zhao and colleagues (2011) observed N400 effects for rhyming words, but importantly did not find differences in N400 modulation between mismatches with word-final overlap and mismatches with word-initial overlap. So while Malins and Joanisse (2010) concluded that overlapping rimes might not affect word recognition in Mandarin to the same extent that they do in English (or that their study was not powerful enough to detect them), Zhao and colleagues (2011) inferred the opposite: rimes play a much more important role in Mandarin than they do in English – a role that is in fact equivalent to the role of onsets. Furthermore, Zhao et al. (2011) concluded that onset versus rime violations are not temporally distinct from one another like the way they are in English.

Next, to test the relative effects of partial versus whole-syllable mismatches in Mandarin, one can look at within-syllable mismatches (e.g., such as cohorts and rhymes) compared to completely unrelated words. As mentioned previously, in English, Desroches et al. (2009) observed differences between all three types of mismatch (cohort versus rhyme versus unrelated words). However, these effects were related only to the point in the word at which mismatch was signaled. Unrelated mismatches differed in onset consonants from expected words in the same way that rhyme mismatches did, and importantly there was no difference in earlier-going ERP components between these two conditions. This suggests that whole-syllable mismatches did not give rise to effects above and beyond respective violations of syllabic components.

The findings in Chinese have been quite different. Zhao et al. (2011) observed an N400 effect for whole-syllable mismatches that was not only considerably earlier than the N400 effect for within-syllable mismatches (in onset, rime, or tone), but was also of a greater magnitude. This complemented evidence from Schirmer et al. (2005), who observed an earlier N400 effect for whole-syllable mismatches compared to within-syllable violations in either rime or tone. Importantly, this view is somewhat complicated by a similar ERP experiment showing that whole-syllable mismatches were in fact detected *later* than mismatches in only segments or only tone (Brown-Schmidt & Canseco-Gonzalez, 2004). Nevertheless, these types of results have been taken to suggest that whole-syllable violations have an effect on word recognition that goes above and beyond mismatches in individual components, and that syllables might therefore be processed more holistically in Mandarin than they are in English. However, it is important to note that the whole-syllable mismatches in Schirmer et al. (2005) were the only condition in which onset consonants differed from expected words, and therefore this might explain why these words led to earlier N400 effects compared to within-syllable violations. Concerning the

Zhao et al. (2011) study, as we noted above, the task did not require subjects to actively process spoken stimuli, and therefore might have failed to elicit strong modulation based on subtler within-syllable violations as a result. This might explain why the only strong effect they observed was for whole-syllable mismatches, which were detected with little effort, even in the absence of active processing.

3.1.3 Rationale for the Current Study

The current study used ERPs to address the following three issues: (1) whether the processes underlying tonal versus phonemic access differ; (2) whether onsets and rimes play similar roles in Mandarin word recognition as they do in English; (3) whether Mandarin syllables are processed more holistically than their English counterparts. To do this, we adapted the paradigm used by Desroches et al. (2009), which looked at processing of English monosyllables, by instead using Mandarin monosyllables and also including several new conditions in addition to cohort, rhyme, and unrelated mismatches. These included a segmental mismatch condition, in which all phonemes were shared with expected words but the tone was different, and a tonal mismatch condition, in which all phonemes were different from what was expected but the tone was shared. Table 3.1 gives examples of stimuli from each of these five mismatch conditions. Note that conditions are named according to what is shared between expected words and mismatches, as per Desroches et al. (2009), Malins & Joanisse (2010) and Allopenna et al. (1998).

Table 3.1: The Five Mismatch Conditions

Mismatch Condition	*Sample Stimulus Pair (Picture/Word)*	*Shared Information*	*Divergent Information*
Segmental	hua1/hua4	All phonemes	Tone
Cohort	hua1/hui1	Word-initial phonemes and tone	Word-final phonemes
Rhyme	hua1/gua1	Word-final phonemes and tone	Word-initial consonants
Tonal	hua1/jing1	Tone	All phonemes
Unrelated	hua1/lang2	None	All phonemes and tone

Our analyses focused on two components: the phonological mapping negativity (PMN), and the N400. The PMN is thought to be associated with pre-lexical processing, and is modulated when word-initial phonemes differ from those of expected words (Connolly & Phillips, 1994; Newman & Connolly, 2009). The N400 is instead more associated with later-going, word-level processing, such as accessing meaning (Kutas & Hillyard, 1984). Both the PMN and N400 can be thought of as indexing expectancy modulation; rather it is the stage in word recognition at which this happens that distinguishes these two components. For example, we can think of lexical competition being indexed by the N400 and not the PMN because competition amongst words would happen after the pre-lexical stage once a competitor set has been activated.

For each of the mismatch types, the onset of modulation of ERP components can be thought of as indexing when the expectancy modulation was first detected; further, the amplitude and scalp distribution of these effects can offer some insight into underlying cognitive processes. The question of access to tonal versus phonemic information was addressed via the conditions with word-initial overlap with expected words; specifically, the segmental and cohort conditions. In line with Schirmer et al. (2005), Malins and Joanisse (2010), and Zhao et al. (2011) we predicted that there would be no difference in the timing of detection of expectancy violation between the segmental and cohort conditions (and therefore no difference between conditions in terms of the onset of effects for mismatching words); however, we predicted these two conditions would differ in the magnitude and/or scalp distribution of components if in fact there was an underlying difference in processing tonal versus phonemic information. The question regarding the relative roles of onsets and rimes in Mandarin spoken word recognition was addressed by comparing cohort and rhyme mismatches. For the cohort condition, we predicted that, in line with Desroches et al. (2009), we would see a delayed and prolonged N400 effect compared to rhyming words; for the rhyme condition, we predicted we would instead see a larger PMN response based on different word-initial phonemes, but a smaller late N400 effect compared to cohort mismatches. To address the question of whether Mandarin syllables are processed more holistically than their English counterparts, we constrained our analysis to only look at mismatch types that had different word-initial consonants from expected words, yet differed in how much word-final overlap they shared with expected words. By doing this, we were able to compare the effects of mismatches in cases of partial overlap with expected words (rime and tone in the rhyme condition, tone only in the tonal condition) to cases of zero overlap (unrelated condition). We predicted that effects for whole-syllable violations would not lead to any differences in ERP effects above and beyond respective violations of syllabic components in any window of interest.

3.2 Methods

3.2.1 Subjects

Nineteen native speakers of Mandarin (14 female; mean age = 23 years) participated; two additional individuals were tested but removed due to excessive EEG artifacts. All were natives of Mainland China, and had been living in North America for a mean of 4.0 years. Subjects were screened for native speaking ability by having a brief telephone conversation with a research assistant who is a native Mandarin speaker. This assistant also ensured that subjects were fluent in the standard Beijing dialect prior to participation in the experiment.

3.2.2 Stimuli

Stimulus sets were constructed by deriving mismatching items from a set of twelve common object names (henceforth called 'critical targets'; see Appendix). This was done so that all critical items participated equally in all conditions. As it is known that most Mandarin syllables map to a number of homophones, this helped us ensure that any influences of homophony, along with any other attendant lexical factors, were equally distributed across all conditions. For each critical target item, the following mismatches were derived: a segmental mismatch, in which all phonemes were the same as those in the expected word but tone was different (e.g., picture: *hua1* 'flower'; sound: *hua4* 'painting'); a cohort mismatch, in which word-initial phonemes as well as tone were the same but word-final phonemes were different (e.g., picture: *hua1* 'flower'; sound: *hui1* 'grey'); a rhyme mismatch, in which word-final phonemes and tone were the same but word-initial consonants were different (e.g., picture: *hua1* 'flower'; sound: *gua1* 'melon'); a tonal mismatch, in which only tone was the same and all phonemes were different (e.g., picture: *hua1* 'flower'; sound: *jing1* 'whale'); an unrelated mismatch, in which all phonemes as well as tone were different between the expectation generated by the picture and the presented auditory stimulus (e.g., picture: *hua1* ‘flower’; sound: *lang2* ‘wolf’). As much as possible, mismatching items were selected from the sets of other items within the experiment. Across all items used in the experiment, critical targets did not differ significantly in logarithmic frequency (as given in *Modern Chinese Frequency Dictionary*, 1986) from mismatching items [$t(59) = -0.631$, $p = .53$, Cohen’s $d = .12$], nor did the five mismatch conditions differ significantly in logarithmic frequency from each other [$F(4,44) = 0.393$, $p = .73$, $\eta_p^2 = .03$].

All auditory stimuli were recordings of monosyllables spoken in isolation, as produced by a male native speaker of Mandarin. Stimuli were digitized and recorded to computer disk at 16 bits quantization, 44,100 Hz sample rate, and all stimuli were volume normalized to -10 dB. Each item was spoken seven times, and the middle five tokens were selected for experimental use. Duration of the auditory stimuli did not differ significantly between critical targets and mismatching items [$t(59) = 0.853$, $p = .40$, Cohen’s $d = .12$], nor between the five mismatch conditions [$F(4,44) = 0.182$, $p = .91$, $\eta_p^2 = .02$].

Pictures matching each word were selected with the assistance of a native Mandarin speaker to ensure that all items were depicted in a way that was culturally appropriate. It should be noted that while we selected objects as much as possible, two of the forty-three words selected were colour words, presented as squares in each respective colour. Because these items may have introduced semantic category effects, we distributed these items across conditions, so even if semantic category effects were present, these would not account for condition-wise differences. Furthermore, we ensured that any items that were paired with these colour words had pictures that did not contain any pixels in these colours.

To verify that subjects were familiar with the names of the pictures of interest, they were asked to complete a naming task prior to participation in the experiment. In this naming task, we presented each picture used within the ERP experiment, and asked subjects to name the single syllable or character that first came to mind when viewing the picture. This approach allowed us to ensure that monosyllabic names were associated with all of the pictures, rather than disyllabic alternatives (for example, 'willow' could be denoted as *liu3shu4*, but the syllable *shu4* simply indicates that it is a kind of tree, similar to how a 'willow' could be named 'willow tree'; in this case *liu3* is the character uniquely associated with this particular item, and is the label we used). In cases where the given monosyllable differed from the intended name used in the experiment (7.5% of the time across subjects and items, with a standard error of .76%), subjects were given the intended name, in accordance with prior picture naming studies (Brooks & MacWhinney, 2000).

3.2.3 Procedure

All testing was conducted by a native Mandarin speaker, who provided all instructions to subjects in Mandarin. Efforts were made to keep the experimental setting as monolingual as possible to reduce interference from English. In the experimental task, subjects completed a picture-word matching task while continuous EEG activity was recorded from the surface of their scalp. Visual stimuli were presented using a 19-inch CRT monitor; auditory words were presented to the right ear via ER-3A insert headphones (Etymotic Research, Elk Grove Village, IL). In each trial, subjects viewed a fixation cross for 250 ms, after which a picture of an item appeared on screen for 1500 ms. While this picture remained on screen, they heard an auditory word and were required to indicate via button press using their right hand whether the word they heard matched or mismatched the picture. Following the button press, a blank screen was presented for 1000 ms until the onset of the next trial.

Subjects completed four blocks of 120 trials for a total of 480 trials. Of these 480 trials, 240 were match trials, and 240 were mismatch trials; therefore the match to mismatch ratio was 1:1 across the experiment. The 240 mismatch trials were equally distributed across the five mismatch types (segmental, cohort, rhyme, tonal, and unrelated; 48 trials each). Across the experiment, each critical target item (e.g., *hua1* ‘flower’) was presented 20 times as a picture, with the auditory stimulus associated with the picture matching it in half the trials (10 out of the 20), and mismatching it in the other half (2 trials each for the 5 mismatch types). In addition, the 5 mismatching items in each set were also presented as pictures for 20 trials (2 match trials for each, and 2 mismatching trials for each where the critical stimulus item was the auditory stimulus). In this way, every item was presented as a picture, and each time an item was present on the screen, it was equally likely to be a match or mismatch trial. In addition, each time a sound was heard, it was equally likely to be a match or mismatch trial. Furthermore, subjects were assigned to one of two pseudorandom stimulus sequences that controlled for whether items occurred in a match versus mismatch trial the first time each was encountered. A training block of eight trials was provided prior to experimental trials, using items not presented in the actual experiment.

After completing the picture-word matching task, a language history and experience questionnaire was administered, which verified native speaking ability by requesting information regarding daily language use and exposure as well as self-ratings of proficiency.

3.2.4 EEG Data Acquisition

EEG data were collected at a sampling rate of 500 Hz using Acquire 4.2 (Neurosoft Inc., El Paso, TX) and a 32-channel cap with sintered Ag/AgCl electrodes (Quik-Caps; Neurosoft Inc., El Paso, TX) oriented according to the international 10-20 system and referenced to the nose tip. Impedances were kept below 5 kΩ. Data were amplified at a gain of 500 using a SynAmps amplifier and filtered online using 60 Hz notch and 0.1-100 Hz bandpass filters. ERPs were segmented into epochs spanning from 100 ms pre-stimulus to 650 ms post-stimulus onset, time-locked to the onset of the auditory stimulus. Data were then filtered offline using a 12 dB zero phase shift digital bandpass filter (0.1 – 30 Hz), and baseline corrected to the mean voltage of the pre-stimulus interval. Trials containing blinks and other artifacts were removed using a maximum voltage criterion of ±75 μV at any scalp electrode. Following artifact rejection, the average number of accepted trials per condition was 200/240 for the match trials, 38/48 for segmental mismatches, 41/48 for cohort mismatches, 40/48 for rhyme mismatches, 39/48 for tonal mismatches, and 40/48 for the unrelated mismatch condition.

3.2.5 Analysis of ERP Data

ERP analyses focused on four components of interest: the N100, the PMN, the N400, and the late N400 (Desroches et al., 2009). The time intervals of each component were identified by calculating the global field power of all scalp electrodes for the grand average of all trial types, and then identifying the full-width at half maximum of resulting voltage peaks (Lehmann & Skrandies, 1980). This yielded the following time windows: 74 to 126 ms (N100), 250 to 310 ms (PMN), 310 to 406 ms (N400), and 406 to 560 ms (late N400). The mean voltage within each time window was calculated for each of the six word type conditions (match, segmental mismatch, cohort mismatch, rhyme mismatch, tonal mismatch, unrelated mismatch). Statistical analyses were conducted for fifteen electrode sites (Fz, F3, F4, F7, F8, Cz, C3, C4, T7, T8, Pz, P3, P4, P7, P8), which provided sufficient coverage across the scalp to differentiate the components of interest (Newman et al., 2003; Desroches et al., 2009). A separate repeated-measures ANOVA using conservative degrees of freedom (Greenhouse & Geisser, 1959) was performed for each window to test for potential electrode site by word type interactions. Significant interactions were followed up on using post-hoc two-tailed *t*-tests at Fz, Cz, and Pz to identify which mismatch conditions were significantly different from the match condition.

Next, for each specific hypothesis, we conducted further analyses by taking the mean voltage within either the PMN, N400, or late N400 windows for six groups of electrodes distributed across the head. These electrodes were grouped as follows: anterior-left (F7 & F3), anterior-right (F8 & F4), central-left (T7 & C3), central-right (T8 & C4), posterior-left (P7 & P3), and posterior-right (P4 & P8). For each test, we took the values from the word type conditions relevant to each specific hypothesis and subjected them to a repeated-measures ANOVA. First, to more directly compare the timing of detection of expectancy violation on the basis of divergent tones (the segmental condition) versus divergent phonemes (the cohort condition), the segmental, cohort, and unrelated mismatch conditions were analyzed within the PMN window, the earliest point at which differences between conditions were anticipated. The unrelated mismatch condition was the most straightforward test for a PMN, as the auditory word was completely unrelated to what was expected. Thus this condition served as the baseline against which other types of phonological mismatch were compared.

Next, to address the hypothesis concerning the relative weighting of onsets versus rimes, we specifically analyzed mismatch effects for the cohort, rhyme, and unrelated conditions in the PMN and late N400 windows. Recall that in English (Desroches et al., 2009), temporally distinct patterns of modulation were observed between these conditions in these windows. Again, the unrelated condition served as the baseline against which other types of phonological mismatch could be compared.

Finally, to examine effects of partial versus whole-syllable mismatches, we tested the effect of zero overlap with expected words (unrelated condition) compared to partial overlap (rime and tone in the rhyme condition, tone only in the tonal condition). In this case we examined all three windows, to determine whether they were differentiated at any of the time points of interest.

3.3 Results

3.3.1 Behavioral Data

Mean accuracy and mean decision latencies for the button press response are presented in Table 3.2. Trials for which decision latencies exceeded +/- 2.5 SD of condition-wise means for each subject were rejected, as were trials with decision latencies less than 100 ms. In addition, only correct trials were included in the decision latency analysis. Repeated-measures ANOVAs showed a main effect of word type on decision latency [$F(5,90) = 6.91$, $p < .001$, $\eta_p^2 = .28$], as well as accuracy [$F(5,90) = 2.84$, $p < .05$, $\eta_p^2 = .14$]. Post-hoc two-tailed *t*-tests for each mismatch condition contrasted separately against the match condition revealed that the segmental, cohort, and rhyme mismatch conditions had significantly longer decision latencies than the match condition [segmental: $t(18) = -3.23$, $p < .01$, Cohen's $d = .41$; cohort: $t(18) = -3.66$, $p < .01$, $d = .44$; rhyme: $t(18) = -2.87$, $p < .01$, $d = .31$], while the tonal and unrelated mismatch conditions were not significantly different from the match condition in decision latency [tonal: $t(18) = -1.26$, $p = .22$, $d = .17$; unrelated: $t(18) = 0.09$, $p = .93$, $d = .01$]. This difference in decision latencies across conditions likely reflected the greater degree of

Table 3.2: Mean Decision Latency and Mean Percent Accuracy for the Matching Judgment Relative to Word Onset

Condition	*Decision Latency (ms)*	*Percent Accuracy*
Match	773.2 (27.3)	97.9 (0.41)
Segmental mismatch	826.0 (32.3)	98.4 (0.60)
Cohort mismatch	829.3 (30.6)	98.1 (0.61)
Rhyme mismatch	813.8 (32.9)	99.2 (0.34)
Tonal mismatch	796.4 (34.5)	98.9 (0.50)
Unrelated mismatch	771.4 (29.5)	99.4 (0.46)

Note: Values in parentheses represent standard error.

phonological overlap with expected words in the segmental, cohort, and rhyme conditions compared to tonal and unrelated words, with subjects requiring longer processing times to recognize words as mismatches in conditions with a high degree of phonological overlap.

In terms of accuracy, the match condition was not significantly different from the cohort and segmental conditions [segmental: $t(18) = -0.85$, $p = .41$, $d = .21$; cohort: $t(18) = -0.29$, $p = .78$, $d = .06$], but accuracy was significantly lower in the match condition than it was in the rhyme, tonal, and unrelated mismatch conditions [rhyme: $t(18) = -2.61$, $p < .05$, $d = .79$; tonal: $t(18) = -2.32$, $p < .05$, $d = .47$; unrelated: $t(18) = -2.82$, $p < .05$, $d = .80$]. This effect of condition on accuracy is likely a result of the difference between ‘yes’ and ‘no’ responses. Because subjects were in a task set where they were encountering many different types of mismatches, they might have been more inclined to say ‘no’ than they might otherwise have been, and thus the accuracy of their ‘yes’ responses was lower than their ‘no’ responses. The only exceptions to this were the segmental and cohort conditions, which likely experienced lower accuracy due to word-initial overlap with expected words.

3.3.2 ERP Data

Mismatch conditions are plotted separately against the match condition in Figure 3.1. In Figure 3.2, the subtraction maps for each mismatch condition minus the match condition are shown for each of the PMN, N400, and late N400 windows.

For the N100, there was no significant interaction between electrode site and word type [$F(70,1260) = 1.26$, $p = .27$, $\eta_p^2 = .07$], nor a main effect of word type [$F(5,90) = 1.88$, $p = .13$, $\eta_p^2 = .10$]. However, there was a main effect of electrode site [$F(14,252) = 53.54$, $p < .001$, $\eta_p^2 = .75$], which was anticipated as the N100 for auditory stimuli is typically strongest over central sites.

For the PMN, there was a significant interaction between electrode site and word type [$F(70,1260) = 2.40$, $p < .05$, $\eta_p^2 = .12$], as well as a main effect of word type [$F(5,90) = 3.20$, $p < .05$, $\eta_p^2 = .15$]. Post-hoc analyses for the electrodes Fz, Cz, and Pz (Table 3.3) showed that the rhyme, tonal, and unrelated conditions all gave rise to a significantly greater negativity from match in the PMN window at Fz, Cz, and Pz, which is consistent with the fact that all three of these conditions had word-initial phonemes that differed from what was expected. The segmental and cohort conditions did not show significant differences from match in Fz, Cz, or Pz, consistent with the fact that both these conditions had word-initial phonemes matching the expected word.

Analysis of the N400 identified a significant interaction between electrode site and word type [$F(70,1260) = 3.30$, $p < .01$, $\eta_p^2 = .16$], but not a main effect of word type [$F(5,90) = 1.76$, $p = .16$, $\eta_p^2 = .09$]. Post-hocs (Table 3.3) revealed that all five of the mismatch conditions showed significantly greater negativities than match at Cz and at least one of either Fz or Pz. This was

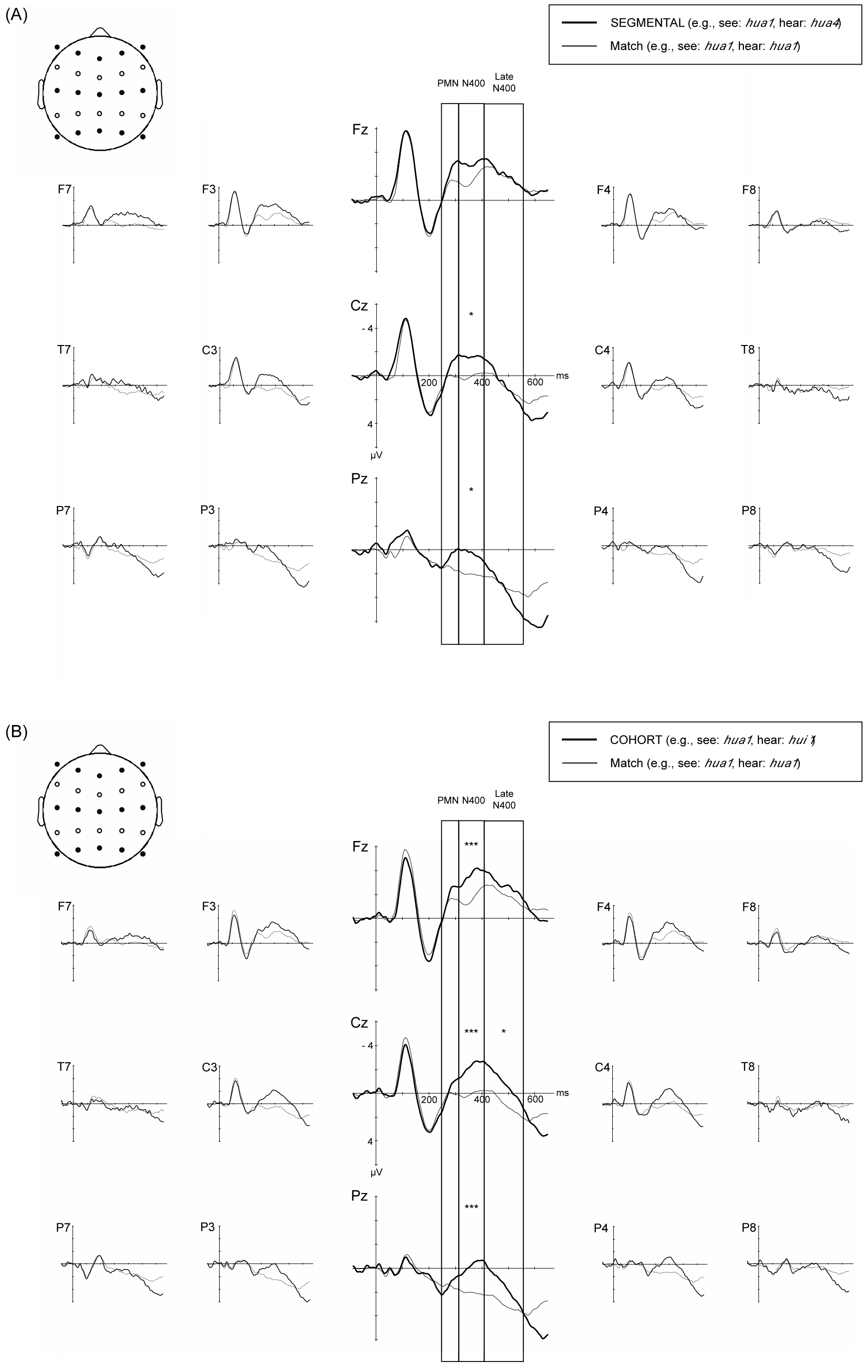


Figure 3.1A and 3.1B: ERP data for the segmental (A) and cohort (B) mismatch conditions versus the match condition showing results of post-hoc analyses for the electrodes Fz, Cz, and Pz. Results indicate an N400 effect in (A) and N400 and late N400 effects in (B).

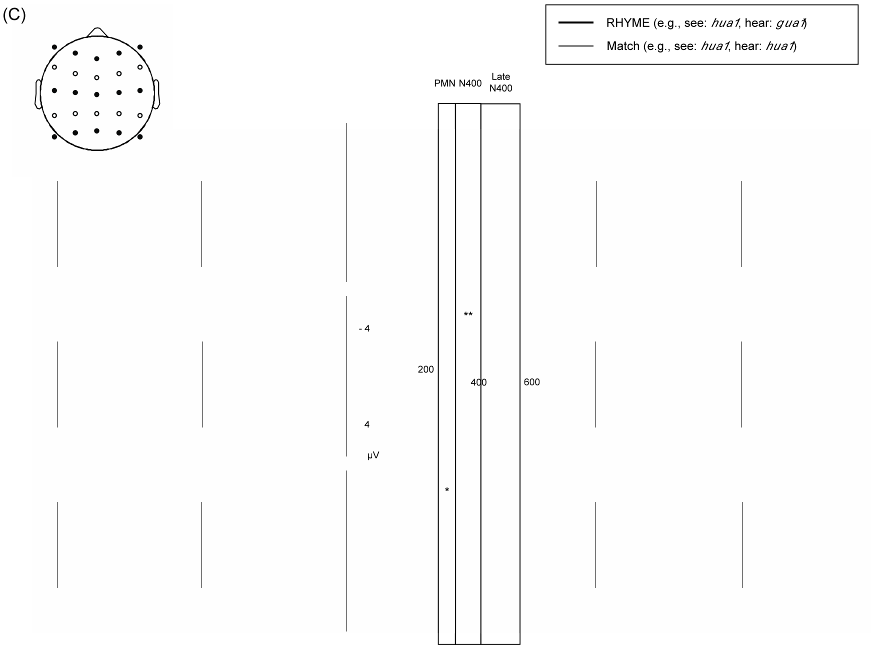


Figure 3.1C and 3.1D: ERP data for the rhyme (C) and tonal (D) mismatch conditions versus the match condition, showing results of post-hoc analyses for the electrodes Fz, Cz, and Pz. Results indicate PMN and N400 effects in both conditions.

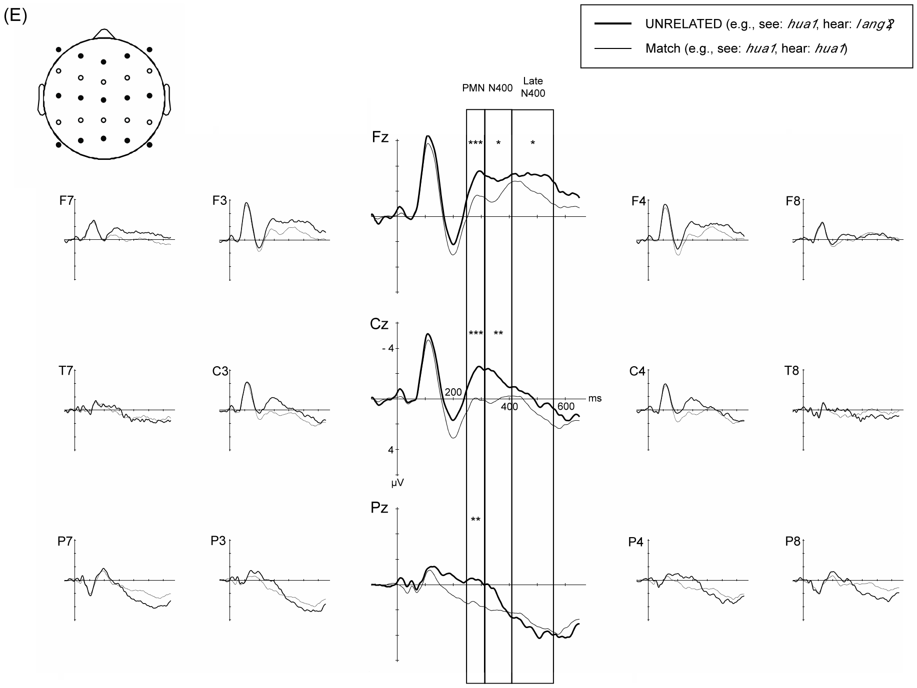


Figure 3.1E: ERP data for the unrelated mismatch condition versus the match condition, showing results of post-hoc analyses for the electrodes Fz, Cz, and Pz. Results indicate effects in all three of the PMN, N400, and late N400 windows.

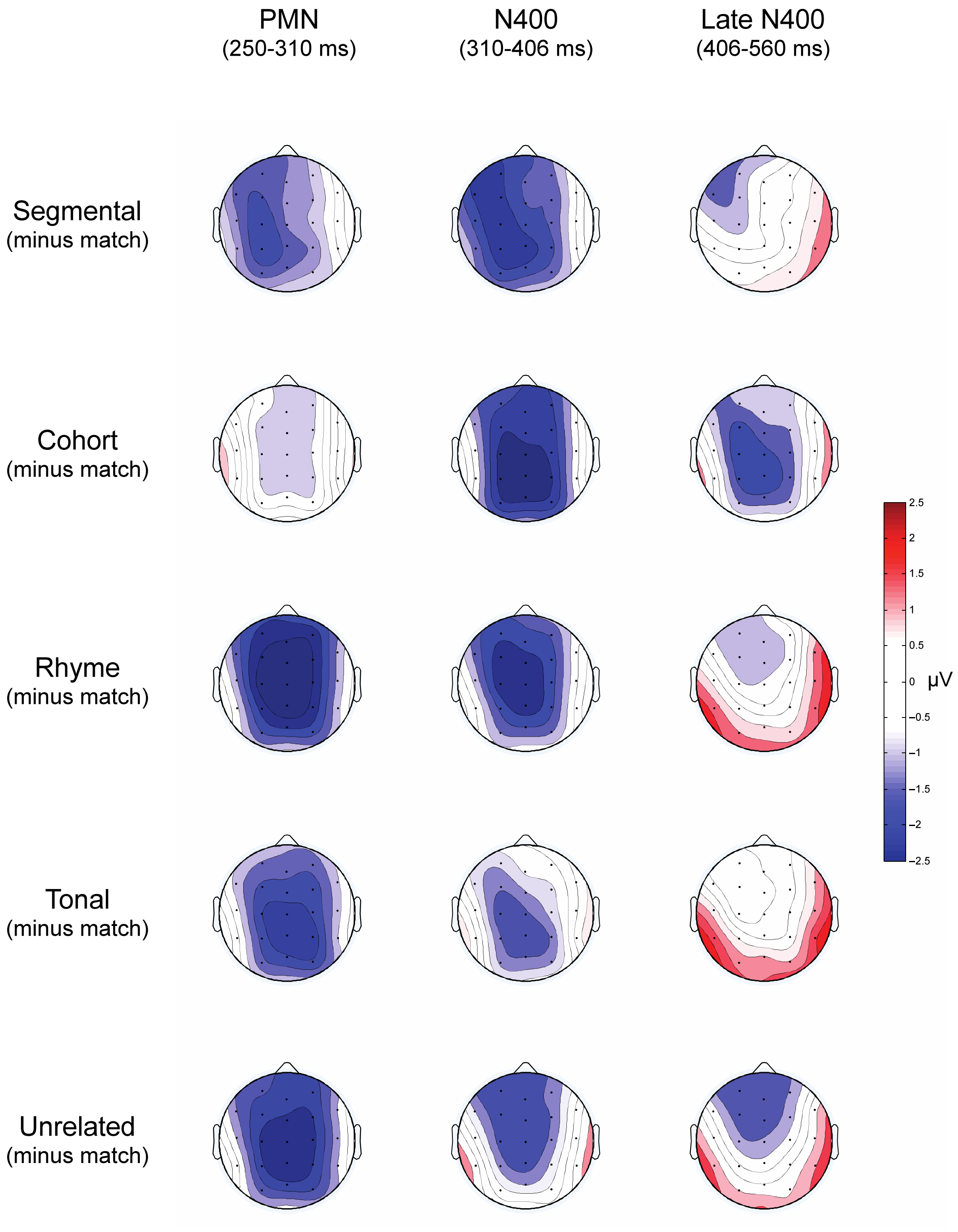


Figure 3.2: Scalp maps for the subtractions of each mismatch condition minus match for the PMN, N400, and late N400 windows.

Table 3.3: Comparison of Each Mismatch Condition Against Match in the PMN, N400, and Late N400 Windows

Component/Contrast	*Electrode*	*t*	*Effect Size (Cohen's d)*
PMN			
Segmental vs. Match	Fz	1.63	.34
	Cz	1.65	.33
	Pz	1.80†	.38
Cohort vs. Match	Fz	1.59	.32
	Cz	1.17	.27
	Pz	0.78	.15
Rhyme vs. Match	Fz	4.05***	.90
	Cz	3.90**	.93
	Pz	2.65*	.61
Tonal vs. Match	Fz	2.83*	.45
	Cz	3.83**	.68
	Pz	3.38**	.56
Unrelated vs. Match	Fz	4.14***	.66
	Cz	4.90***	.89
	Pz	3.48**	.64
N400			
Segmental vs. Match	Fz	1.96†	.34
	Cz	2.27*	.37
	Pz	2.18*	.39
Cohort vs. Match	Fz	4.72***	.59
	Cz	4.30***	.65
	Pz	4.11***	.62
Rhyme vs. Match	Fz	3.00**	.54
	Cz	3.33**	.63
	Pz	1.94†	.36
Tonal vs. Match	Fz	0.82	.12
	Cz	2.46*	.35
	Pz	2.15*	.33
Unrelated vs. Match	Fz	2.36*	.36
	Cz	3.60**	.46
	Pz	1.57	.21
Late N400			
Segmental vs. Match	Fz	0.13	.02
	Cz	0.001	< .01
	Pz	-0.62	.10
Cohort vs. Match	Fz	1.26	.21
	Cz	2.37*	.39
	Pz	1.81†	.32
Rhyme vs. Match	Fz	1.36	.24
	Cz	0.83	.12
	Pz	-1.49	.21
Tonal vs. Match	Fz	-0.01	< .01
	Cz	-0.03	< .01
	Pz	-2.01†	.24
Unrelated vs. Match	Fz	2.14*	.35
	Cz	1.56	.20
	Pz	-0.65	.10

* $p < .05$(two-tailed) ** $p < .01$ *** $p < .001$ † $p < .10$

anticipated as in all five conditions the auditory word mismatched from what was expected at the whole-word level.

Finally, analysis of the late N400 yielded a significant interaction between electrode site and word type [$F(70,1260) = 4.05$, $p < .001$, $\eta_p^2 = .18$], but not a main effect of word type [$F(5,90) = 0.94$, $p = .44$, $\eta_p^2 = .05$]. Post-hocs (Table 3.3) revealed that the cohort condition showed a greater negativity than match in this window at Cz, and that the unrelated condition showed a greater negativity than match at Fz, suggesting that the effect of expectancy violation was more persistent over midline electrodes in the cohort and unrelated mismatch conditions than it was in the other three mismatch conditions.

3.3.2.1 Hypothesis 1: Different cognitive processes underlie tonal versus phonemic access

To address whether the timing and distribution of pre-lexical access differs for tonal versus phonemic information, we looked more closely at the point of detection of expectancy violation in the segmental and cohort conditions compared to the unrelated mismatch condition (Figure 3.3A). A repeated-measures ANOVA revealed a significant electrode group by word type interaction in the PMN window [$F(10,180) = 2.47$, $p < .05$, $\eta_p^2 = .12$], with post-hoc paired sample *t*-tests showing a significantly greater negativity in the segmental condition than in the cohort condition over left central sites [$t(18) = -2.35$, $p < .05$, Cohen's $d = .41$]. Thus it appears that expectancy violations were detected earlier in the segmental condition than in the cohort condition, as the segmental condition showed a greater effect of mismatch in the PMN window than the cohort condition, which did not show a significant mismatch effect until the N400 window. In addition, this effect was strongest in left central sites.

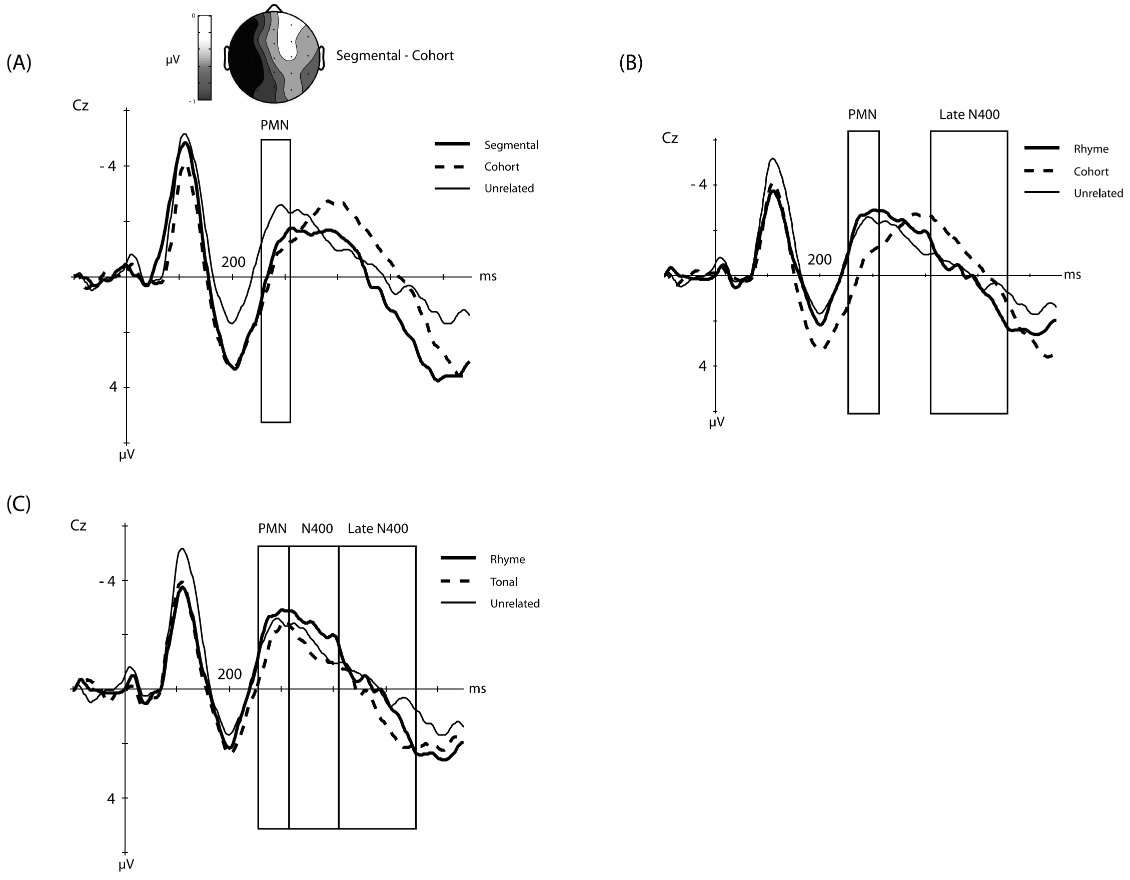


Figure 3.3: Direct comparisons of ERP effects for mismatch conditions relevant to our three hypotheses. (A) For hypothesis one, we directly compared the segmental, cohort, and unrelated mismatch conditions in the PMN window; the scalp map shows the subtraction of the segmental minus the cohort condition in the PMN window. (B) For hypothesis two, we directly compared the cohort, rhyme, and unrelated mismatch conditions in the PMN and late N400 windows. (C) For hypothesis three, we directly compared the rhyme, tonal, and unrelated mismatch conditions in all three of the PMN, N400, and late N400 windows.

3.3.2.2 Hypothesis 2: The relative weighting of onsets versus rimes is different in Mandarin compared to English

To investigate the relative weighting of onsets versus rimes in Mandarin, we looked at whether the ERP responses to mismatching words that overlapped in onset (cohort condition) differed from those of words overlapping in rime (rhyme condition), versus completely unrelated words (Figure 3.3B). Specifically, we looked at the earlier-going PMN and the late N400 windows, guided by what has previously been observed in English (Desroches et al., 2009). In the PMN window, we did not find a significant electrode group by word type interaction [$F(10,180) = 0.69$, $p = .55$, $\eta_p^2 = .04$]; however, there was a main effect of word type [$F(2,36) = 4.81$, $p < .05$, $\eta_p^2 = .21$]. Follow-up paired sample *t*-tests, collapsed across electrode groups, showed that the rhyme and unrelated conditions had more negative amplitudes in the PMN window than the cohort condition [rhyme vs. cohort: $t(18) = -3.74$, $p < .01$, Cohen's $d = .46$; unrelated vs. cohort: $t(18) = -2.25$, $p < .05$, Cohen's $d = .4133$]. In the late N400 window, we found a significant electrode group by word type interaction [$F(10,180) = 3.52$, $p < .01$, $\eta_p^2 = .16$], with follow-up analyses showing that the cohort condition had a more negative amplitude in the late N400 window compared to the rhyme condition in left central [$t(18) = -2.59$, $p < .05$, Cohen's $d = .33$], left posterior [$t(18) = -3.26$, $p < .01$, Cohen's $d = .44$], and right posterior sites [$t(18) = -2.39$, $p < .05$, Cohen's $d = .35$]. From these results, it is evident that the cohort and rhyme conditions showed temporally distinct patterns of effects in the PMN and late N400 windows.

3.3.2.3 Hypothesis 3: Whole-syllable mismatches give rise to mismatch effects above and beyond those of mismatches in individual components

To compare partial versus whole syllable mismatch effects, we examined the rhyme, tonal, and unrelated conditions (Figure 3.3C), on the logic that in each case the target word differed from expectation with respect to the onset consonants, but also that the forms differed from each other with respect to how many *additional* components also differed from expectations (none in the rhyme condition, rime in the tonal condition, and rime and tone in the unrelated condition). In all three of the PMN, N400, and late N400 windows, we did not find a significant electrode group by word type interaction [PMN: $F(10,180) = 0.79$, $p = .50$, $\eta_p^2 = .04$; N400: $F(10,180) = 1.86$, $p = .13$, $\eta_p^2 = .09$; Late N400: $F(10,180) = 0.75$, $p = .56$, $\eta_p^2 = .04$], nor a main effect of word type [PMN: $F(2,36) = 0.58$, $p = .55$, $\eta_p^2 = .03$; N400: $F(2,36) = 0.59$, $p = .55$, $\eta_p^2 = .03$; Late N400: $F(2,36) = 0.45$, $p = .63$, $\eta_p^2 = .02$]. Thus it seems that word-initial mismatch in the PMN window was treated in the same way across conditions regardless of whether there was later overlap with expectations; furthermore, we failed to see a gradation later on in the N400 and late N400 windows based on the degree of overlap with expectations (rime and tone in the rhyme condition, tone only in the tone condition, and no overlap in the unrelated condition). This suggests that whole-syllable mismatches did not give rise to mismatch effects above and beyond those of mismatches in individual components, as these three conditions could not be differentiated in any of the windows of interest.

3.4 Discussion

We used ERPs to study (1) whether different neurocognitive processes underlie tonal versus phonemic access during auditory word recognition; (2) whether onsets and rimes play similar roles in Mandarin as they do in English, and (3) whether Mandarin syllables are processed in a generally holistic manner. In our task paradigm, picture cues generated phonological expectations, which were then subsequently confirmed or violated by a spoken word. Mismatching words overlapped phonologically with expected words in varying ways, allowing us to address each hypothesis.

3.4.1.1 Hypothesis 1: Different cognitive processes underlie tonal versus phonemic access

This hypothesis was addressed using the segmental and cohort conditions. Mismatching words in both conditions shared word-initial overlap with targets up to a point of divergence. However the nature of the information signaling the divergence was different: in the segmental condition, auditory words differed from expectations in tone (e.g., picture: *hua1* ‘flower’; sound: *hua4* ‘painting’), while in the cohort condition, auditory words differed from expectations in word-final phonemes (e.g., picture: *hua1* ‘flower’; sound: *hui1* ‘grey’). In both cases, the divergence was signaled within the rime of the syllable, and this close approximation in timing allowed us to look more closely at the issue of timing of access to tonal versus phonemic information in Mandarin. As is evident in Figure 3.1A /3.1B, we observed that the segmental and cohort conditions showed similar magnitude PMNs at midline sites; however, as Figure 3.2 shows, we did see a difference in terms of the broader scalp distribution of this effect, which was confirmed by the follow-up analysis at different scalp sites that showed a larger amplitude PMN in the segmental condition than the cohort condition over left central sites (also shown in the direct subtraction between the segmental and cohort conditions in Figure 3.3A). This suggests that tonal information is processed very early in auditory word recognition, and that listeners use tonal information as soon as it becomes available, which is consistent with prior work by Schirmer et al. (2005), Malins and Joanisse (2010), and Zhao et al. (2011). It is possible that the PMN modulation in the segmental condition that was not observed in the cohort condition is due to the fact that listeners were potentially able to detect tonal information carried by voiced onset consonants (see Howie, 1974, for a review). This possibility seems unlikely given that only two out of the twelve stimulus sets used in this experiment began with voiced consonants (/l/ and /m/). Nevertheless, such an explanation does not undermine our central claim; rather it is fully consistent with our argument that listeners do access tonal information very early in word recognition and use it to constrain word recognition as soon as it becomes available (just like phonemic information). Furthermore, it would only serve to strengthen our assertion that tone is not a weaker cue to word recognition like Cutler and Chen (1997) and others suggest.

In addition to addressing this question of timing, the cortical distribution of the PMN component in the segmental mismatch condition suggests that tonal processing might also be more left lateralized than phonemic processing, as a more centralized scalp distribution was observed in the cohort condition when the divergence was detected (around the N400 window) than was observed in the segmental condition. Likewise, later-going effects in the segmental and cohort conditions, especially during the late N400, suggest further processing differences between tonal and phonemic processing: the cohort condition was significantly different from match at Cz in the late N400 window while the segmental condition was not. This suggests that the effect of word-initial overlap was more persistent in the cohort condition than it was in the segmental condition, and that competition among lexical items was greater in the cohort condition than in the segmental condition (see Desroches et al., 2009). Taken together, these results suggest at least partially separate mechanisms underlie processing of tonal and phonemic information in Mandarin, given that competitive effects were different across the two conditions. For instance, the cohort N400 effect may have been more prolonged than the segmental N400 effect because mismatching tones in the segmental condition constrained the competitor pool to a smaller size than mismatching phonemes did in the cohort condition.

The observed differences in cortical distribution are supported by Luo et al. (2006), who found differences in the lateralization of MMNs (an ERP index of perceptual categorization) to tonal versus phonemic contrasts. This is also supported by Gandour et al. (2003), who examined fMRI activation when subjects actively judged onset consonants, rimes, or tones of Mandarin syllables. Their data indicated greater activation in the left posterior middle frontal gyrus for rhymes than for onsets and tones, while greater activation was seen bilaterally in the posterior inferior frontal gyrus for tones compared to onsets and rhymes.

3.4.1.2 Hypothesis 2: The relative weighting of onsets versus rimes is different in Mandarin compared to English

This hypothesis was addressed by examining the rhyme and cohort conditions, in which the auditory word differed from expectation either in onset (rhyme condition; picture: *hua1* ‘flower’; sound: *gua1* ‘melon’), or rime (cohort condition; picture: *hua1* ‘flower’; sound: *hui1* ‘grey’). As is presented in Figures 3.1B and 3.1C, these two conditions showed temporally distinct patterns of effects; the rhyme condition was significantly different from match over midline sites in the PMN window and N400 window but not the late N400 window, while the cohort condition was significantly different from match in the N400 and late N400 windows, but not the PMN window.

These differences between conditions were confirmed by a follow-up analysis at different scalp sites looking at the cohort, rhyme, and unrelated mismatch conditions (Figure 3.3B), which showed that PMN amplitude was overall more negative in the rhyme and unrelated conditions compared to the cohort condition. This difference in timing was expected, as violations in onset in the rhyme and unrelated conditions were signaled earlier than violations in rime in the cohort condition. Interestingly though, the cohort and rhyme conditions also differed in the late N400 window, as the cohort condition showed a more negative late N400 compared to the rhyme condition over left central and left and right posterior sites. This large late N400 in the cohort condition has been observed previously in Desroches et al. (2009), marked by a more negative late N400 for cohort mismatches compared to completely unrelated mismatches. This was attributed to initial confirmatory evidence in the cohort condition leading to increased activation of the expected word. The late N400 was thought to reflect a later-going increase in processing required to overcome this initial confirmatory evidence, which extended beyond what was required for completely unrelated words. We take the same position here, as similar results were observed in this study, and on the basis of this evidence we would propose that onsets and rimes are weighted similarly between English and Mandarin, in contrast to what some authors have proposed (Zhao et al., 2011).

3.4.1.3 Hypothesis 3: Whole-syllable mismatches give rise to mismatch effects above and beyond those of mismatches in individual components

This hypothesis was addressed by comparing effects in the rhyme, tonal, and unrelated conditions. All three of these conditions differed from expected words in onsets, but the rhyme and tonal conditions shared some partial overlap with expected words – rime and tone in the rhyme condition (e.g., picture: *hua1* ‘flower’; sound: *gua1* ‘melon’) or only tone in the tonal condition (e.g., picture: *hua1* ‘flower’; sound: *jing1* ‘whale’) – while words in the unrelated condition differed in all components from expected words. As illustrated in Figure 3.1C/3.1D/3.1E, all three conditions showed a greater negativity compared to match in the PMN window, which was expected given that onsets differed from anticipated words in these cases. Importantly though, when these conditions were compared directly to each other (Figure 3.3C), it was found that the conditions could not be differentiated from one another in any of the three windows that were analyzed. This suggests that whole-syllable mismatches do not show effects above and beyond mismatches in individual components, supporting the idea that Mandarin syllables are processed in an incremental fashion similar to what is observed in English, rather than in a purely holistic fashion as some have posited (Zhou & Marslen-Wilson, 1994; Zhao et al., 2011).

3.4.2 Implications for Models of Spoken Word Recognition

The findings from the current study have a number of implications for models of spoken word recognition. Namely, these results can help adjudicate amongst competing models such as Cohort (Marslen-Wilson, 1987), TRACE (McClelland & Elman, 1986), and the neighborhood activation model (NAM; Luce & Pisoni, 1998) in terms of which is a most appropriate starting point for modeling spoken word recognition in Mandarin. First, our data seem to suggest that the neighborhood activation model is not viable for Mandarin. This is because NAM weights the degree of phonological competition among words equally, regardless of where these differences occur (word-initial, word-medial, or word-final), while we found distinct patterns of modulation for violations in onset, rime, and tone (the rhyme, cohort, and segmental conditions respectively). It is important to note that this conclusion is different from the one reached by Zhao et al. (2011), as these authors observed equivalent ERP mismatch effects for diverging onsets, rimes, and tones. The likely reason for this difference between studies is that our task required subjects to actively process the target words and subsequently make decisions about them, which attuned them to subtle differences between the auditory words and the picture stimuli.

Our data are more consistent with online processing models, either feedforward in nature such as Cohort, or continuous mapping (interactive) such as TRACE. That being said, the data from this study suggest that to fully accommodate tonal languages, a working model of spoken word recognition requires tonal feature detectors (as suggested by Zhou, Qu, Shu, Gaskell, & Marslen-Wilson, 2004), instantiated at the same level as phonetic feature detectors. If these are activated over a similar time course as phonemic feature detectors, this change could account for the similar timing of the segmental and cohort effects observed in this study. In addition, these should be instantiated such that they are distinct processing structures from feature level and sublexical phoneme units (i.e., there do not need to be reciprocal connections between tonal processors and phonemic processors), to reflect the potential mechanistic difference of tonal versus phonemic processing observed in this study.

Last, the current data suggest that a model of Mandarin spoken word recognition does not merit a special layer for syllables as some have posited (Zhou & Marslen-Wilson, 1994; Zhao et al., 2011), as we did not observe effects for whole-syllable mismatches above and beyond effects of individual components of syllables. This again suggests that the three-layer architecture posited by the TRACE model for languages like English (which does not have a separate layer for syllables) might be sufficient for Mandarin as well.

It is noteworthy that the decision latency data yielded a different pattern of results than what was observed in the ERP data. Even though there were different patterns of effects in the segmental, cohort, and rhyme conditions in the ERP data, decision latencies for these three conditions were similar, which is consistent with NAM. However, we take this result cautiously because there were similar ERP effects for the rhyme and tonal conditions, yet in this case the decision latencies were considerably different for these two conditions. This finding that ERP differences are not clearly echoed in subsequent decision latencies is not altogether unexpected given prior studies showing similar effects (Brown & Hagoort, 1993). The explanation is that decision latencies fall well outside the analysis range of the ERP data, and therefore capture later-going effects in addition to the earlier-going ones that are of particular interest with respect to the hypotheses we are addressing. Thus, decision latencies represent the culmination of a number of different processes, any one of which could have differed between conditions, versus the ERP analysis, which looked at differences in component processes early on while ignoring later differences leading up to the response.

3.5 Conclusions

We used ERP to examine effects of phonological similarity on Mandarin spoken word recognition. Our goals were to establish whether the mechanisms underlying tonal access were congruent with those governing phonemic access, whether onsets and rhymes are weighted similarly in Mandarin and English, and whether Mandarin syllables are processed more holistically than their English counterparts. Results support the view that tonal and phonemic information are both accessed and used as soon as they become available during the unfolding of a spoken word. Nonetheless, the somewhat different scalp distributions of segmental versus cohort mismatch effects suggest that the mechanisms underlying access to tonal versus phonemic information differ to some extent, an intriguing possibility that warrants further study. In addition, we found evidence that onsets and rimes are weighted similarly in Mandarin as they are in English. Furthermore, we did not observe effects of whole-syllable mismatches above and beyond mismatches in individual components, suggesting that Mandarin syllables are processed incrementally just like they are in English. Overall, these findings offer support from a tonal language for online processing models of spoken word recognition, and suggest several refinements necessary for these models to fully account for spoken word recognition across the world’s languages.

3.6 References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38* (4), 419-439.

Brooks, P. J., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language*, *27*(2), 335-366.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5* (1), 34-44.

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research*, *33*(2), 103-135.

Connolly, J. F., & Phillips, N. A. (1994). Event related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6* (3), 256-266.

Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59* (2), 165-179.

Cutler, A. & Otake, T. (1999). Pitch accent in spoken word recognition in Japanese. *Journal of the Acoustical Society of America*, *105*(3), 1877-1888.

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, *100*, B32-B42.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21* (10), 1893-1906.

Gandour, J., et al. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*, 185-200.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, 30 (3), 129-148.

Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, *13*, 646-653.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography & Clinical Neurophysiology, 48* (6), 609-621.

Li, P., Tan, L. H., Bates, E., & O. J. L. Tzeng (Eds.). (2006). *The Handbook of East Asian Psycholinguistics Volume 1: Chinese*. New York, USA: Cambridge University Press.

Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context: Evidence from Chinese ERP analyses. *Brain and Language*, *96*, 37-48.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1-36.

Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, *103* (51), 19558-19563.

Malins, J. G. & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62* (4), 407-420.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition Special Issue: Spoken Word Recognition, 25* (1-2), 71-102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18* (1), 1-86.

Modern Chinese Frequency Dictionary. (1986). Beijing: Beijing Language Institute Publisher (in Chinese).

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology*, *80*, 114-121.

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potential. *Psychophysiology*, *40*, 640–647.

Newman, R. L., Forbes, K., & Connolly, J.F. (2012). Event-related potentials and magnetic fields associated with spoken word recognition. In M. Spivey, M. Joanisse & K. McRae (Eds.), *Cambridge Handbook of Psycholinguistics*. Cambridge, England: Cambridge University Press.

Repp, B. H., & Lin, H. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18* (4), 481-495.

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

Taft, M., & Chen, H. (1992). Judging homophony in Chinese: The influence of tones. In H. Chen, & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 151-172). Oxford, England: North-Holland.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *58A*(2), 251-273.

Wang, M., & Cheng, C. (2008). Subsyllabic unit preference in young Chinese children. *Applied Psycholinguistics*, *29*, 291-314.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and* Cognitive *Processes Special Issue: Processing East Asian Languages, 14* (5-6), 609-630.

Zhao, J., Guo, J. Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49* (7), 1761-1770.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes, and syllables in the Chinese mental lexicon. *Language and Cognitive Processes*, *9*(3), 393-422.

Zhou, X., Qu, Y., Shu, H., Gaskell, G., & Marslen-Wilson. (2004). Constraints of lexical tone on semantic activation in Chinese spoken word recognition. *Acta Psychologica Sinica*, *36* (4), 379-392.

3.7 Appendix – Experimental Stimuli

Set	*Critical Target*	*Segmental Mismatch*	*Cohort Mismatch*	*Rhyme Mismatch*	*Tonal Mismatch*	*Unrelated Mismatch*
1	bing3 (pastry) /piŋ3/ [93]	bing1 (soldier) /piŋ1/ [700]	biao3 (watch) /piɑu3/ [1705]	ling3 (collar) /liŋ3/ [1605]	gou3 (dog) /kəu3/ [245]	hua4 (painting) /xua4/ [602]
2	chuang2 (bed) /tʂʰuɑŋ2/ [454]	chuang1 (window) /tʂʰuɑŋ1/ [401]	chuan2 (ship) /tʂʰuan2/ [1016]	huang2 (yellow) /xuɑŋ2/ [508]	hou2 (monkey) /xəu2/ [105]	liu3 (willow) /liəu3/ [68]
3	deng4 (stool) /təŋ4/ [92]	deng1 (lamp) /təŋ1/ [585]	dou4 (beans) /təu4/ [182]	feng4 (phoenix) /fəŋ4/ [25]	hua4 (painting) /xua4/ [602]	biao3 (watch) /piɑu3/ [1705]
4	feng4 (phoenix) /fəŋ4/ [25]	feng1 (wind) /fəŋ1/ [1474]	fei4 (lungs) /fei4/ [75]	deng4 (stool) /təŋ4/ [92]	gui4 (wardrobe) /kuei4/ [120]	tao2 (peach) /tʰɑu2/ [100]
5	gou1 (hook) /kəu1/ [56]	gou3 (dog) /kəu3/ [245]	gen1 (roots) /kən1/ [1304]	dou1 (pocket) /təu1/ [51]	jin1 (gold) /tɕin1/ [695]	xin4 (envelope) /ɕin4/ [1202]
6	gui4 (wardrobe) /kuei4/ [120]	gui3 (ghost) /kuei3/ [276]	guan4 (jar) /kuan4/ [63]	hui4 (meeting) /xuei4/ [7000]	miao4 (temple) /miɑu4/ [60]	jiao1 (glue) /tɕiɑu1/ [75]
7	hua1 (flower) /xua1/ [1574]	hua4 (painting) /xua4/ [602]	hui1 (gray) /xuei1/ [284]	gua1 (melon) /kua1/ [174]	jing1 (whale) /tɕiŋ1/ [82]	lang2 (wolf) /lɑŋ2/ [105]
8	jing1 (whale) /tɕiŋ1/ [82]	jing3 (well) /tɕiŋ3/ [207]	jiao1 (glue) /tɕiɑu1/ [75]	bing1 (soldier) /piŋ1/ [700]	gua1 (melon) /kua1/ [174]	dou4 (beans) /təu4/ [182]
9	ling3 (collar) /liŋ3/ [1605]	ling2 (bells) /liŋ2/ [113]	liu3 (willow) /liəu3/ [68]	jing3 (well) /tɕiŋ3/ [207]	gui3 (ghost) /kuei3/ [276]	gua1 (melon) /kua1/ [174]
10	miao2 (sprout) /miɑu2/ [107]	miao4 (temple) /miɑu4/ [60]	mian2 (cotton) /mian2/ [464]	qiao2 (bridge) /tɕʰiɑu2/ [621]	chuan2 (ship) /tʂʰuan2/ [1016]	gui3 (ghost) /kuei3/ [276]
11	tang2 (candy) /tʰɑŋ2/ [127]	tang1 (soup) /tʰɑŋ1/ [75]	tao2 (peach) /tʰɑu2/ [100]	lang2 (wolf) /lɑŋ2/ [105]	niu2 (cow) /niəu2/ [373]	fei4 (lungs) /fei4/ [75]
12	xin1 (heart) /ɕin1/ [3963]	xin4 (envelope) /ɕin4/ [1202]	xia1 (shrimp) /ɕia1/ [54]	jin1 (gold) /tɕin1/ [695]	dou1 (pocket) /təu1/ [51]	gou3 (dog) /kəu3/ [245]

Note. Values in parentheses represent the morphemic frequency of each item, as indicated in *Modern Chinese Frequency Dictionary*. Transcriptions are in IPA.

4 Investigating the Time Course of Mandarin Spoken Word Processing in Typically Developing Children Using ERPs

4.1 Introduction

Over time, child readers change in their sensitivity to different types of phonological relationships (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Ziegler & Goswami, 2005). This is related to the concept of phonological awareness, or knowledge of the sounds that make up spoken words (Wagner & Torgesen, 1987; McBride-Chang, 1996). Phonological awareness can refer to phonemic awareness, or awareness of the individual consonants and vowels in words; alternatively, it can also refer to awareness of larger units such as rimes (Treiman, 1986; Bryant, MacLean, Bradley, & Crossland, 1990). In a tonal language like Mandarin Chinese, there is an additional component to awareness of the sound structure of words: tone awareness. Tone can be defined as fluctuation in the pitch of a speaker's voice that is used in a lexically contrastive sense. For example, the syllable *tang* can mean 'soup' when pronounced in a high and level pitch (tone 1), while it can mean 'candy' when pronounced in a pitch that begins in the mid-range of a speaker's register and rises throughout articulation of the syllable (tone 2). Tone awareness then refers to one's ability to recognize, identify, or manipulate the tones of a language.

Phonological awareness is typically measured using overt behavioral tasks such as the oddity task (Burnham et al., 2011; Chen et al., 2004). In this task, subjects are presented with triplets of syllables, two of which share a critical feature, and are asked to select the odd one out. Using these types of measures, the developmental progression of phonological awareness for the components of the Mandarin syllable has been fairly well characterized (Chen et al., 2004; Shu, Peng, & McBride-Chang, 2008). These studies have provided evidence that awareness of larger units, such as syllables and tones, precedes awareness of smaller units such as onsets and rimes (Shu et al., 2008). However, a limitation of these types of measures is that they index metalinguistic knowledge of spoken words but fail to capture potential differences between individuals in the cognitive processing leading up to a behavioral response. For example, using eyetracking, Desroches, Robertson, and Joanisse (2006) showed that two groups of English-speaking children differed in online processing of rhyming words, even though they failed to show a difference in accuracy on a rhyme judgment task.

These types of results indicate that measures of online spoken word processing offer information that is not always apparent in behavorial measures. For this reason, we recorded online measures of spoken word processing in a group of Mandarin-speaking children during a task that captures processing of Mandarin syllable components. Our rationale is that of late there has been a considerable amount of interest in developing theories of spoken word processing in Mandarin and other tonal languages (Li, Tan, Bates, & Tzeng, 2006), and online language measures can offer important contributions to these types of theories (Allopenna, Magnuson, & Tanenhaus, 1998). However, prior investigations of Mandarin spoken word processing have tended to focus on adults (Ye & Connine, 1999; Lee, 2007; Malins & Joanisse, 2010; Zhao et al., 2011; Malins & Joanisse, 2012). It is our view that comparing spoken word processing in children and adults therefore offers important developmental constraints on these theories.

One way to test whether children and adults differ in how they process spoken words is to examine how they resolve different types of phonological competition. This competition can be incurred by setting up an expectation for a spoken word, which will spread activation to a set of lexical candidates that share a phonological relationship with the spoken word. As auditory input unfolds, these candidate items then compete for recognition. The manner in which listeners resolve this competition is related to their sensitivity to phonological relationships, among other factors (McMurray, Samelson, Lee, & Tomblin, 2010).

In the current study we examined how children versus adults resolved phonological competition based on shared onsets, rimes, or tones. Examining responses to these different types of phonological competition allowed us to draw conclusions regarding two key issues associated with theories of spoken word processing. The first concerns whether spoken word processing in children is incremental or holistic. The second concerns whether or not spoken word recognition in Mandarin is entirely feedforward.

Regarding the first issue, theories of spoken word processing differ as to how they compute phonological similarity among spoken words. In theories like TRACE (McClelland & Elman, 1986) and Cohort (Marslen-Wilson, 1987), temporal information is taken into account when calculating phonological similarity. Therefore, these theories predict different levels of competition among words sharing onset (*cat*-*cab*) versus words sharing rhyme (*cat*-*hat*). The reason for this is that shared phonological information among these types of words is signaled at different time points during the unfolding of a spoken word, and onset-based competition is thought to be stronger in nature than rhyme-based competition because onset competitors cannot be disambiguated until later on in the spoken word (Marslen-Wilson & Zwitserlood, 1989). Cohort and TRACE differ in that Cohort predicts that onset-based competition is the only type of competition amongst spoken words, while TRACE allows for rhyme effects but predicts they are much weaker in nature than onset-based competition. However, both theories are similar in that the temporal structure of a word affects the dynamics of lexical competition. This type of processing can be considered incremental, as differences between the acoustic signal and word-level representations are mapped as the spoken word unfolds (Magnuson, Tanenhaus, Aslin, &

Dahan, 2003). In contrast, the neighborhood activation model (NAM; Luce & Pisoni, 1998) defines a competitor as differing from a spoken word in one phoneme in any position, whether initial, medial, or final. Therefore, like TRACE it predicts that rhyming forms are part of the competitor set of a spoken word; however, they are considered to give rise to competitive effects that are equivalent to those of onset-based competitors. This type of processing has sometimes been termed 'holistic', as a type of global similarity among spoken words is computed without taking temporal information into account (Zhao et al., 2011).

We have previously reported evidence that spoken word processing in Mandarin-speaking adults is incremental in nature (Malins & Joanisse, 2012). In children however, there has been prior evidence for incremental processing in English (Desroches et al., 2006; Desroches, Newman, Robertson, & Joanisse, 2013; Malins et al., 2013), but this has yet to be tested in Mandarin. However, given that Mandarin-speaking children appear to be differentially sensitive to onsets, rimes, and tones behaviorally (Shu et al., 2008), we predict this should also be reflected in online processing of spoken word forms. Therefore, we hypothesize that both children and adults should show evidence of incremental processing.

The second theoretical issue relates to the proposal that there are feedback connections that influence online spoken word processing. As mentioned previously, Cohort and TRACE differ in that TRACE predicts rhyme effects while Cohort does not. The reason for this is that TRACE incorporates feedback connections from the lexical to phoneme layer (McClelland & Elman, 1986). In TRACE, expecting to hear a certain word form leads to partial activation of the phonemes that comprise it, and these spread activation in turn to words that contain these phonemes. Therefore, rhyming words that overlap with expectations in word-final phonemes receive partial activation, and hence are part of a word's competitor set. For this reason, evidence for competition among rhyming words has often been taken as support for the TRACE model and others like it (Allopenna et al., 1998). However, this evidence has mostly come from Indo-European languages (Allopenna et al., 1998; Praamstra, Meyer, & Levelt, 1994; Radeau, Besson, Fontenau, & Castro, 1998; Desroches, Newman, & Joanisse, 2009).

In contrast, rhyme effects in Mandarin speakers have been elusive. With respect to adults, rhyme effects have been observed for disyllabic words (Liu, Shu, & Wei, 2006) but there has been mixed evidence in monosyllables (Malins & Joanisse, 2010; Malins & Joanisse, 2012; Zhao et al, 2011). Therefore, there is some support for the TRACE model in Mandarin-speaking adults, although it is inconclusive. In children, however, no experiment has yet recorded online measures of spoken word processing to test for rhyme effects during rapid processing. Therefore it remains unclear whether there are developmental differences for these types of effects. In English, responses to rhymes become adult-like in typically developing children at least by age seven (Coch, Grossi, Coffey-Corina, Holcomb, & Neville, 2002) and are well-established in children around the age of nine or ten (Desroches et al., 2006; Desroches, Newman, Robertson, & Joanisse, 2013; Malins et al., 2013). One way that these rhyme effects manifest is that individuals show evidence for facilitation of rhyming versus non-rhyming forms (Praamstra et al., 1994; Radeau et al., 1998; Coch et al., 2002; Desroches et al., 2009). The idea here is because the phonemes that comprise the rime are partially active, when they finally arrive in the acoustic signal they are processed more easily than word-final phonemes in non-rhyming forms, which have not received this prior activation. For Mandarin-speaking individuals however, we hypothesize that rhyme effects will be stronger in adults compared to children. The reason we hypothesize this is that recent work using functional magnetic resonance imaging (fMRI) has shown that top-down modulation is weaker in children compared to adults during tasks of spoken word processing (Bitan et al., 2006; Bitan, Cheon, Lu, Burman, & Booth, 2009). It could be the case that this effect is especially prominent in Mandarin (Cao et al., 2011).

In the present study, we analyzed event related potentials (ERPs), an online processing measure which has not yet been used to assess spoken word processing in Mandarin-speaking children. ERPs are advantageous as they provide a continuous measure of speech perception as spoken words unfold, and furthermore certain components of the ERP waveform have been reliably associated with discrete cognitive processes (Newman, Connolly, & Forbes, 2012). We employed a picture-word matching paradigm that we have used to investigate spoken word recognition in adult native Mandarin speakers (Malins & Joanisse, 2012). In this task, subjects are presented with pictures of items, which set up expectations of subsequent auditory input (Desroches, Newman, & Joanisse, 2009). Following this, subjects hear an auditory word that either matches or mismatches expectations, and their task is to indicate whether or not the auditory word matches the picture. These mismatching forms share some phonological relationship with the expected word, such as onset or rime, but differ in other components, such as tone. In the current study, mismatches overlapped with expectations in either word-initial phonemes and tone (cohort mismatches; see: *tang2* ‘candy’, hear: *tao2* ‘peach’), word-final phonemes and tone (rhyme mismatches: see: *tang2* ‘candy’, hear: *lang2* ‘wolf’), all phonemes but not tone (segmental mismatches; see: *tang2* ‘candy’, hear: *tang1* ‘soup’), only tone (tonal mismatches: see: *tang2* ‘candy’, hear: *niu2* ‘cow’), or were completely unrelated (unrelated mismatches; see: *tang2* ‘candy’, hear *xia1* ‘shrimp’). We were interested in how these mismatches would modulate several components thought to index distinct stages of the word recognition process: the phonological mapping negativity (PMN; Connolly & Phillips, 1994; Newman & Connolly, 2009), thought to index pre-lexical phoneme mapping, and early and late portions of the N400 (Kutas & Hillyard, 1984; Connolly & Phillips, 1994; Desroches et al., 2009), thought to index whole-word recognition. Previous research has indicated that analysis of these components reveals temporally dissociable effects for different mismatch types, such as cohorts versus rhymes (Desroches et al., 2009; Archibald & Joanisse, 2011; Malins & Joanisse, 2012; Desroches et al., 2013; Malins et al., 2013). This approach is therefore well-suited to the current investigation, as we aimed to identify which specific aspects of spoken word processing differed between the children and the adults.

To summarize, we were interested in whether Mandarin-speaking children process spoken words incrementally, as well as the extent of the influence of top-down processing during word recognition in children versus adults. We hypothesized that children and adults both process spoken words incrementally, and therefore predicted that they would show a similar pattern of responses to words differing from expectations in either onset, rime, or tone. Namely, words diverging from expectations in word-initial information (i.e., rhyme mismatches) were expected to modulate the PMN response to a greater extent than words overlapping expectations in word-initial information (i.e., cohort and segmental mismatches). In addition, words differing from expectations in onsets versus rimes versus tones were expected to lead to differential modulation of the N400 and late N400. Next, as we hypothesized that top-down processing is weaker in children compared to adults, we predicted that the adults should show greater evidence of facilitated processing of rhyming forms. We expected this to result in an attenuation of the late N400 for rhyme mismatches in the adults compared to the children.

4.2 Methods

4.2.1 Subjects

Testing took place at the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University in Beijing, China. Overall, 29 children and 21 adults participated. Children were recruited from local schools in Beijing, and adults were recruited from Beijing Normal University and the surrounding community. Twelve children and four adults were removed due to either excessive EEG artifacts or equipment failure, for a total of 17 children and 17 adults included in the experiment (children: 12 female, mean age 10;5, see Table 4.1 for full details; adults: 11 female, mean age 24). As it has been shown that there is a relationship between phonology and reading ability in Chinese children (Cheung et al., 2009; Liu, Shu, & Yang, 2009), we wished to screen out atypical readers to make sure that our results were not contaminated by differences in reading ability across children. To do this, we screened the children using a test of reading ability as well as a test of phoneme awareness; the latter was included because deficits in phoneme awareness have also been implicated in reading impairment in Chinese (Ho, Law, & Ng, 2000; Newman, Tardif, Huang, & Shu, 2011). All children were identified as typically developing readers based on scoring at or above the expected average for their grade level on these two tests. The reading test was a single character reading task taken from Lei et al. (2011), which required children to read aloud 150 Chinese characters, all of which are expected to be learned by the sixth grade in Beijing (Shu, Chen, Anderson, Wu, & Xuan, 2003). The characters were arranged from the simplest to the most challenging based on complexity and the grade level at which a character is first learned. The phoneme awareness test was a phoneme deletion task, taken from Li et al. (2012), in which

Table 4.1: Descriptive Statistics for the Group of Typically Developing Children (*N* = 17; 12 Female)

Measure	*Mean (SD)*
Age (months)	124.8 (13.7)
Verbal IQ[a]	114.5 (13.1)
Performance IQ[a]	114.9 (11.3)
Phoneme deletion[b]	22.7 (3.3)
Single character reading[b]	118.0 (14.1)

[a] Standard scores.

[b] Raw score out of a maximum of 28 for phoneme deletion and 150 for single character reading.

IQ = Intelligence Quotient

children were asked to produce a new syllable after the deletion of a target phoneme in either the initial, medial, or final position. There were 28 trials in this task. In addition, cognitive ability was assessed using the Chinese version of the Wechsler Intelligence Scale for Children (WISC-R; Wechsler, 1974), which verified that all children had normal or above normal IQ (IQ > 90).

All materials and procedures were approved by the Institutional Review Board at Beijing Normal University. When testing children, each child's caregivers gave written consent to participate in the study, and each family received a small gift as well as a performance report after attending the experimental session.

4.2.2 Stimuli and Procedures

All stimuli were monosyllabic Mandarin words referring to common, highly imageable objects. Sixteen sets of items (as listed in the Appendix) were created by changing single Mandarin words in ways that created phonologically related competitor words; for example, for the critical word *tang2* 'candy', we derived competitor forms that were different in either tone (segmental mismatch: *tang1* 'soup'), rime (cohort mismatch: *tao2* 'peach'), onset (rhyme mismatch: *lang2* 'wolf'), all segments but not tone (tonal mismatch: *niu2* 'cow'), or all segments and tone (unrelated mismatch: xia1 'shrimp'). By deriving stimuli from these critical words, we were able to control as much as possible attendant psycholinguistic factors influencing single-word recognition, as each item participated equally in all conditions. In addition, as much as possible, mismatching items were selected from the other sets in the experiment in order to have a closed set of stimuli. Items in the five mismatch conditions did not differ from one another in logarithmic frequency [$F(4,60) = .686$, $p = .58$, $\eta_p^2 = .04$].

Auditory stimuli were single-syllable words spoken in isolation by an adult male speaker of Beijing Mandarin, digitized and recorded to disk at 16 bits quantization, with a 44,100 Hz sampling rate, and volume normalized to -10 dB of maximum amplitude. Each item was spoken seven times, and the middle five tokens were selected for use in the experiment. Items in the five mismatch conditions did not differ from one another in duration [$F(4,60) = 1.51$, $p = .22$, $\eta_p^2 = .09$].

Culturally appropriate pictures depicting these items were selected with the assistance of native Mandarin speakers. To verify that subjects were familiar with the names of these pictures, we asked subjects at the beginning of the experimental session to name aloud each of the pictures by saying the first Mandarin syllable/character that came to mind upon viewing the picture. This allowed us to ensure they were comfortable with the names we had selected for each picture and associated them with monosyllabic words rather than potential disyllabic alternatives. Subjects showed high levels of accuracy on this task [adults: 89.5% (SE = .77%); children: 83.8% (SE = 1.42%)], indicating that the selected names for the pictures were familiar and appropriate for the subjects. This was further supported by high levels of accuracy for the subsequent picture-word matching task in both groups. In cases where the given name differed from the intended name, subjects were told the intended name, as in prior picture naming studies (Brooks & MacWhinney, 2000).

In the experimental trials, subjects performed a picture-word matching task while we recorded continuous EEG activity from the surface of their scalp. In each trial, subjects viewed a fixation cross for 250 ms, following which a picture of an item appeared on 21-inch CRT monitor for 1500 ms. Next, while this item remained on screen, subjects heard an auditory word presented binaurally via Sennheiser HD201 headphones. Following responses, a blank screen was presented for 1000 ms prior to the onset of the next trial.

Subjects were asked to indicate via button press whether the auditory word matched or mismatched the picture present on the screen. Across all trials in the experiment, the match to mismatch ratio was one to one, and the five mismatch trial types (cohort, etc.) were presented with equal probability. Subjects completed 320 trials (160 match trials; 32 trials for each of the five mismatch types). For all mismatch trials, the roles of pictures and words were counterbalanced (e.g., for the pair *tang2*-*tang1*, in one trial *tang2* was the picture and *tang1* the sound, and in another trial *tang1* was the picture and *tang2* the sound). This ensured that every time a picture or sound item was presented, it was equally likely to be a match or a mismatch trial. Testing was divided into four blocks of trials with short rests between blocks. Block order was counterbalanced across subjects, which allowed us to control for whether each item was a match or mismatch the first time it was encountered. Prior to the experimental trials, subjects performed a training block of six trials containing items not presented in the actual experiment.

4.2.3 EEG Data Acquisition

EEG data were collected at a sampling rate of 1000 Hz using Acquire 4.2 (Neurosoft Inc., El Paso, TX) and 32-channel caps with sintered Ag/AgCl electrodes (Quik-Caps; Neurosoft Inc., El Paso, TX) oriented according to the international 10–20 system and referenced to the nose tip. Across subjects, the impedance of each channel was kept below 5 kΩ.

Data were amplified at a gain of 500 using a SynAmps amplifier and filtered online using 60 Hz notch and .1 – 100 Hz bandpass filters. ERPs were then segmented into epochs spanning from 200 ms pre-stimulus to 750 ms post-stimulus onset, time-locked to the onset of the auditory stimulus. Following this, data were filtered offline using a 24 dB zero phase shift digital bandpass filter (0.1 – 30 Hz), and baseline corrected to the mean voltage of the pre-stimulus interval. Trials containing blinks and other artifacts were removed using a maximum voltage criterion of ±100 μV for the portion of the waveform subjected to statistical analysis (0 - 600 ms) on the following electrodes: Fz, F3, F4, F7, F8, Cz, C3, C4, T7, T8, Pz, P3, P4, P7, P8. In addition, incorrect trials were rejected. Following artifact rejection, the average number of accepted trials per condition for the children was 121/160 for the match trials, 23/32 for segmental mismatches, 23/32 for cohort mismatches, 24/32 for rhyme mismatches, 25/32 for tonal mismatches, and 24/32 for the unrelated mismatch condition, while for adults the average number of accepted trials per condition was 149/160 for the match trials, 29/32 for segmental mismatches, 29/32 for cohort mismatches, 30/32 for rhyme mismatches, 30/32 for tonal mismatches, and 29/32 for the unrelated mismatch condition. During data recording, there were several bad channels; namely, F7 for two of the children, T8 for two of the children, and C3 for one of the children and eight of the adults. Values for these channels were calculated using the mean of the surrounding electrodes (for F7: F3 and FT7; for T8: FT8, C4, and TP8; for C3: T7, FC3, Cz, and CP3).

Because the ratio of accepted trials differed between groups [$t(32) = -6.41$, $p < .001$, Cohen's $d = 2.20$], we calculated signal to noise ratios for the fifteen electrodes subjected to data analysis. This was done by dividing the root mean square of the voltage for the period following the auditory stimulus (0 – 600 ms) by the root mean square of the pre-stimulus baseline period (-200 – 0 ms). Comparison between groups showed that SNRs were not different between the children and the adults [$t(32) = .252$, $p = .80$, Cohen's $d = .09$]. The results indicate that any observed differences between adults and children are not likely due to differences in the signal to noise ratio of our ERP recordings.

4.2.4 Analysis of ERP Data

Because our aim was to compare the adults and children, we analyzed only difference waves when comparing groups. Our rationale was that performing between-groups statistics on relative voltage differences between the word type conditions rather than on absolute voltages helps offset potential differences in the absolute magnitude of ERP waveforms. These difference waves were generated by subtracting each group's match condition from respective mismatch conditions.

For both groups, we analyzed the following ERP components: the PMN, the N400, and the late N400, as in prior picture-word matching studies (Desroches et al., 2009; Archibald & Joanisse, 2011; Malins & Joanisse, 2012; Desroches et al., 2012; Malins et al., 2013). Windows of analysis were determined by visual inspection, and were defined as follows: PMN (260-320 ms), N400 (350-500 ms), and late N400 (500-600 ms). We calculated the magnitude of each ERP response for each respective mismatch condition as the difference in mean voltage compared to match over the specified time window. For each analysis window in each group, we selected fifteen electrodes (Fz, F3, F4, F7, F8, Cz, C3, C4, T7, T8, Pz, P3, P4, P7, P8), which provided full scalp coverage in a way that would allow us to differentiate the components of interest (Newman, Connolly, Service, & McIvor, 2003). We then analyzed mean amplitude values using a ‘column’ approach to ERP analysis similar to that described in Holcomb and Grainger (2006). That is, we conducted separate repeated-measures analyses of variance for lateral, medial, and midline electrodes. Midline electrodes were further divided into anterior (Fz), central (Cz), and posterior (Pz) regions, yielding a 3-way mixed design ANOVA with two within-subjects factors (3 region, 5 mismatch type), and a between-subjects factor of group (2; adults and children). Medial electrodes were also divided into anterior (F3/F4), central (C3/C4), and posterior (P3/P4) regions, and there was an additional within-subjects factor of hemisphere (left or right). Similarly, for lateral electrodes, we divided electrodes according to region (anterior: F7/F8; central: T7/T8; posterior P7/P8), and hemisphere. Thus, lateral and medial electrodes were analyzed with a four-way mixed design ANOVA (within-subjects factors: hemisphere, region, and word type; between-subjects factor: group). All analyses of variance were conducted using conservative degrees of freedom (Greenhouse & Geisser, 1959).

For each component, we focused on interactions between word type and group; if any of these were significant, we performed step-down repeated-measures ANOVA separately for each group. If there were no significant interactions between word type and group, we then focused on interactions between word type and the within-subjects factors of region and hemisphere. If none of these were significant, we then focused on main effects of word type.

Hypothesis 1: Both Children and Adults Process Spoken Words Incrementally. To test this hypothesis, we followed up on significant interactions and main effects in the PMN, N400, and late N400 windows with pairwise comparisons between the following mismatch types: segmental, cohort, and rhyme. The reason for this is that these three conditions differed as to which component of the syllable diverged from expectations: tone for the segmental condition, word-final vowels for the cohort condition, and onset consonants for the rhyme condition. Because these mismatches are signaled at different times during the unfolding of the syllable, we predicted that we would see differences among these conditions in the PMN, N400, and late N400 windows, which can be considered evidence of incremental processing. As we hypothesized that both children and adults process spoken words incrementally, we predicted both groups would show a similar pattern of mismatch effects for these three conditions.

Hypothesis 2: Top-Down Processing is Weaker in Children Compared to Adults. To test this hypothesis, we followed up on significant interactions in the late N400 window by performing simple effects analyses that contrasted the rhyme condition between groups. The reason we did this is that in English, an attenuation of the late N400 for rhyming forms has been observed in multiple studies, and has been taken as evidence of an influence of top-down processing (Praamstra, Meyer, & Levelt, 1994; Radeau, Besson, Fontenau, & Castro, 1998; Coch et al., 2002; Desroches et al., 2006; Desroches et al., 2009; Desroches et al., 2013; Malins et al., 2013). Because we hypothesized that top-down processing is weaker in Mandarin-speaking children compared to adults, we expected that the late N400 amplitude would be attenuated in adults compared to children.

4.3 Results

4.3.1 Behavioral Data

Table 4.2 presents behavioral performance for the picture-word matching task for both the adults and children. Trials were removed from behavioral analysis if decision latencies were over 2.5 standard deviations above or below condition-wise means, or were less than 100 ms. Furthermore, decision latencies were only analyzed for correct trials. For decision latencies, a mixed repeated-measures ANOVA with word type as a within-subjects factor (6) and group as a between-subjects factor (2), showed that there was a significant interaction between group and word type [$F(5,170) = 3.052$, $p = .02$, $\eta_p^2 = .08$], as well as a main effect of group [$F(1,34) = 31.388$, $p = .001$, $\eta_p^2 = .48$], as decision latencies were generally longer for the children. For accuracy, there was a main effect of word type [$F(5,170) = 4.116$; $p = .01$; $\eta_p^2 = .11$] as well as a main effect of group [$F(1,34) = 15.743$; $p = .001$; $\eta_p^2 = .32$], as accuracy was slightly lower in the children (although still well over 90% in all word type conditions); however, there was no significant interaction between word type and group [$F(5,170) = .664$; $p = .59$; $\eta_p^2 = .02$]. The values in Table 4.2 clearly show that both the adults and the children performed well on the task.

4.3.2 ERP Data

Condition-wise waveforms for the adults and children for the different mismatch conditions are shown in Figures 4.1 though 4.6, while difference waveforms are plotted in Figures 4.7 through 4.10. Table 4.3 presents the ANOVA analyses for effects of age group, word type, and electrode position for each component.

Table 4.2: Mean Decision Latencies and Mean Percent Accuracy for the Matching Judgment Relative to Word Onset

	Adults (N = 17)		*Children (N = 17)*	
Word Type Condition	*Decision Latency (ms)*	*Percent Accuracy*	*Decision Latency (ms)*	*Percent Accuracy*
Match	744.1 (24.4)	99.3 (0.11)	996.3 (30.2)	94.5 (0.90)
Segmental mismatch	835.0 (22.3)*	97.3 (0.84)*	1097.1 (45.7)*	94.5 (1.31)
Cohort mismatch	850.8 (23.0)*	96.8 (0.95)*	1137.2 (41.3)*	93.8 (1.61)
Rhyme mismatch	792.4 (22.0)*	99.2 (0.34)	1041.0 (33.1)*	96.2 (1.01)
Tonal mismatch	769.5 (19.0)	99.8 (0.19)*	1055.5 (43.4)*	96.4 (0.89)
Unrelated mismatch	761.3 (19.5)	99.3 (0.43)	1058.2 (42.7)*	97.0 (0.79)*

Note. Values in parentheses represent standard errors. Asterisks represent pairwise comparisons between respective mismatch conditions and the match condition that were significant at $p < .05$.

4.3.2.1 Hypothesis 1: Both Children and Adults Process Spoken Words Incrementally

To test this hypothesis, we examined responses to the different mismatch conditions in all three of the PMN, N400, and N400 windows. For the PMN, as is shown in Table 4.3, analyses did not reveal an interaction between group and word type in any column. This suggests the relationship between the magnitude of the PMN response and mismatch type was comparable across the adults and children. However, when the two groups were collapsed, there was a significant interaction between region and word type in all three columns [midline: $F(8,256) = 4.195$, $p = .002$, $\eta_p^2 = .12$; medial: $F(8,256) = 2.804$, $p = .03$, $\eta_p^2 = .08$; lateral: $F(8,256) = 2.552$, $p = .05$, $\eta_p^2 = .07$], as well as an interaction between hemisphere and word type in the lateral column [$F(4,128) = 2.793$, $p = .04$, $\eta_p^2 = .08$]. Pairwise comparisons between the segmental,

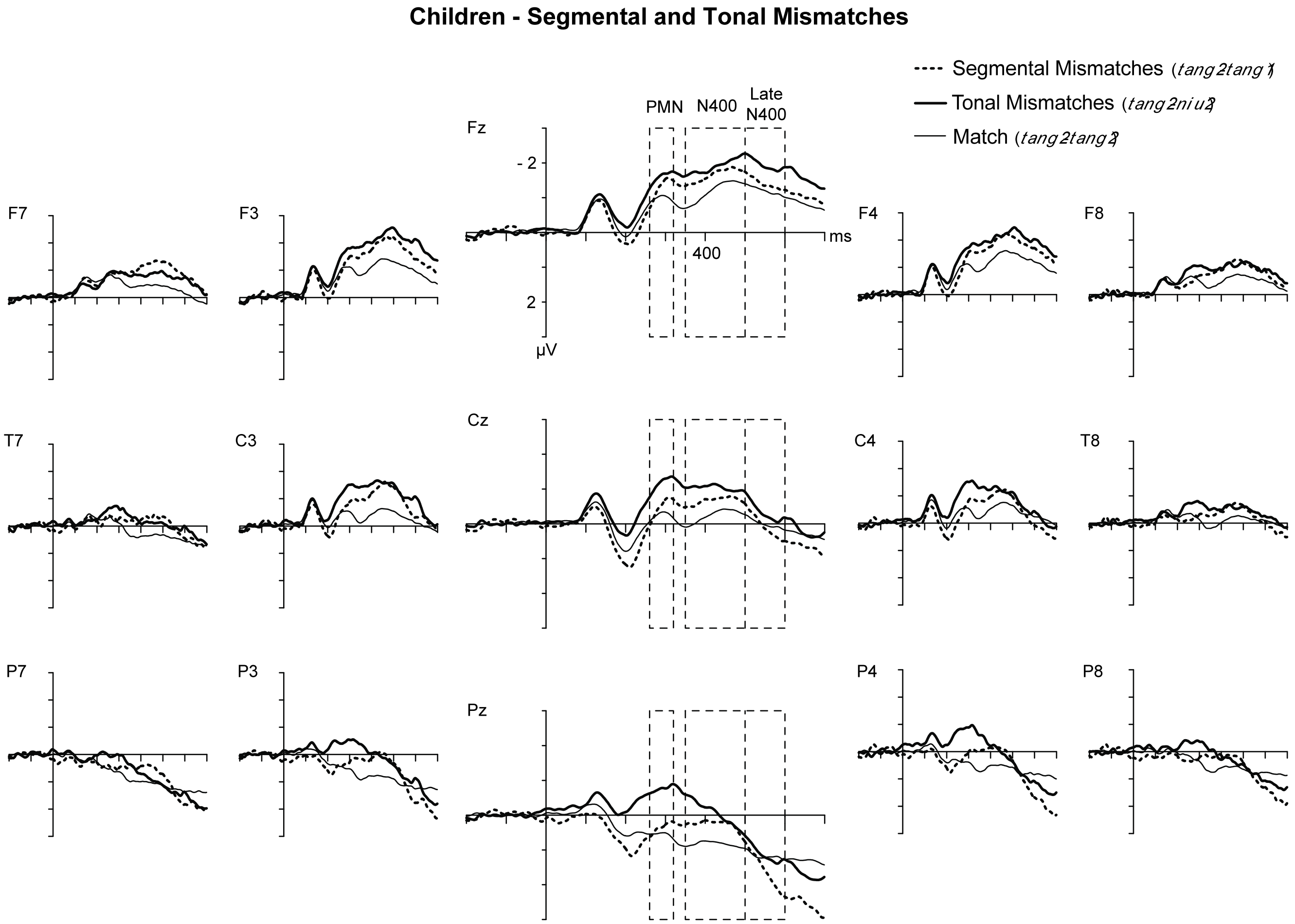


Figure 4.1: Condition-wise waveforms for the segmental, tonal, and match conditions for the group of children. The boxes delineate the PMN, N400, and late N400 windows, which were assessed via statistical analysis of subtracted waveforms.

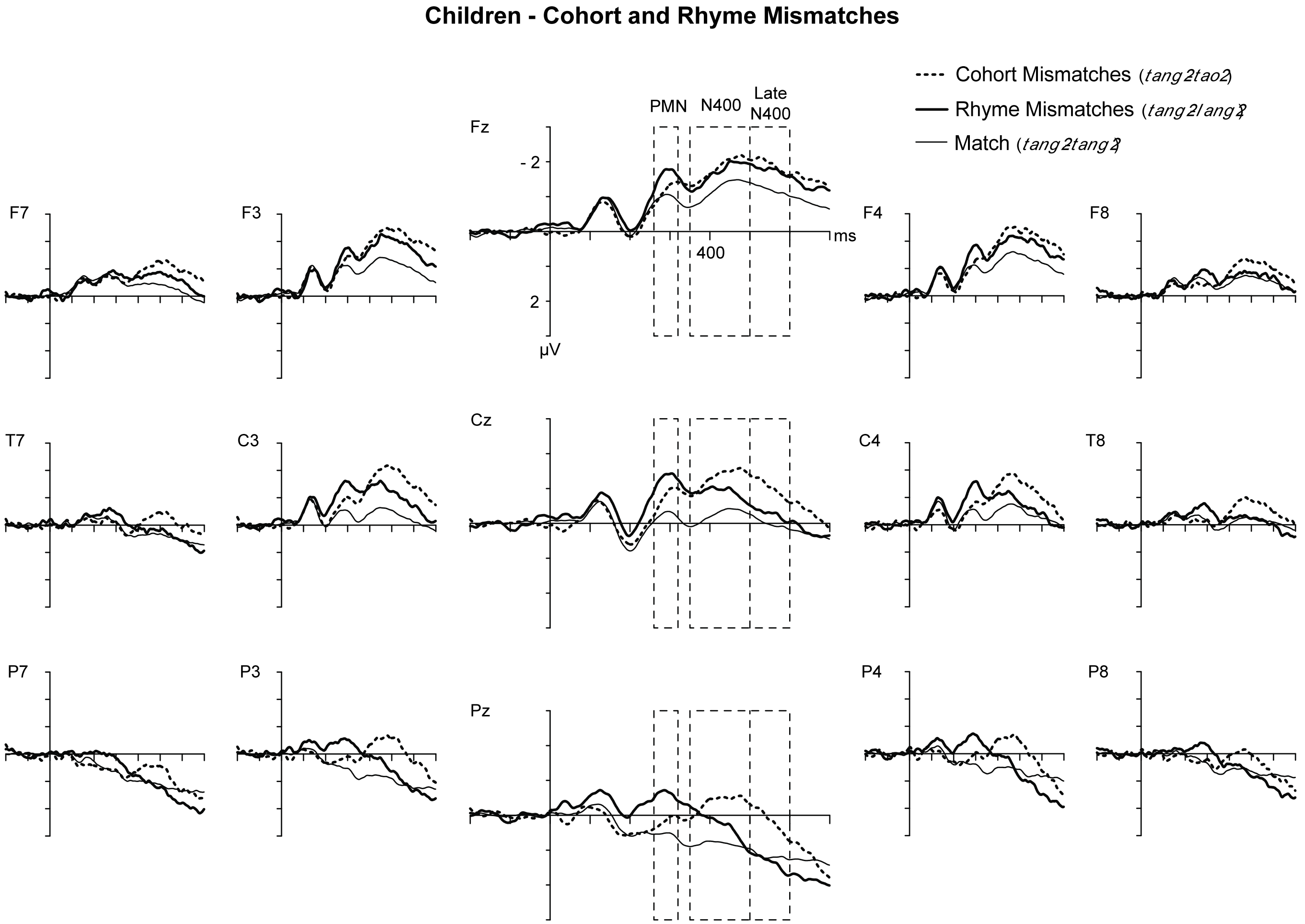


Figure 4.2: Condition-wise waveforms for the cohort, rhyme, and match conditions for the group of children. The boxes delineate the PMN, N400, and late N400 windows, which were assessed via statistical analysis of subtracted waveforms.

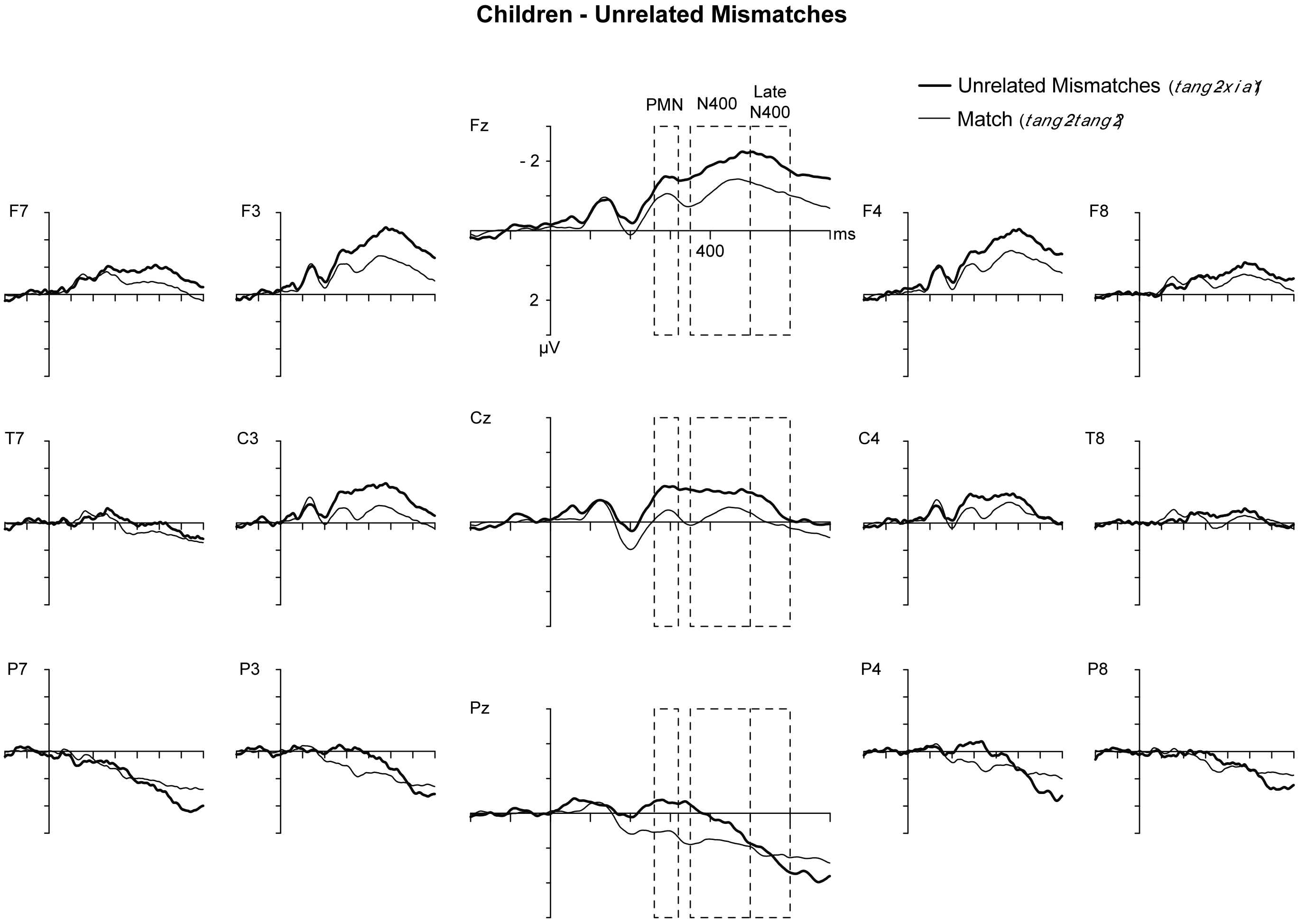

Children - Unrelated Mismatches

Figure 4.3: Condition-wise waveforms for the unrelated and match conditions for the group of children. The boxes delineate the PMN, N400, and late N400 windows, which were assessed via statistical analysis of subtracted waveforms.

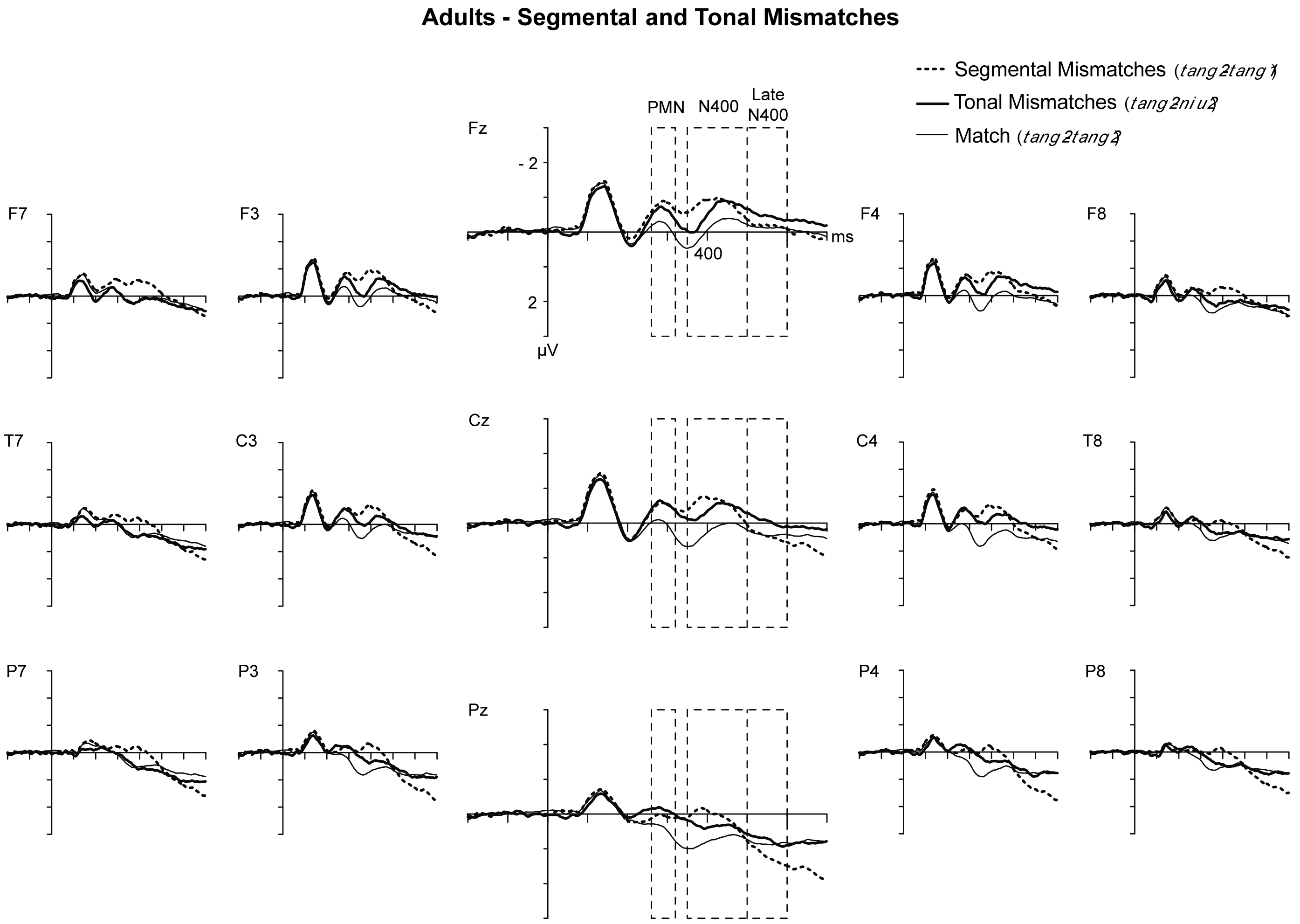


Figure 4.4: Condition-wise waveforms for the segmental, tonal, and match conditions for the group of adults. The boxes delineate the PMN, N400, and late N400 windows, which were assessed via statistical analysis of subtracted waveforms.

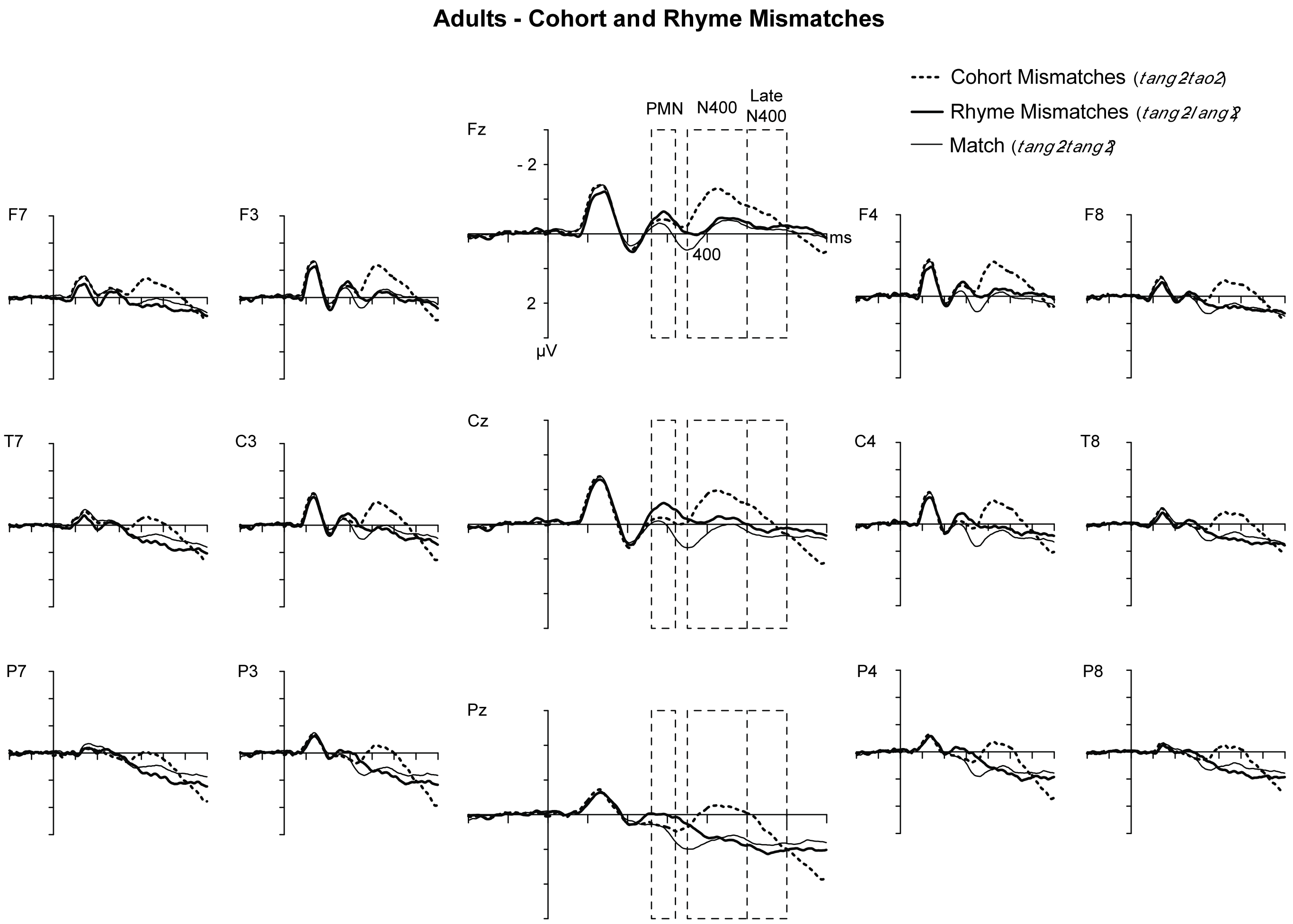


Figure 4.5: Condition-wise waveforms for the cohort, rhyme, and match conditions for the group of adults. The boxes delineate the PMN, N400, and late N400 windows, which were assessed via statistical analysis of subtracted waveforms.

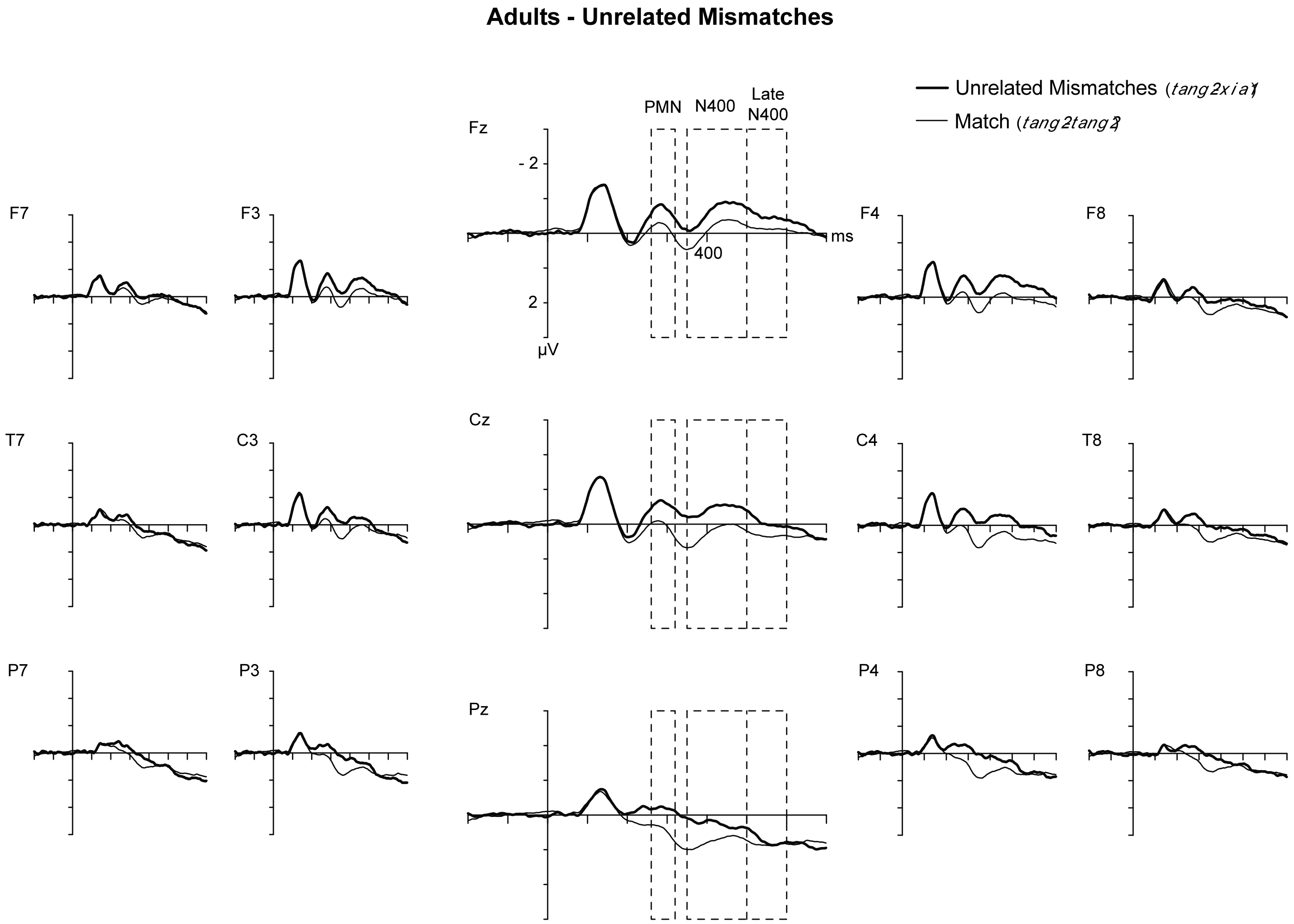


Figure 4.6: Condition-wise waveforms for the unrelated and match conditions for the group of adults. The boxes delineate the PMN, N400, and late N400 windows, which were assessed via statistical analysis of subtracted waveforms.

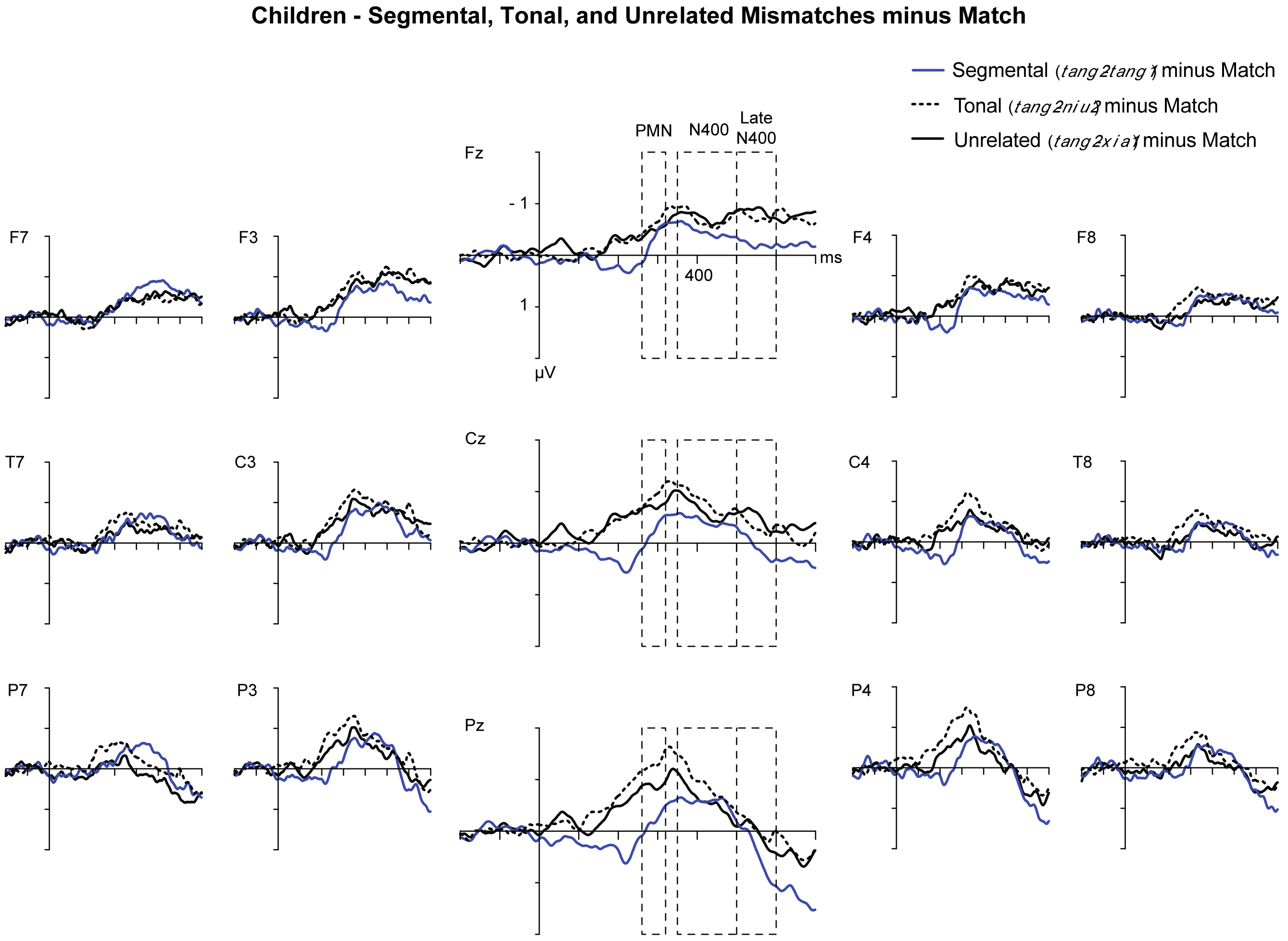


Figure 4.7: Difference waveforms for the segmental, tonal, and unrelated conditions in the children, which were generated by subtracting the match condition from each respective mismatch condition. The boxes delineate the PMN, N400, and late N400 windows, which were subjected to statistical analysis.

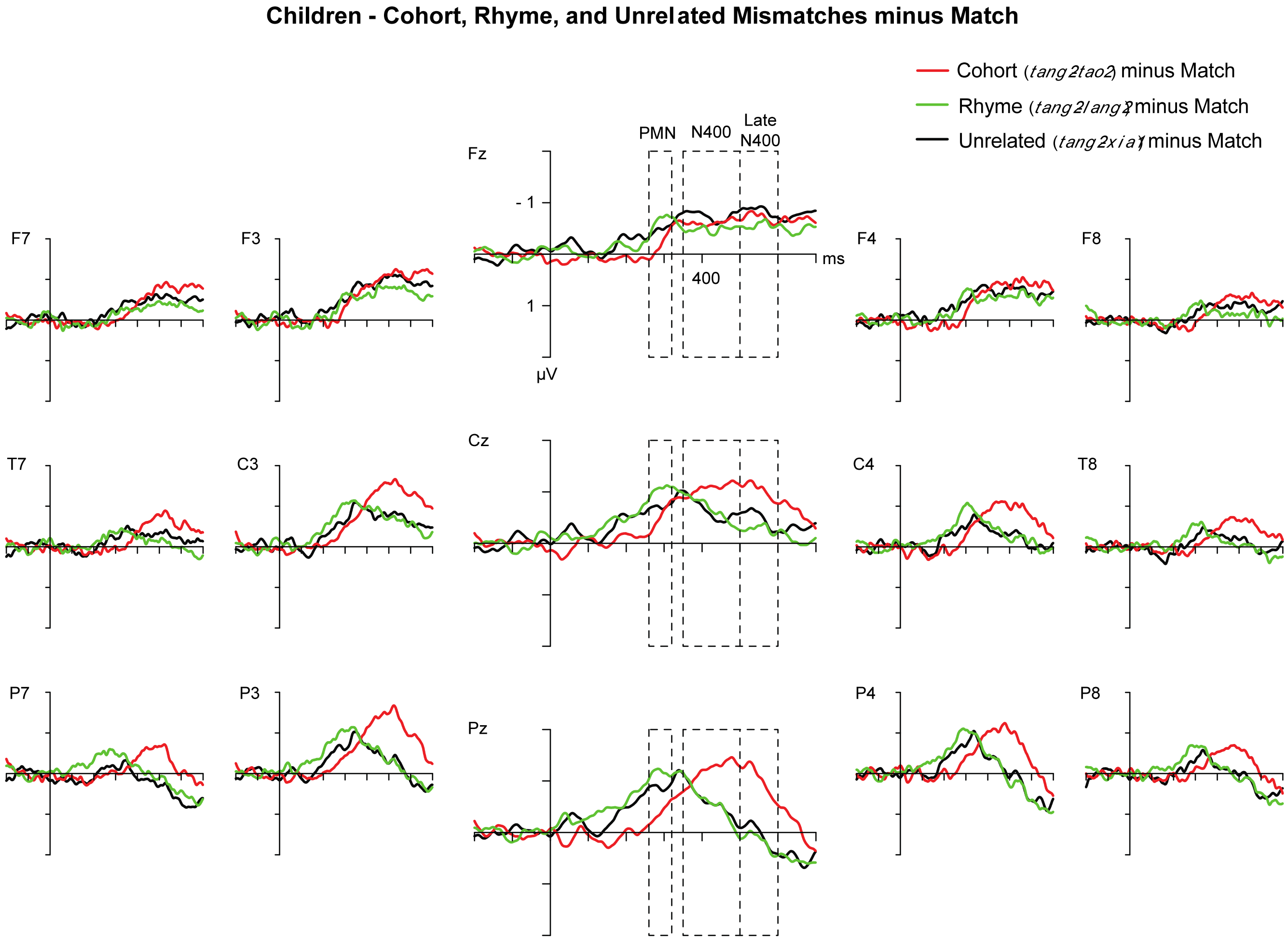
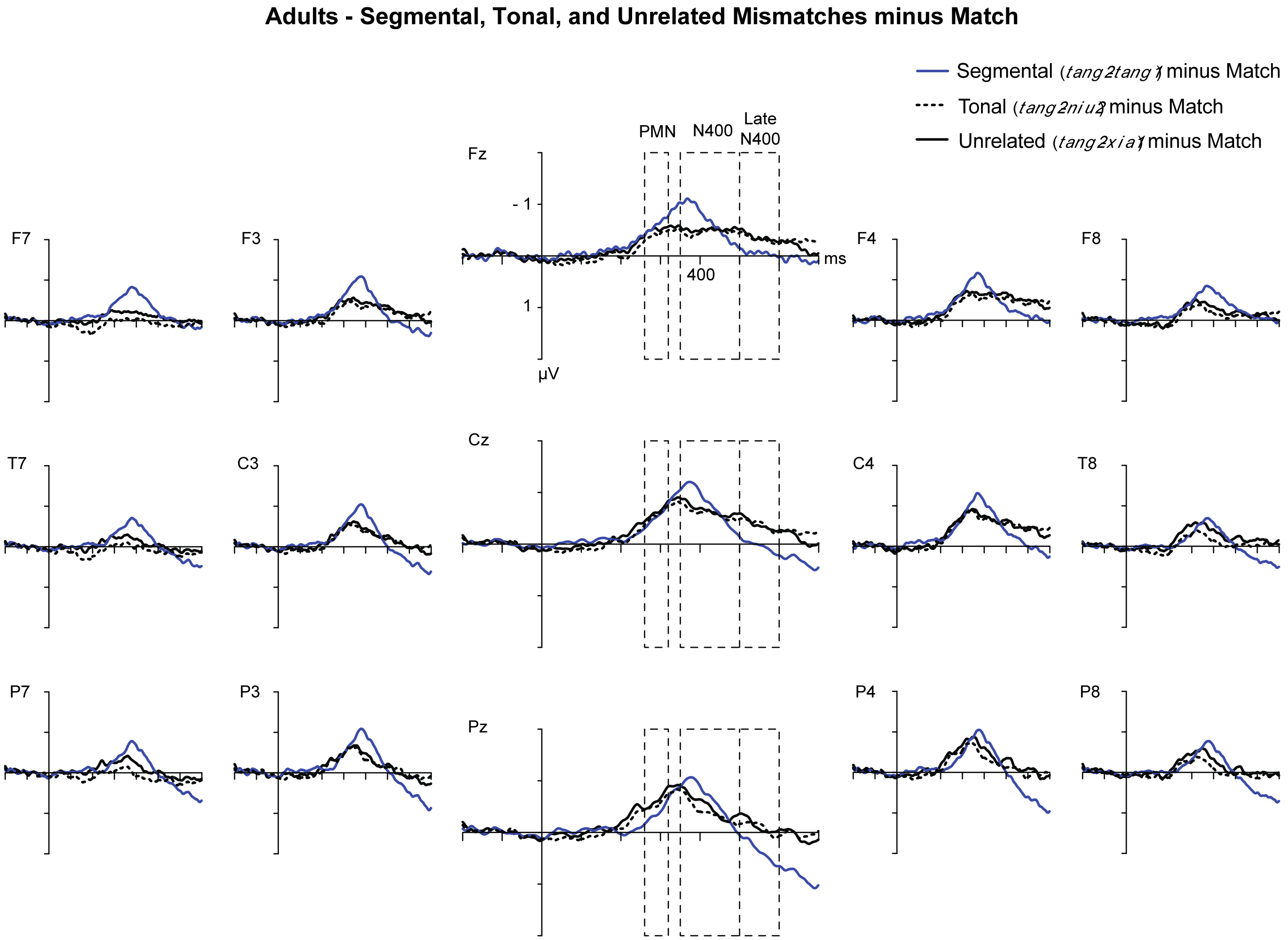


Figure 4.8: Difference waveforms for the cohort, rhyme, and unrelated conditions in the children, which were generated by subtracting the match condition from each respective mismatch condition. The boxes delineate the PMN, N400, and late N400 windows, which were subjected to statistical analysis.



Adults - Segmental, Tonal, and Unrelated Mismatches minus Match

Figure 4.9: Difference waveforms for the segmental, tonal, and unrelated conditions in the adults, which were generated by subtracting the match condition from each respective mismatch condition. The boxes delineate the PMN, N400, and late N400 windows, which were subjected to statistical analysis.

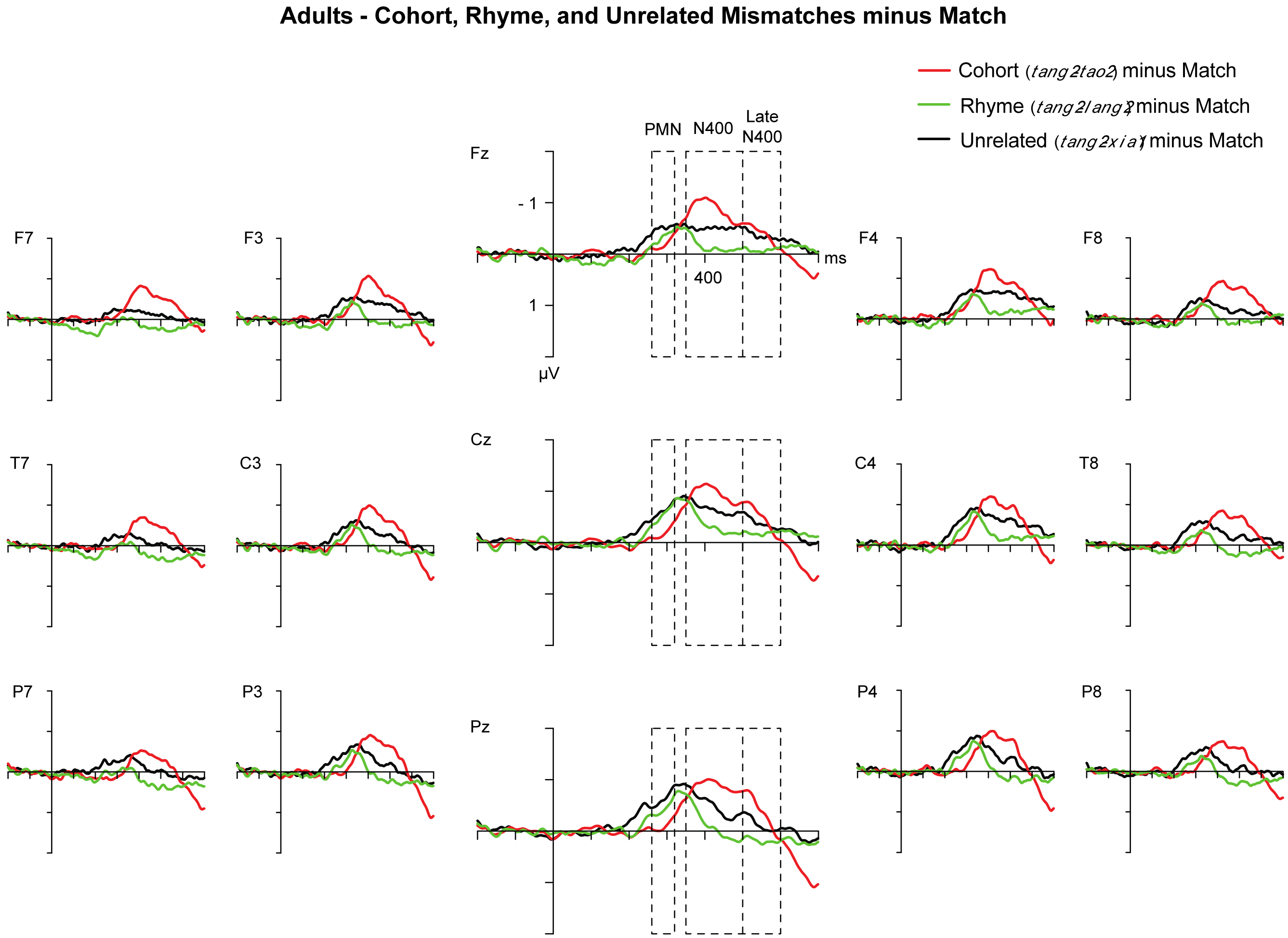


Figure 4.10: Difference waveforms for the cohort, rhyme, and unrelated conditions in the adults, which were generated by subtracting the match condition from each respective mismatch condition. The boxes delineate the PMN, N400, and late N400 windows, which were subjected to statistical analysis.

Table 4.3: Summary of ANOVAs for mean amplitude of the PMN, N400, and late N400, with one between-subjects factor of group (2), and either one within-subjects factor of anterior-posterior region (3) for the midline column, or two within-subjects factors of hemisphere (2) and anterior-posterior region (3) for medial and lateral columns.

Effect	*df*	*PMN*			*N400*			*Late N400*		
		Midline	*Medial*	*Lateral*	*Midline*	*Medial*	*Lateral*	*Midline*	*Medial*	*Lateral*
Group *F*	1,32	1.665	1.189	.245	.541	2.127	.861	1.190	5.114	2.175
p/η_p^2		.21/.05	.28/.04	.62/.01	.47/.02	.16/.06	.36/.03	.28/.04	.03/.14	.15/.06
Word type	4,128	3.676	3.745	2.058	**3.180**	**4.286**	5.593	5.323	4.851	**3.589**
		.01/.10	.01/.11	.11/.06	**.03/.09**	**.01/.12**	.001/.15	.001/.14	.003/.13	**.01/.10**
Region	2,64	5.205	3.676	1.193	1.921	2.404	5.225	18.368	34.741	35.492
		.02/.14	.06/.10	.30/.04	.17/.06	.12/.07	.02/.14	.001/.37	.001/.52	.001/.53
Hemisphere	1,32	–	4.136	3.772	–	.595	.572	–	2.716	.059
			.05/.11	.06/.11		.45/.02	.46/.02		.11/.08	.81/.01
Word type × Group	4,128	1.948	2.596	2.023	1.076	1.149	1.198	.551	.448	.259
		.12/.06	.06/.08	.11/.06	.36/.03	.33/.04	.32/.04	.67/.02	.73/.01	.87/.01
Region × Group	2,64	3.943	6.933	3.765	1.566	.080	.444	2.712	5.460	7.344
		.04/.11	.01/.18	.05/.11	.22/.05	.86/.01	.58/.01	.09/.08	.01/.15	.003/.19
Hemisphere × Group	1,32	–	6.699	1.521	–	30.158	1.650	–	47.778	6.970
			.01/.17	.23/.05		.001/.49	.21/.05		.001/.60	.01/.18
Region × Word Type	8,256	**4.195**	**2.804**	**2.552**	1.796	1.329	.941	5.628	3.465	1.695
		.002/.12	**.03/.08**	**.05/.07**	.11/.05	.26/.04	.45/.03	.001/.15	.01/.10	.15/.05
Region × Word Type × Group	8,256	.766	.617	1.100	1.623	1.478	**2.368**	**2.632**	**2.510**	1.930
		.56/.02	.65/.02	.36/.03	.15/.05	.21/.04	**.05/.07**	**.03/.08**	**.04/.07**	.10/.06
Hemisphere × Word Type	4,128	–	1.554	**2.793**	–	.135	1.386	–	.102	1.323
			.20/.05	**.04/.08**		.95/.01	.25/.04		.97/.01	.27/.04
Hemisphere × Type × Group	4,128	–	.751	.197	–	.812	.080	–	.513	.088
			.55/.02	.90/.01		.50/.03	.98/.01		.69/.02	.97/.01
Hemisphere × Region	2,64	–	.597	.125	–	2.287	1.253	–	9.005	1.110
			.52/.02	.81/..01		.11/.07	.29/.04		.001/.22	.33/.03
Hemisphere × Region × Group	2,64	–	4.909	.110	–	11.466	1.704	–	6.995	3.153
			.02/.13	.82/.01		.001/.26	.20/.05		.002/.18	.06/.09
Hemisphere × Region × Type	8,256	–	.719	1.261	–	.149	.467	–	.324	.729
			.61/.02	.28/.04		.99/.01	.84/.01		.90/.01	.63/.02
Hemi × Region × Type × Group	8,256	–	.881	.977	–	.772	.952	–	.800	1.038
			.50/.03	.44/.03		.58/.02	.46/.03		.56/.02	.40/.03

Note. Bold values indicate significant main effects/interactions that are further explored in the text.

cohort, and rhyme conditions (presented in Table 4.4) revealed that the PMN response was larger in the rhyme condition than in the segmental and cohort conditions.

For the N400, there was no significant interaction between word type and group in the midline and medial columns, nor interactions between word type and region or hemisphere. However, there was a main effect of word type in both columns [midline: $F(4,128) = 3.180$, $p = .03$, $\eta_p^2 = .09$; medial: $F(4,128) = 4.286$, $p = .01$, $\eta_p^2 = .12$]. Follow-up pairwise comparisons (Table 4.4) revealed that the N400 was larger in amplitude for cohort and segmental mismatches compared to rhyme mismatches.

The spatial extent of this effect was slightly larger in the adults, as there was a three-way interaction between region, word type, and group in the lateral column [$F(8,256) = 2.368$, $p = .05$, $\eta_p^2 = .07$]. Separate repeated-measures ANOVAs for each group revealed a significant interaction between word type and region in the adults [$F(4,64) = 9.041$, $p = .001$, $\eta_p^2 = .36$]; however, in the children there was no significant interaction between word type and region nor a main effect of word type. Pairwise comparisons (Table 4.4) revealed that the adults showed a larger N400 for cohort and segmental mismatches compared to rhyme mismatches, the same pattern observed for both the children and the adults in the midline and medial columns.

Last, for the late N400, there were significant three-way interactions between region, word type, and group in the midline and medial columns [midline: $F(8,256) = 2.632$, $p = .03$, $\eta_p^2 = .08$; medial column: $F(8,256) = 2.510$, $p = .04$, $\eta_p^2 = .07$]. Separate repeated-measures ANOVAs within each group revealed an interaction between region and word type in the children [midline: $F(8,128) = 4.828$, $p = .002$, $\eta_p^2 = .23$; medial column: $F(8,128) = 3.499$, $p = .01$, $\eta_p^2 = .18$], and a

Table 4.4: Results of Pairwise Comparisons Relevant to Hypothesis 1

Component	*Comparison*	*Group*	*Location*	*df*	*t*	*Cohen's d*
PMN	Rhy > Seg	Both	Cz	33	-2.66**	.46
			C3 and C4	33	-2.18*	.37
			Pz	33	-3.11**	.53
			P3 and P4	33	-2.63**	.45
	Rhy > Cht	Both	Cz	33	-3.06**	.52
			C3 and C4	33	-3.30**	.57
			T7 and T8	33	-2.40*	.41
			Pz	33	-3.23**	.55
			P3 and P4	33	-3.50**	.60
			P7 and P8	33	-2.83**	.49
N400	Cht > Rhy	Both	midline	33	-4.06**	.70
			medial column	33	-5.19***	.89
		Adults	lateral column	16	-6.27***	1.52
	Seg > Rhy	Both	medial column	33	-3.20**	.55
		Adults	lateral column	16	-5.29***	1.28
Late N400	Cht > Rhy	Children	Cz	16	-2.53*	.61
			C3 and C4	16	-2.75**	.67
			Pz	16	-3.43**	.83
			P3 and P4	16	-4.37**	1.06
		Adults	midline	16	-2.27*	.55
			medial column	16	-3.24**	.79
		Both	lateral column	33	-3.76**	.64
	Cht > Seg	Children	Cz	16	-3.06**	.74
			C3 and C4	16	-2.79**	.68
			Pz	16	-3.51**	.85
			P3 and P4	16	-3.00**	.73
		Adults	midline	16	-2.60*	.63
			medial column	16	-2.65*	.64
		Both	lateral column	33	-2.15*	.37

Note: Cht = Cohort Mismatch; Rhy = Rhyme mismatch; Seg = Segmental mismatch
*$p < .05$ (two-tailed) **$p < .01$ ***$p < .001$

main effect of word type in the adults [midline: $F(4,64) = 2.710$, $p = .05$, $\eta_p^2 = .15$; medial column: $F(4,64) = 3.357$, $p = .02$, $\eta_p^2 = .17$]. For both groups, pairwise comparisons (Table 4.4) revealed larger late N400 amplitudes for the cohort condition compared to the rhyme condition, and the cohort condition compared to the segmental condition. In the lateral column, there was a main effect of word-type [$F(8,128) = 3.589$, $p = .01$, $\eta_p^2 = .10$], and the same condition-wise effects were observed as in the midline and medial columns.

Overall, these results suggest that the children and adults showed similar patterns of mismatch effects: first, a larger PMN response for rhyme mismatches compared to cohort and segmental mismatches; second, larger N400 responses for cohort and segmental mismatches compared to rhyme mismatches; third, larger late N400 responses for cohort mismatches compared to segmental and rhyme mismatches. These effects are apparent in Figure 4.11.

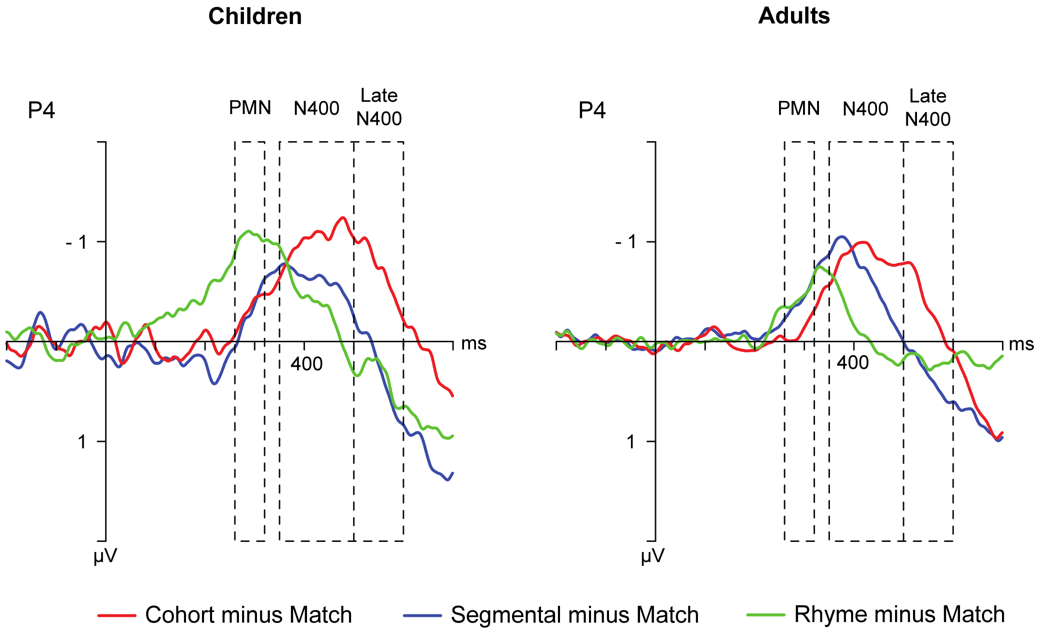


Figure 4.11: Responses to cohort, segmental, and rhyme mismatches in the children and adults. Across these three conditions, there was a dissociation in terms of which components of the waveform were modulated compared to the match condition. A similar pattern of effects was observed between the children and adults.

4.3.2.2 Hypothesis 2: Top-Down Processing is Weaker in Children Compared to Adults

As there were significant three-way interactions between group, region, and word type in the late N400 window, we focused on the rhyme condition and tested for group by region interactions in the midline and medial columns. This analysis revealed a significant group by region interaction in the medial column [$F(2,64) = 5.638, p = .006, \eta_p^2 = .15$]. A simple effects analysis in each region showed that the late N400 for rhymes was reduced in amplitude (i.e., more positive) in the adults compared to the children in frontal sites [$F(1,32) = 7.769, p = .01, \eta_p^2 = .20$]. This effect is illustrated in Figure 4.12.

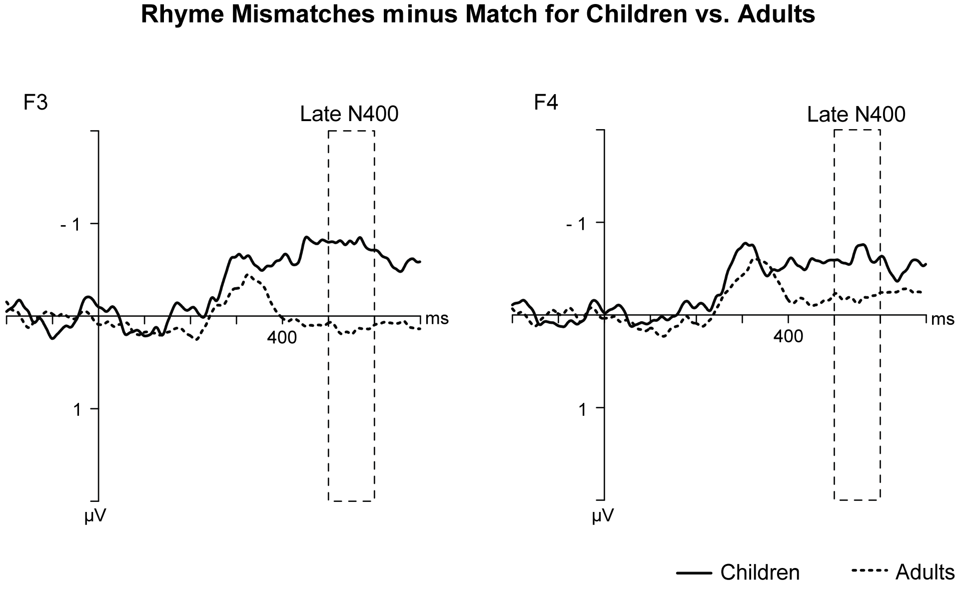


Figure 4.12: Responses to rhyme mismatches in the children and adults. The late N400 was attenuated in the adults compared to the children.

4.4 Discussion

We investigated whether Mandarin-speaking children process spoken words incrementally, and whether top-down processing influences spoken word recognition in children to the same extent that it does in adults. To do this, we compared typically developing children and adults on a task that required subjects to determine whether or not auditory words matched visual pictures (Desroches et al., 2009). Our rationale is that when viewing a particular object, a subject generates a name for it and activates its phonological form. This sets up a strong expectation for subsequent auditory input; however, during this process competitor words are also partially activated. Next, when auditory information arrives, a subject maps auditory input onto the previously activated phonological form to perform the matching judgment. This requires the subject to use incoming acoustic information to disambiguate targets from competitors. Critically, we designed stimuli such that we expected different levels of phonological and/or lexical competition based on the nature of the relationship between expected words and mismatches in the different word type conditions. As a result, our analyses focused on two components thought to be associated with phonological and lexical processing during auditory word recognition; namely, the phonological mapping negativity (PMN) response, which is associated with processing word-initial phonemes (Connolly & Phillips, 1994; Newman & Connolly, 2009), and early and late portions of the N400, a component associated with processing at the word level (Kutas & Hillyard, 1984).

4.4.1 Hypothesis 1: Both Children and Adults Process Spoken Words Incrementally

The first hypothesis concerned whether or not children process spoken words incrementally. By incremental processing, we are referring to competition amongst spoken forms that takes the temporal structure of words into account (Magnuson et al., 2003). This can be considered the opposite of holistic processing, which computes phonological similarity among spoken words in a global fashion (Luce & Pisoni, 1998). We compared responses to mismatches that diverged from expectations at different time points during the unfolding of a spoken word. Specifically, we compared responses to the cohort, rhyme, and segmental mismatch conditions in the children and adults. Rhyme mismatches (e.g., *tang2-lang2*) were signaled very early during the unfolding of spoken forms, as word-initial phonemes differed from expectations. Conversely, word-initial phonemes in the cohort (e.g., *tang2-tao2*) and segmental (e.g., *tang2-tang1*) conditions matched expectations, and therefore differences from expectations were not signaled until later on in the spoken form. In the case of the cohort condition, these were different vowels, whereas in the case of the segmental condition, this was a different tone, which is carried on the vowel cluster (Howie, 1974). We expected that if subjects process words incrementally, the PMN response, which is sensitive to pre-lexical phoneme mapping, should be more negative in the rhyme condition compared to the segmental and cohort conditions (Connolly & Phillips, 1994; Newman & Connolly, 2009; Desroches et al., 2009; Archibald & Joanisse, 2011; Desroches et al., 2013). As can be seen in Figure 4.11, this is exactly what was observed in both children and adults.

Evidence for incremental processing was also observed for both adults and children in the N400 and late N400 windows. First, the cohort and segmental conditions both showed larger N400 responses compared to the rhyme condition. We think that the size of the N400 and late N400 response is associated with the amount of indecision and/or interference a listener experiences when recognizing an auditory word, which can be considered a reflection of the extent of competition experienced from other lexical items (O’Rourke & Holcomb, 2002). A larger N400 in the cohort and segmental conditions thus suggests that individuals experienced more competition in this window, likely because this was when the mismatch was first signaled in these conditions. This is different from the rhyme condition, in which divergence from expectation was signaled earlier. As a result, competition was reduced in this window for rhyming forms, as initial divergence from expectation led to suppression of representations of competitor words that were incompatible with the spoken input.

Second, the cohort condition showed a larger late N400 amplitude compared to segmental and rhyme mismatches for both adults and children. A large late N400 amplitude in the cohort condition has been observed previously in this task in both Mandarin (Malins & Joanisse, 2012) and English (Desroches et al., 2009); it suggests that competition in this condition persisted for a longer duration of time than it did for the segmental and rhyme mismatches. The difference between cohorts and rhymes was already apparent in the N400 window; however a difference between cohort and segmental mismatches was unique to this time window. In the cohort condition, expected and presented words only differed in word-final vowels, while they differed in tone in the segmental condition. As there are thirty-six vowel clusters in Mandarin versus only four tones, it is conceivable that vowel-based competition in the cohort condition was more extensive than tone-based competition in the segmental condition (Hu, Gao, Ma, & Yao, 2012).

At any rate, as is shown in Figure 4.11, there was a dissociation in responses to cohort, rhyme, and segmental mismatches in terms of which components of the waveform were modulated. Importantly, this pattern held for both the adults and children, and can be taken as evidence of incremental processing of spoken words in both groups. This replicates prior work we have done that has offered evidence for incremental processing in English-speaking adults (Desroches et al., 2009), English-speaking children (Desroches et al., 2013; Malins et al., 2013), and Mandarin-speaking adults (Malins & Joanisse, 2012), but this is the first study to extend these results to Mandarin-speaking children.

4.4.2 Hypothesis 2: Top-Down Processing is Weaker in Children Compared to Adults

This second hypothesis concerned the influence of top-down processing on spoken word recognition in children compared to adults. As discussed previously, theories of spoken word processing differ as to whether they allow for top-down processing from the lexical to phoneme layer to result in prior activation of rhyming forms. In Indo-European languages such as English, a number of ERP studies have offered evidence for facilitated processing of rhyming compared to non-rhyming forms (Praamstra et al., 1994; Radeau et al., 1998; Coch et al., 2002; Desroches et al., 2006; Desroches et al., 2009; Desroches et al., 2013; Malins et al., 2013). In these studies, this has manifested as an attenuated (late) N400 component. The explanation for this is that because the phonemes in the rhyme receive some partial activation, they are processed more easily than word-final phonemes in non-rhyming forms, which do not benefit from this prior activation.

As is shown in Figure 4.12, the late N400 was attenuated in the adults to a greater extent than it was in the children. The result for the adults complements a prior ERP experiment using disyllabic words, in which spoken forms overlapping expectations in the second syllable showed a reduced N400 compared to words in which both characters diverged from expectations (Liu et al., 2006). However, this is the first time this type of effect has been observed for monosyllables. We have failed to observe these effects even in our own prior work employing the same methodology (Malins & Joanisse, 2012). There are a few potential reasons for this. First, there were more sets of items in the current study than in our prior work, and so the power to detect these differences was greater. Second, because we grouped electrodes into columns in the current study, we were better able to uncover this rhyme effect, which was observed only in the medial column and not the midline or lateral columns.

Importantly, the difference in rhyme processing between children and adults suggests that top-down processing does not influence spoken word recognition in children to the same extent that it does in adults. This complements earlier work using fMRI showing that top-down modulation from prefrontal areas of the brain is weaker in children compared to adults during spoken word processing tasks (Bitan et al., 2006; Bitan et al., 2009; Cao et al., 2011). However, there is an alternative explanation. It could be that the children activated rhyming forms to the same extent as adults, yet had more trouble suppressing these on the basis of bottom-up acoustic input. A consequence of this is that rhyming forms were still active in the late N400 window for the children, and so they were still experiencing competition among these forms; meanwhile, the adults had suppressed these forms and instead only experienced facilitation in processing word-final phonemes. However, we do not think this fully accounts for the effect given that children and adults did not differ in bottom-up acoustic processing of any of the other word type conditions, suggesting both groups suppressed competitors in a similar way. For example, as is shown in Figure 4.11 and reviewed in the previous section, the children experienced similar mismatch effects as the adults for the segmental and cohort conditions.

Last, it should be noted that behaviorally, children much younger than those tested in the current study show ceiling performance on tests of rhyme awareness (Siok & Fletcher, 2001; Shu et al., 2008). Therefore the present data suggest that children ten years of age differ from adults in their sensitivity to rhyming words during rapid processing, even though they do not differ from adults in overt sensitivity to rhyming relationships. This is consistent with the behavioral data reported in Table 2, as we observed similar patterns of effects for rhyming words between the adults and children, even though ERP responses to rhyming words were different between groups. This finding highlights the utility of ERP investigations in uncovering potential differences in spoken word processing between groups of subjects that are not always apparent in behavioral measures (Spivey, 2007).

4.4.3 Implications for Theories of Spoken Word Processing

These findings have important implications for theories and models of spoken word processing in Mandarin. First, they offer support for theories such as Cohort and TRACE that are based on incremental processing, and refute theories such as NAM that compute phonological similarity in a global sense. Furthermore, they suggest that this incremental processing is intact in children as young as ten years of age. Second, the rhyme effects in the adults support continuous mapping models such as TRACE that allow for top-down activation of rhyming forms. However, the lack of rhyme effects in the children suggests this top-down processing is slow to develop in Mandarin. This represents an important constraint on theories of spoken word processing in Mandarin-speaking children. We therefore recommend that TRACE is a viable starting point for a theory of Mandarin spoken word processing in both adults and children, provided that this developmental difference is taken into account.

4.5 Conclusions

We used an ERP picture-word matching task to assess how typically developing children differ from adults in how they resolve phonological competition in Mandarin. Specifically, we focused on responses to monosyllables that mismatched expectations in either onset, rime, or tonal information. First, we uncovered evidence that both children and adults process spoken words incrementally, as both groups showed sensitivity to the temporal relationship between expected word forms and the different mismatch types. Second, we observed that compared to adults, children showed less evidence of facilitated processing of rhyming forms. This suggests that in children ten years of age, top-down processing has a weaker influence on spoken word processing than it does in adults. These findings can help inform theories and models of spoken word processing by offering important developmental constraints.

4.6 References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi‐parallel progression of word structure units and cognitive operations. *Reading Research Quarterly*, *38*(4), 470-487.

Archibald, L. M., & Joanisse, M. F. (2011). Electrophysiological responses to coarticulatory and word level miscues. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1275-1291.

Bitan, T., Burman, D. D., Lu, D., Cone, N. E., Gitelman, D. R., Mesulam, M., & Booth, J. R. (2006). Weaker top–down modulation from the left inferior frontal gyrus in children. *NeuroImage*, *33*(3), 991-998.

Bitan, T., Cheon, J., Lu, D., Burman, D. D., & Booth, J. R. (2009). Developmental Increase in Top–Down and Bottom–Up Processing in a Phonological Task: An Effective Connectivity, fMRI Study. *Journal of Cognitive Neuroscience*, *21*(6), 1135-1145.

Brooks, P. J., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language*, *27*(2), 335-366.

Cao, F., Khalid, K., Lee, R., Brennan, C., Yang, Y., Li, K., Bolger, D. J., & Booth, J. R. (2011). Development of brain networks involved in spoken word processing of Mandarin Chinese. *NeuroImage*, *57*(3), 750-759.

Cheung, H., Chung, K. K., Wong, S. W., McBride‐Chang, C., Penney, T. B., & Ho, C. S. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry*, *50*(6), 726-733.

Coch, D., Grossi, G., Coffey–Corina, S., Holcomb, P. J., & Neville, H. J. (2002). A developmental investigation of ERP auditory rhyming effects. *Developmental Science*, *5*(4), 467-489.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256-266.

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, *100*(3), B32-B42.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893-1906.

Desroches, A. S., Newman, R. L., Robertson, E. K., & Joanisse, M. F. (2013). Electrophysiological indices of phonological impairments in dyslexia. *Journal of Speech, Language and Hearing Research*, *56*(1), 250-264.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Ho, C. S. H., Law, T. P. S., & Ng, P. M. (2000). The phonological deficit hypothesis in Chinese developmental dyslexia. *Reading and Writing*, *13*(1-2), 57-79.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, *18*(10), 1631-1643.

Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and vowel processing in Mandarin idioms. *Psychophysiology*, *49*(9), 1179-1190.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163.

Lei, L., Pan, J., Liu, H., McBride‑Chang, C., Li, H., Zhang, Y., & Shu, H. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *Journal of Child Psychology and Psychiatry*, *52*(2), 212-220.

Li, H., Shu, H., McBride‑Chang, C., Liu, H., & Peng, H. (2012). Chinese children's character recognition: Visuo‑orthographic, phonological processing and morphological skills. *Journal of Research in Reading*, *35*(3), 287-307.

Liu, W., Shu, H., & Yang, Y. (2009). Speech perception deficits by Chinese children with phonological dyslexia. *Journal of Experimental Child Psychology*, *103*(3), 338-354.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420.

Malins, J. G., Desroches, A. S., Robertson, E. K., Newman, R. L., Archibald, L., & Joanisse, M. F. (2013). ERPs reveal the temporal dynamics of auditory word recognition in Specific Language Impairment. *Developmental Cognitive Neuroscience*, *5*, 134-148.

Malins, J.G., Joanisse, M.F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*, 2032-2043.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71-102.

McBride‑Chang, C. (1996). Models of speech perception and phonological processing in reading. *Child Development*, *67*(4), 1836-1856.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

Modern Chinese Frequency Dictionary. (1986). Beijing: Beijing Language Institute Publisher (in Chinese).

Newman, E. H., Tardif, T., Huang, J., & Shu, H. (2011). Phonemes matter: The role of phoneme-level awareness in emergent Chinese readers. *Journal of Experimental Child Psychology*, *108*(2), 242-259.

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114-121.

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event‐related brain potentials. *Psychophysiology*, *40*(4), 640-647.

Newman, R.L., Forbes, K., & Connolly, J.F. (2012). Event-related potentials and magnetic fields associated with spoken word recognition. In M. Spivey, M. Joanisse, & K. McRae, (Eds.), *Cambridge handbook of psycholinguistics*. New York, NY: Cambridge University Press.

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology*, *60*(2), 121-150.

Praamstra, P., Meyer, A. S., & Levelt, W. J. (1994). Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component timelocked to phonological mismatch. *Journal of Cognitive Neuroscience*, *6*(3), 204-219.

Radeau, M., Besson, M., Fonteneau, E., Castro, S.L. (1998). Semantic, repetition and rime priming between spoken words: Behavioural and electrophysiological evidence. *Biological Psychology*, *48*, 183–204.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development*, *74*(1), 27-47.

Shu, H., Peng, H., & McBride‐Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science*, *11*(1), 171-181.

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology*, *37*(6), 886-899.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192-212.

Wechsler, D. (1974). Wechsler intelligence scale for children-revised. San Antonio, TX: The Psychological Corporation.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761-1770.

4.7 Appendix – Experimental Stimuli

Set	*Critical Target*	*Segmental Mismatch*	*Cohort Mismatch*	*Rhyme Mismatch*	*Tonal Mismatch*	*Unrelated Mismatch*
1	bing3 (pastry) /piŋ3/ [93]	bing1 (soldier) /piŋ1/ [700]	biao3 (watch) /piɑu3/ [1705]	ling3 (collar) /liŋ3/ [1605]	xue3 (snow) /ɕyɛ3/ [516]	hou2 (monkey) /xəu2/ [105]
2	chuang2 (bed) /tʂʰuɑŋ2/ [454]	chuang1 (window) /tʂʰuɑŋ1/ [401]	chuan2 (ship) /tʂʰuan2/ [1016]	huang2 (yellow) /xuɑŋ2/ [508]	hou2 (monkey) /xəu2/ [105]	pai1 (racket) /pʰai1/ [258]
3	deng1 (lamp) /təŋ1/ [585]	deng4 (stool) /təŋ4/ [92]	dou1 (pocket) /təu1/ [51]	feng1 (wind) /fəŋ1/ [1474]	gui1 (turtle) /kuei1/ [34]	biao3 (watch) /piɑu3/ [1705]
4	gou1 (hook) /kəu1/ [56]	gou3 (dog) /kəu3/ [245]	gen1 (roots) /kən1/ [1304]	dou1 (pocket) /təu1/ [51]	qian1 (lead) /tɕʰian1/ [83]	xin4 (envelope) /ɕin4/ [1202]
5	gui3 (ghost) /kuei3/ [276]	gui4 (wardrobe) /kuei4/ [120]	guan3 (straws) /kuan3/ [1120]	tui3 (leg) /tʰuei3/ [321]	cao3 (grass) /tsʰau3/ [803]	chao2 (nest) /tʂʰɑu2/ [32]
6	hua1 (flower) /xua1/ [1574]	hua4 (painting) /xua4/ [602]	hui1 (gray) /xuei1/ [284]	gua1 (melon) /kua1/ [174]	dao1 (knife) /tɑu1/ [363]	niu2 (cow) /niəu2/ [373]
7	jiao3 (foot) /tɕiɑu3/ [746]	jiao1 (glue) /tɕiɑu1/ [75]	jing3 (well) /tɕiŋ3/ [207]	niao3 (bird) /niɑu3/ [234]	ban3 (board) /pan3/ [589]	pan2 (disc) /pʰan2/ [321]
8	jing1 (whale) /tɕiŋ1/ [82]	jing3 (well) /tɕiŋ3/ [207]	jiao1 (glue) /tɕiɑu1/ [75]	bing1 (soldier) /piŋ1/ [700]	guo1 (pot) /kuo1/ [148]	lou2 (building) /ləu2/ [222]
9	lian2 (curtain) /lian2/ [62]	lian3 (face) /lian3/ [1036]	ling2 (bells) /liŋ2/ [113]	tian2 (field) /tʰian2/ [406]	chao2 (nest) /tʂʰɑu2/ [32]	tui3 (leg) /tʰuei3/ [321]
10	ling3 (collar) /liŋ3/ [1605]	ling2 (bells) /liŋ2/ [113]	liu3 (willow) /liəu3/ [68]	jing3 (well) /tɕiŋ3/ [207]	dao3 (island) /tɑu3/ [204]	gua1 (melon) /kua1/ [174]
11	miao2 (sprout) /miɑu2/ [107]	miao4 (temple) /miɑu4/ [60]	mian2 (cotton) /mian2/ [464]	qiao2 (bridge) /tɕʰiɑu2/ [621]	lun2 (wheel) /luən2/ [303]	feng1 (wind) /fəŋ1/ [1474]
12	pai2 (card) /pʰai2/ [102]	pai1 (racket) /pʰai1/ [258]	pan2 (disc) /pʰan2/ [321]	chai2 (firewood) /tʂʰai2/ [189]	lin2 (forest) /lin2/ [411]	jin1 (gold) /tɕin1/ [695]

Set	*Critical Target*	*Segmental Mismatch*	*Cohort Mismatch*	*Rhyme Mismatch*	*Tonal Mismatch*	*Unrelated Mismatch*
13	qian2 (money) /tɕʰian2/ [953]	qian1 (lead) /tɕʰian1/ [83]	qiu2 (ball) /tɕʰiəu2/ [550]	mian2 (cotton) /mian2/ [464]	lou2 (building) /ləu2/ [222]	gou3 (dog) /kəu3/ [245]
14	tang2 (candy) /tʰɑŋ2/ [127]	tang1 (soup) /tʰɑŋ1/ [75]	tao2 (peach) /tʰɑu2/ [100]	lang2 (wolf) /lɑŋ2/ [105]	niu2 (cow) /niəu2/ [373]	xia1 (shrimp) /ɕia1/ [54]
15	ting2 (pavilion) /tʰiŋ2/ [75]	ting1 (hall) /tʰiŋ1/ [80]	tian2 (field) /tʰian2/ [406]	ling2 (bells) /liŋ2/ [113]	mao2 (spear) /mɑu2/ [291]	hua4 (painting) /xua4/ [602]
16	xin1 (heart) /ɕin1/ [3963]	xin4 (envelope) /ɕin4/ [1202]	xia1 (shrimp) /ɕia1/ [54]	jin1 (gold) /tɕin1/ [695]	mao1 (cat) /mɑu1/ [70]	tao2 (peach) /tʰɑu2/ [100]

Note. Values in parentheses represent the morphemic frequency of each item, as indicated in *Modern Chinese Frequency Dictionary*. Transcriptions are in IPA.

5 Towards a Theory of Mandarin Spoken Word Processing

Despite the global prevalence of tonal languages such as Mandarin, current theories and models do not account for tonal languages in their current state. In this dissertation, I addressed this problem by characterizing how spoken words are processed in Mandarin Chinese, with the aim of making specific recommendations as to how current theories could be modified to include tonal languages. I used fMRI and ERP to investigate how adult native Mandarin speakers process the different components of the Mandarin syllable, focusing on the brain regions and cognitive processes involved. I subsequently used the knowledge I gained of the adult system as a platform for investigating the development of Mandarin spoken word processing in children.

In this final chapter, I first briefly summarize the important findings from the studies reported in Chapters 2 and 3. I then outline a proposed theory of Mandarin spoken word processing that is informed by these findings. Subsequently, I use this theory to explain the findings reported for the children tested in Chapter 4. Last, I close with a few considerations of future research directions.

5.1 Relevant Findings

5.1.1 The Brain Regions Involved in Representing Tones versus Vowels are Somewhat Distinct

For current theories of spoken word processing to capture spoken word processing in tonal languages, it is essential to consider how tones are represented in the brains of native speakers. To do this, I elected to compare tonal processing to vowel processing in adult native Mandarin speakers. My reasoning for this is that vowels carry tone in Mandarin syllables (Howie, 1974), and therefore vowels and tones arrive at a similar time point acoustically during the unfolding of a syllable. Importantly though, tones and vowels differ in that vowels are segmental, or phonemic, while tones are thought to be suprasegmental, as they can span multiple phonemes. Because tones are carried on vowels, I was interested in whether tonal variations are treated as a phonological feature of vowels, or whether tones are represented separately from the vowels on which they are carried.

The fMRI study reported in Chapter 2 offers evidence that the brain regions involved in representing tones versus vowels are not entirely overlapping. In this study, subjects passively listened to trains of Mandarin syllables, consisting of a repeated standard followed by a deviant that differed from the repeated standard in either tone or vowels. Analyses revealed several brain regions that differed in activity between tones and vowels in native Mandarin speakers, including right posterior middle temporal gyrus, right inferior frontal gyrus, and left insular cortex. Furthermore, these areas followed a different pattern of activity in native English speakers unfamiliar with phonetic distinctions in Mandarin. Therefore, modulation of activity in these areas can be ascribed to native language experience. These findings add to an existing body of evidence showing that the neurobiological areas involved in tonal versus segmental processing are partially dissociable (Hsieh, Gandour, Wong, & Hutchins, 2001; Gandour et al., 2003; Li et al., 2010).

Based on these findings, I argue that a successful theory of spoken word recognition in Mandarin should incorporate some processing units that are solely responsible for handling tonal information. It is my view that these tonal representations are similar in nature to phonemic representations in that they represent linguistic categories. That is, they are stored patterns associated with each of the Mandarin tones onto which listeners map acoustic input. Ye and Connine (1999) initially proposed these representations be denoted as ‘tonemes’, a term I adopt here.

Support for the idea of tonemic representations has also come from studies of categorical perception. For phonemes, this type of perception has been well-documented using a wide range of behavioral and neuroimaging measures (Joanisse, Zevin, & McCandliss, 2007; Liberman, Harris, Hoffman, & Griffith, 1957; McMurray, Tanenhaus, & Aslin, 2002). An example in English is the phoneme pair /b/ and /p/, which have the same place of articulation, yet differ in voicing (/b/ is voiced; /p/ is voiceless). Thus English listeners can use voice onset time (an acoustic correlate of voicing in stop consonants) to distinguish the two. Interestingly, when one varies VOT as a continuum between /p/ and /b/, listeners do not linearly transition between hearing /b/ and hearing /p/; rather, there is a sharp boundary at which they stop hearing /b/ and start hearing /p/.

Recently, Xi, Zhang, Xu, Zhang, and Li (2010) examined whether or not Mandarin tones are perceived categorically in an ERP experiment using the mismatch negativity (MMN). The MMN is ERP component that is modulated in response to auditory stimuli that differ in some way from a trace in echoic memory (Garrido, Kilner, Stephan, & Friston, 2009). In these types of experiments, subjects are presented with trains of repeated stimuli (called standards), punctuated by stimuli that differ from repeated items in one or more features (called deviants). The amplitude of the ensuing MMN is scaled according to how much the deviant stimulus differs from the standard. In their study, Xi et al. (2010) first varied the pitch contour of a Mandarin syllable ‘pa’ from rising to falling, and then had subjects perform an identification and discrimination task on tokens spanning this continuum. From this, the authors determined the category boundary between tone 2 and tone 4, and then selected pairs of tokens that either fell within this boundary or crossed the boundary for a subsequent MMN experiment. In this experiment, Xi et al. (2010) recorded ERPs while subjects heard a train of repeated standards of

Mandarin tones punctuated by within-category and across category deviants. They found that the MMN was larger in amplitude for across-category deviants versus within-category deviants, even though each of these deviants was the same number of steps away from the standard on the continuum. This suggests that tones are perceived in a categorical fashion, just like phonemes are, at least in Mandarin[1].

Importantly, the MMN is thought to index perceptual categorization of linguistic stimuli (Garrido et al., 2009), the same set of processes thought to be involved in the fMRI passive listening task (Zevin & McCandliss, 2005). Therefore both the Xi et al. (2010) study and the current fMRI results point to tonemic representations at the perceptual level. However, I was also interested in how these representations participate in lexical access during an active processing task.

5.1.2 Different Cognitive Processes Underlie Tonal Versus Phonemic Access

In the ERP study reported in Chapter 3, I examined how tonal versus phonemic information participate in lexical access in Mandarin. Specifically, I investigated whether the same or different cognitive processes underlie tonal versus phonemic access during the unfolding of a spoken word. I employed a picture-word matching task that required subjects to map incoming acoustic input onto word-form representations in the brain. I compared ERP responses to words differing from expectations in either tone (segmental mismatches; e.g., *tang2-tang1*) or vowels

---

[1] Some studies have shown that Cantonese tones, which differ from Mandarin tones in that they are register tones rather than contour tones, may not be perceived categorically (Francis, Ciocca, & Ng, 2003; Zheng, Minett, Peng, & Wang, 2010).

(cohort mismatches; e.g., *tang2*-*tao2*). The waveform for segmental mismatches diverged from the waveform for the match condition even earlier than did the waveform for cohort mismatches, as the phonological mapping negativity (PMN) was modulated for segmental mismatches, while it was not modulated for cohort mismatches. As the PMN is thought to index pre-lexical mapping of phonological information (Newman & Connolly, 2009), this finding suggests that tones and vowels slightly differ in the pre-lexical stage of processing. This was confirmed by an analysis of the scalp distribution of the PMN, which was more left-lateralized for the segmental versus cohort condition. While it is difficult to infer the precise brain region that gives rise to an ERP response based solely on scalp distribution (Luck, 2005), an observation of different scalp distributions between conditions suggests that the neural generators underlying tonal versus vowel access are at least partially distinct. This finding extends the fMRI results by showing that not only are tones and vowels represented somewhat separately at the perceptual level, but additionally each type of information makes an independent contribution to lexical access.

5.1.3 Spoken Words are Processed Incrementally in Mandarin

It has been suggested previously that Mandarin listeners do not show differences in the time course over which they resolve onset versus rime-based competition, and therefore process spoken words in a holistic rather than incremental fashion (Zhao, Guo, Zhou, & Shu, 2011). This is different from English, in which listeners have shown differences in the way they process onset versus rime-based similarity between spoken words (Allopenna, Magnuson, & Tanenhaus, 1998; Desroches, Joanisse, & Robertson, 2006; Desroches, Newman, & Joanisse, 2009). Therefore, there is some disagreement as to which current theory of spoken word processing is most appropriate for Mandarin (Malins & Joanisse, 2010; Zhao et al., 2011). For this reason, in the ERP study reported in Chapter 3, I also tested whether Mandarin spoken word processing is incremental or holistic. I compared responses to words sharing onset with expectations (cohort mismatches; see: *tang2* ‘candy’, hear: *tao2* ‘peach’) with responses to words sharing rime with expectations (rhyme mismatches; see: *tang2* ‘candy’, hear: *lang2* ‘peach’). Cohort mismatches modulated the late N400 but not the PMN, while rhyme mismatches modulated the PMN but not the late N400. Based on this observation, I inferred that onset and rime-based similarity do show temporally dissociable effects in Mandarin, just like they do in English. So while Zhao et al. (2011) proposed that the neighborhood activation model (NAM; Luce & Pisoni, 1998) is appropriate for Mandarin because it does not take temporal information into account when calculating phonological similarity, the current data suggest NAM is not appropriate for Mandarin. I instead argue that incremental processing models such as Cohort (Marslen-Wilson, 1987) or TRACE (McClelland & Elman, 1986) are better suited as a starting point for a theory of Mandarin spoken word processing.

5.1.4 Top-Down Information Influences Lexical Competition in Mandarin

Last, there has also been some disagreement as to whether top-down connections should be incorporated into a model of Mandarin spoken word processing. This is because theories that permit these connections predict partial activation of rhyming words when expecting to hear a particular word form, leading to lexical competition amongst rhyming forms (Allopenna et al., 1998). To date, there has been inconsistent evidence for these rhyme effects in Mandarin (Liu, Shu, & Wei, 2006; Malins & Joanisse, 2010; Zhao et al., 2011).

To address this, in Chapter 3, I compared ERP responses to non-rhyming forms (unrelated mismatches; see: *tang2* ‘candy’, hear: *xin1* ‘heart’) with responses to rhyming forms (rhyme mismatches; see: *tang2* ‘candy’, hear: *tao2* ‘peach’). Specifically, I examined whether there was an attenuation of the late N400 for rhyme compared to unrelated mismatches, as this has been observed previously in English, and has been taken as evidence for top-down connections (Desroches et al., 2009). Analyses revealed that there was no difference in late N400 amplitude between these two mismatch types. However, in Chapter 4 I repeated the same experiment with another group of subjects and included more sets of items, and analyzed the data using a slightly different approach. In this case I did find evidence for an attenuation of the late N400 for rhymes in a group of adults. This finding lends support to theories of spoken word processing such as TRACE that allow for an influence of top-down information on lexical access.

5.2 A Proposed Theory of Mandarin Spoken Word Processing

My proposed theory of Mandarin spoken word recognition is illustrated in Figure 5.1. Essentially this is a modified version of the TRACE model that includes toneme units. I elected to start with the TRACE model because it is an incremental processing model that allows for an influence of top-down information on word recognition. As in TRACE, lexical competitors inhibit one another via mutually inhibitory connections. However, the competitor set for any given spoken word includes segmentally identical words differing in tone.

As shown in Figure 5.1, if a listener expects to hear a word such as *tang2*, feedback connections from the lexical to phoneme layer allow the phonemes that comprise the word to become partially active. These then spread activation to the words that contain these phonemes, resulting in the segmentally identical words *tang1*, *tang3*, and *tang4* becoming partially active, as well as words sharing only some of these phonemes, such as *tao2* or *tao3*. In addition, expecting to hear the word *tang2* also results in the tone 2 toneme becoming partially active.

If the actual acoustic input is a segmental mismatch such as *tang1*, distinctive features of the phonemes /t^h/ /ɑ/ and /ŋ/ activate their respective phoneme units, while the distinctive features of

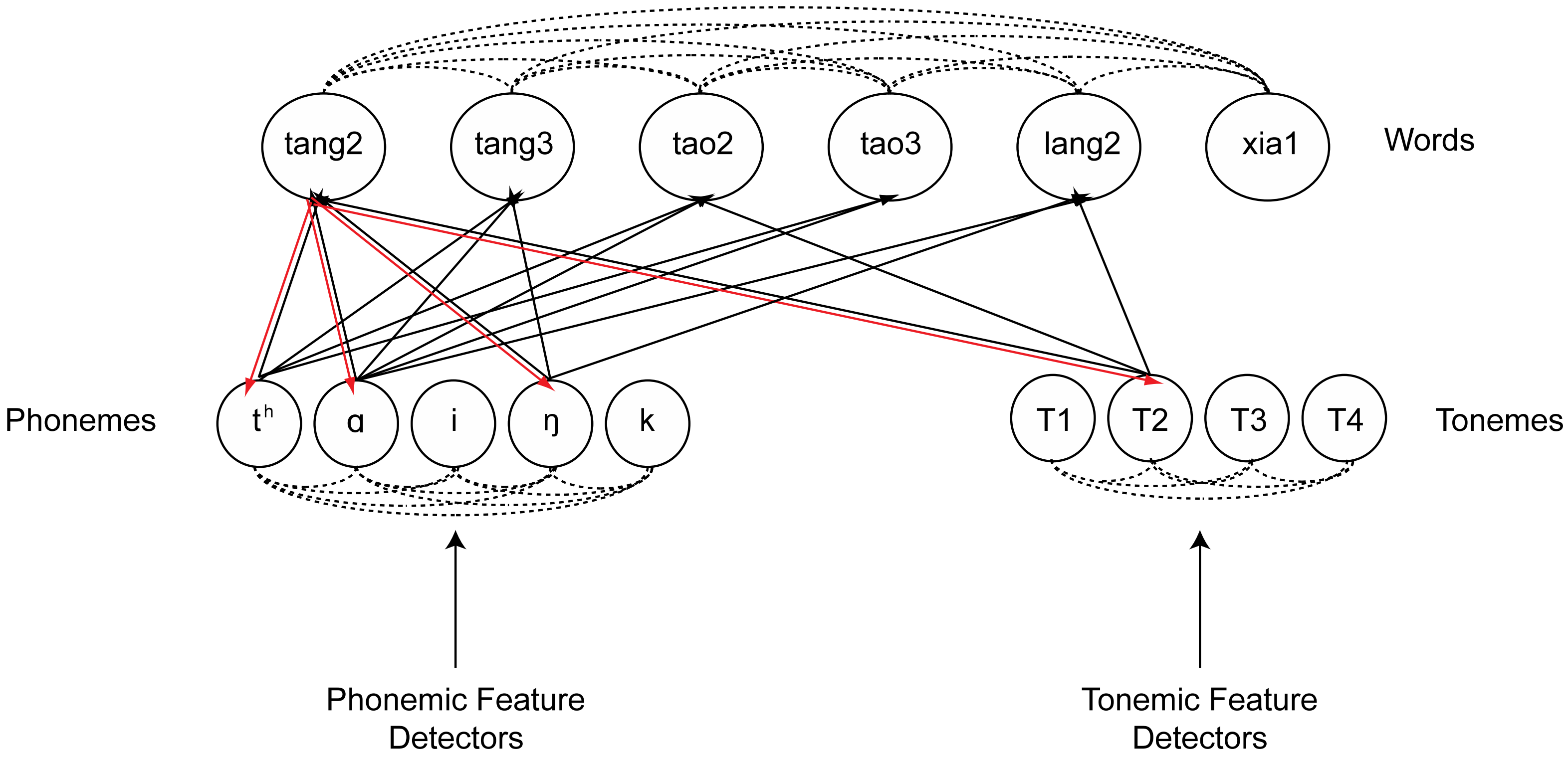


Figure 5.1: Proposed theory of Mandarin spoken word processing outlined in section 5.2. Phonemic and tonemic representations are separate from one another, and both influence word recognition. Connections within layers (dashed lines) are inhibitory, while connections between layers (solid lines) are facilitatory. Red arrows represent top-down connections from the word to phoneme level and word to toneme level. For the sake of simplicity, between-layer connections are only shown for units that are engaged when expecting to hear the word *tang2*.

the level tone activate a tone 1 toneme. Because the tone 2 toneme had received some partial activation, the tone 1 and tone 2 tonemes compete at the toneme level. Furthermore, at the word level *tang1*, *tang2*, *tang3*, and *tang4* are all partially active initially, and this activation is further bolstered when acoustic input comes in, as all of these words share the phonemic units /t^h/ /ɑ/ and /ŋ/. Therefore for a listener to correctly recognize the input, *tang1* must become more highly activated than the rest. It does this by being activated by the tone 1 toneme, as well as by inhibiting its segmental competitors at the word level.

In contrast, cohort mismatches in which tone is shared would not incur tonemic competition, as competitor words differ from one another only in vowels. Because of this lack of tonemic competition, the cognitive processes involved in resolving cohort competitors are not the same as those involved in resolving segmental competitors, explaining the difference in scalp distribution for the mismatch types observed in the ERP experiment reported in Chapter 3.

5.3 Using this Model to Understand the Development of Mandarin Spoken Word Processing

If we now turn our focus to the rhyme condition, we can interpret the data from Chapter 4. In this study, typically developing children and adults performed a picture-word matching study, and I compared ERP responses to different mismatch types across groups. I was interested in whether children would also show evidence for incremental processing, and whether top-down processing has the same influence on spoken word recognition in children as it does in adults. I found that just like adults, children ten years of age process spoken words incrementally, suggesting the theory outlined in the previous section is also viable for typically developing children. However, children and adults differed in responses to rhyming words. Namely, compared to the children, the adults showed evidence of an attenuation of competition in the late N400 window for rhyming forms. I argue this attenuation comes about for the following reason: the feedback connections from the word to phoneme level, and from the word to toneme level, result in prior activation of the phonemes and toneme that comprise a spoken word. As mentioned previously, in the case of *tang2*, this consists of the phonemes /t^h/ /ɑ/ and /ŋ/, as well as the tone 2 toneme. During the unfolding of a rhyme competitor *lang2*, word-initial mismatch signaled by the onset consonant quickly results in *lang2* inhibiting its competitors at the word level. However, the phonemes that comprise the rime, as well as the tone 2 toneme, remain partially active, and so they facilitate bottom-up recognition of word-final phonemes and tone. This results in less interference in the N400 window for rhyme mismatches compared to unrelated mismatches, for which these prior activations have to be overcome.

From the data presented in Chapter 4, it appears the top-down connections between the word level and the phoneme and toneme levels are not as strong in children ten years of age as they are in adults. This explains why the children experienced more interference than did adults in the late N400 window for rhymes. Word-final phonemes and tone did not receive as much prior activation as they did in the adults, and therefore rhyming forms were treated just like unrelated words. This offers an important constraint on the theory outlined in the previous section. I predict that these top-down connections are present in adults, but do not first emerge until adolescence. This fits with previous work using fMRI showing that top-down modulation is weaker in children compared to adults when performing rhyme judgment tasks in both English (Bitan et al., 2006; Bitan, Cheon, Lu, Burman, & Booth, 2009) and Mandarin (Cao et al., 2011). More specifically, this previous research has suggested that there is an age-related enhancement of top-down control processes from inferior frontal cortex to the left temporal regions involved in auditory processing, resulting in greater sensitivity to bottom-up information (Bitan et al., 2009).

5.4 Directions for Future Research

While some steps have been taken towards a theory of Mandarin spoken word processing, the research reported in this dissertation raises further questions and motivates future work. First, it would be useful to more adequately characterize exactly how tones are represented in the brain. Linguistic theories have long proposed that tones are processed in terms of features such as ‘contour’, ‘high’, ‘central’, ‘mid’, ‘rising’, ‘falling’, and ‘convex’ (Wang, 1967). However, it is currently unclear how these map onto auditory processing structures in the brain. Current fMRI methods such as multi-voxel pattern analysis hold particular promise for this endeavour, as this technique has been used to show that several subregions of human auditory cortex are selective to the direction of frequency modulation of sounds (Hsieh, Fillmore, Rong, Hickok, & Saberi, 2012).

Second, our understanding of the development of Mandarin spoken word processing in children is far from complete. For example, a study by Zhang et al. (2012) offered evidence that Mandarin-speaking children with reading impairment do not process tones as categorically as typically developing children do, and therefore have deficient toneme representations. It would be informative to test children with reading impairment to see if they also differ from typically developing children in phonological-lexical processing of tonal information. This is a critical aspect of reading development in Chinese that has been relatively understudied to date (Shu, Peng, & McBride-Chang, 2008).

Last, important theoretical advances could be made by implementing the theory proposed in section 5.2 in jTRACE, a Java-based program that enables simulations of spoken word recognition using the TRACE architecture (Strauss, Harris, & Magnuson, 2007). Tonemes could be implemented in several ways to determine which architecture most adequately captures the findings from the ERP study reported in Chapter 3 as well as other prior studies. The parameters of this model could also be manipulated to account for the data from the children studied in Chapter 4. Furthermore, this model could motivate future studies on Mandarin word recognition that can in turn be used to further refine the model. Together, these endeavors and others like them will help researchers develop theories of spoken word recognition that more capably account for Mandarin and other tonal languages, and therefore are more universal in nature.

5.5 References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439.

Bitan, T., Burman, D. D., Lu, D., Cone, N. E., Gitelman, D. R., Mesulam, M., & Booth, J. R. (2006). Weaker top–down modulation from the left inferior frontal gyrus in children. *NeuroImage*, *33*(3), 991-998.

Bitan, T., Cheon, J., Lu, D., Burman, D. D., & Booth, J. R. (2009). Developmental Increase in Top–Down and Bottom–Up Processing in a Phonological Task: An Effective Connectivity, fMRI Study. *Journal of Cognitive Neuroscience*, *21*(6), 1135-1145.

Cao, F., Khalid, K., Lee, R., Brennan, C., Yang, Y., Li, K., Bolger, D. J., & Booth, J. R. (2011). Development of brain networks involved in spoken word processing of Mandarin Chinese. *NeuroImage*, *57*(3), 750-759.

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, *100*(3), B32-B42.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893-1906.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029-1044.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., & Tong, Y. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*(4), 185-200.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453-463.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30* (3), 129-148.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, *76*(3), 227-252.

Hsieh, I. H., Fillmore, P., Rong, F., Hickok, G., & Saberi, K. (2012). FM-selective networks in human auditory cortex revealed using fMRI and multivariate pattern classification. *Journal of Cognitive Neuroscience*, *24*(9), 1896-1907.

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*(9), 2084-2093.

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., & Dzemidzic, M. (2010). Hemispheric asymmetries in phonological processing of tones vs. segmental units. *Neuroreport*, *21*(10), 690.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358-368.

Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context: Evidence from Chinese ERP analyses. *Brain and Language*, *96*(1), 37-48.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, *19*, 1-36.

Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge, MA: MIT Press.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71-102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33-B42.

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114-121.

Shu, H., Peng, H., & McBride‐Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science*, *11*(1), 171-181.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods*, *39*(1), 19-30.

Wang, W. S. Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, *33*(2), 93-105.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223-231.

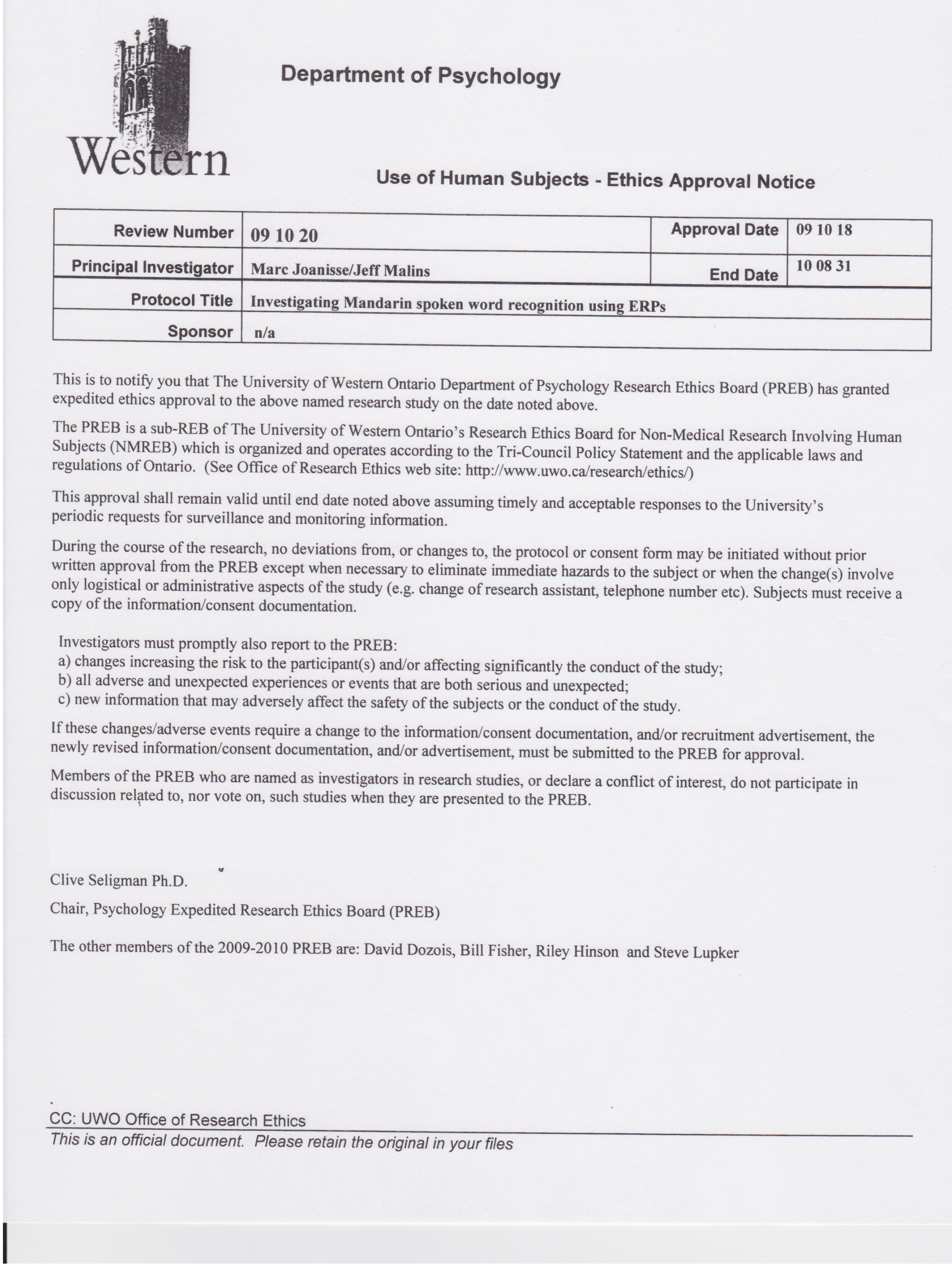
Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and* Cognitive *Processes Special Issue: Processing East Asian Languages, 14*(5-6), 609-630.

Zevin, J. D., & McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions*, *1*(4).

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761-1770.

Zheng, H. Y., Minett, J. W., Peng, G., & Wang, W. S. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes*, *27*(2), 184-209.

Appendix A: Ethics Approval



Western

Department of Psychology

Use of Human Subjects - Ethics Approval Notice

Review Number	09 10 20	**Approval Date**	09 10 18
Principal Investigator	Marc Joanisse/Jeff Malins	**End Date**	10 08 31
Protocol Title	Investigating Mandarin spoken word recognition using ERPs		
Sponsor	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: http://www.uwo.ca/research/ethics/)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:
a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

Clive Seligman Ph.D.

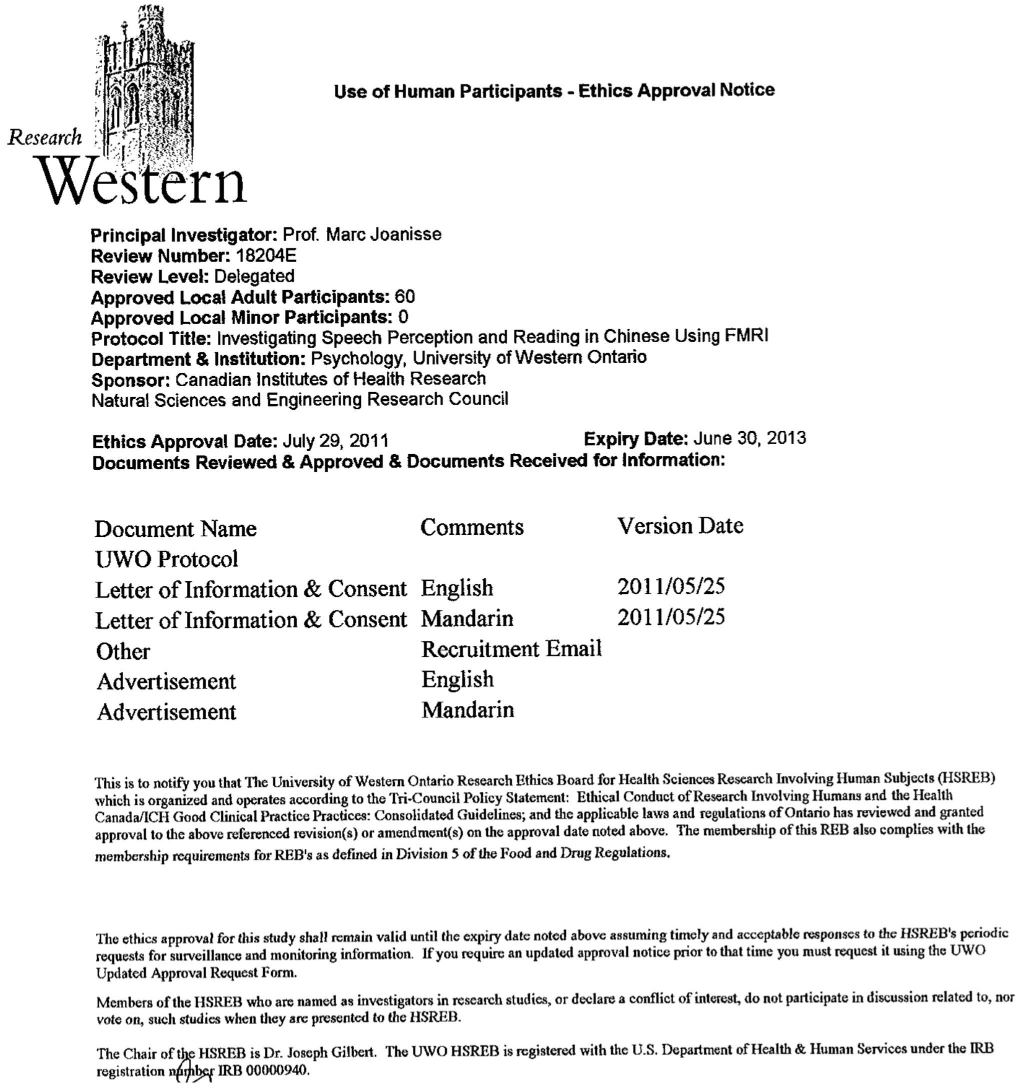
Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2009-2010 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

This is an official document. Please retain the original in your files

Appendix B: Ethics Approval



Research Western

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Prof. Marc Joanisse
Review Number: 18204E
Review Level: Delegated
Approved Local Adult Participants: 60
Approved Local Minor Participants: 0
Protocol Title: Investigating Speech Perception and Reading in Chinese Using FMRI
Department & Institution: Psychology, University of Western Ontario
Sponsor: Canadian Institutes of Health Research
Natural Sciences and Engineering Research Council

Ethics Approval Date: July 29, 2011 **Expiry Date:** June 30, 2013
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
UWO Protocol		
Letter of Information & Consent	English	2011/05/25
Letter of Information & Consent	Mandarin	2011/05/25
Other	Recruitment Email	
Advertisement	English	
Advertisement	Mandarin	

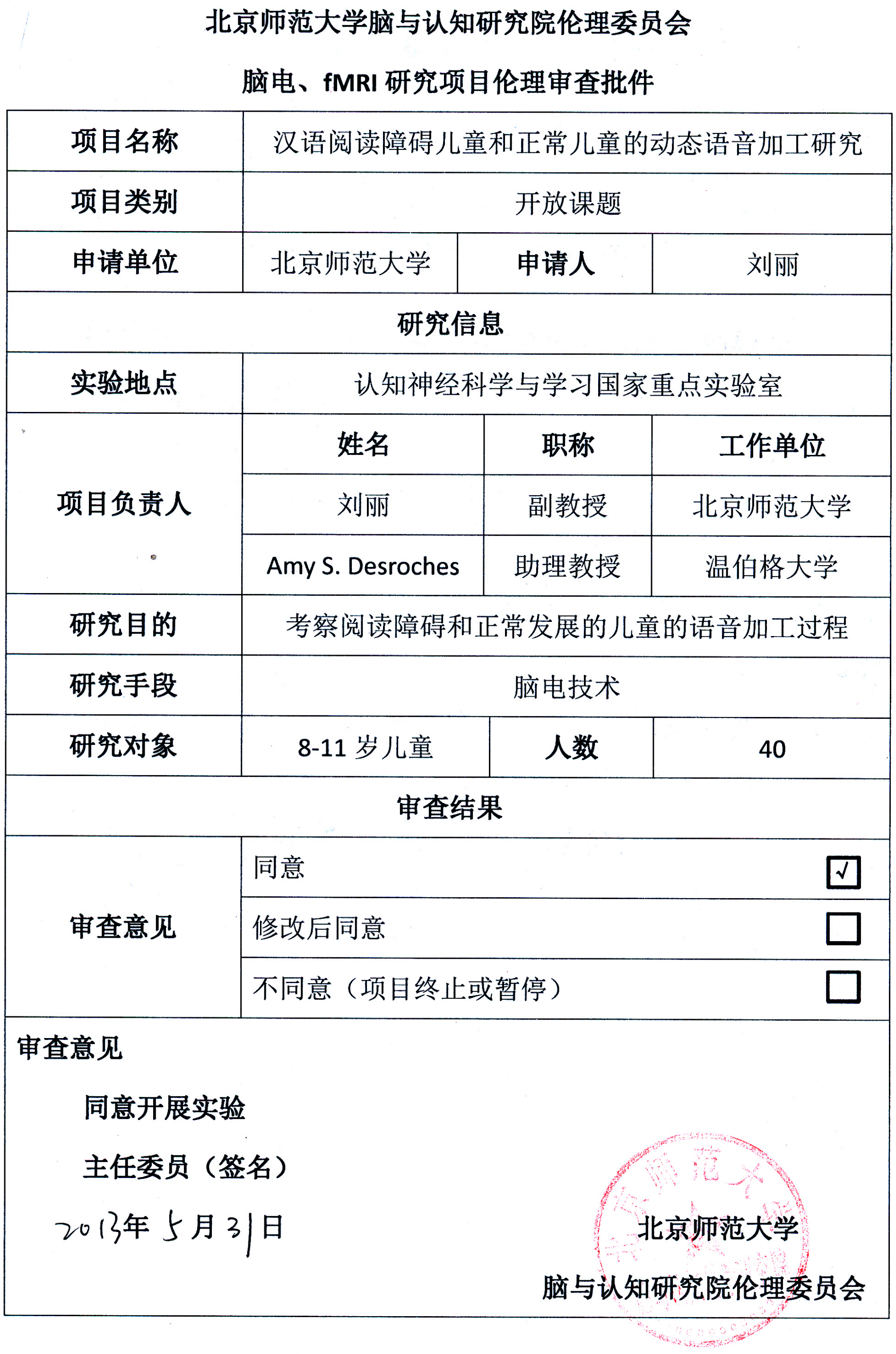
This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The UWO HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Appendix C: Ethics Approval



北京师范大学脑与认知研究院伦理委员会

脑电、fMRI 研究项目伦理审查批件

项目名称	汉语阅读障碍儿童和正常儿童的动态语音加工研究		
项目类别	开放课题		
申请单位	北京师范大学	**申请人**	刘丽
研究信息			
实验地点	认知神经科学与学习国家重点实验室		
项目负责人	**姓名**	**职称**	**工作单位**
	刘丽	副教授	北京师范大学
	Amy S. Desroches	助理教授	温伯格大学
研究目的	考察阅读障碍和正常发展的儿童的语音加工过程		
研究手段	脑电技术		
研究对象	8-11 岁儿童	**人数**	40
审查结果			
审查意见	同意	☑	
	修改后同意	□	
	不同意（项目终止或暂停）	□	

审查意见

同意开展实验

主任委员（签名）

2013年5月31日

北京师范大学

脑与认知研究院伦理委员会

Curriculum Vitae

Jeffrey G. Malins

EDUCATION

2009-2013	The University of Western Ontario, London, Canada Ph.D. Neuroscience
2007-2009	The University of Western Ontario, London, Canada M.Sc. Neuroscience
2003-2007	The University of Guelph, Guelph, Canada Honours B.Sc. Molecular Biology & Genetics

PUBLICATIONS

Malins, J. G., Desroches, A. S., Robertson, E. K., Newman, R. L., Archibald, L. M. D., & Joanisse, M. F. (2013). ERPs reveal the temporal dynamics of auditory word recognition in Specific Language Impairment. *Developmental Cognitive Neuroscience*, *5*, 134-148.

Malins, J. G. & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*(8), 2032-2043.

Malins, J. G. & Joanisse, M. F. (2012). Towards a model of tonal processing during Mandarin spoken word recognition. Invited contribution to the Third International Symposium on Tonal Aspects of Languages, Nanjing, China.

Malins, J. G. & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory & Language*, *62*(4), 407-420.

MANUSCRIPTS IN PROGRESS

Malins, J. G. & Joanisse, M. F. (in preparation). An fMRI investigation of Mandarin tonal processing using short-interval habituation.

Malins, J. G., Gao, D., Tao, R., Booth, J., Shu, H., Joanisse, M., Liu, L., & Desroches, A. (in preparation). Developmental differences in the influence of phonological similarity on spoken word processing in Mandarin Chinese.

Desroches, A. S., **Malins, J. G.**, & Joanisse, M. F. (in preparation). The time course of speech perception deficits in dyslexia: Dynamical vs. end-state measures reveal underspecified representations.

AWARDS

2010-2012	Natural Sciences and Engineering Research Council of Canada (NSERC) Canada Graduate Scholarship (Doctoral)
2012	International Phonetic Association Student Award
2011	G. Keith Humphrey Memorial Award (conferred by The University of Western Ontario)
2009-2010	Ontario Graduate Scholarship (conferred by the Ontario Ministry of Training, Colleges and Universities)
2007-2013	Western Graduate Research Scholarship
2007-2009	Natural Sciences and Engineering Research Council of Canada (NSERC) Canada Graduate Scholarship (Master's)
2007	Roche Award in Molecular Biology and Genetics
2003-2007	University of Guelph Board of Governors' Scholarship
2003-2007	James Hillier Foundation Scholarship
2005	Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award
2004	University of Guelph Dean's Scholarship

CONFERENCE PRESENTATIONS

Malins, J. G. & Joanisse, M. F. (2013, May). Processing tonal categories in Mandarin Chinese: An fMRI investigation using short-interval habituation. Poster presented at the Rovereto Workshop on Concepts, Actions, and Objects, Rovereto, Italy.

Malins, J. G. & Joanisse, M. F. (2012, June). An fMRI investigation of Mandarin tonal processing using short-interval habituation. Poster presented at the 18th Annual Meeting of the Organization for Human Brain Mapping, Beijing, China.

Malins, J. G., Gao, D., Tao, R., Booth, J., Shu, H., Liu, L., & Desroches, A. (2012, June). Online markers of phonological processing in typically developing Mandarin-speaking children. Poster presented at the 18th Annual Meeting of the Organization for Human Brain Mapping, Beijing, China.

Gao, D., **Malins, J. G.**, Tao, R., Booth, J., Shu, H., Liu, L., & Desroches, A. (2012, June). The relationship between online and offline measures of phonology in Chinese children: An ERP study. Poster presented at the 18th Annual Meeting of the Organization for Human Brain Mapping, Beijing, China.

Malins, J. G. & Joanisse, M. F. (2012, June). Towards a model of tonal processing during Mandarin spoken word recognition. Invited talk given at the National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China.

Malins, J. G. & Joanisse, M. F. (2012, May). Towards a model of tonal processing during Mandarin spoken word recognition. Talk given at the Third International Symposium on Tonal Aspects of Languages, Nanjing, China.

Malins, J. G. & Joanisse, M. F. (2011, May). An ERP investigation of tonal versus phonemic influences on spoken word recognition in Mandarin Chinese. Poster presented at the 31st Annual Meeting of the Southern Ontario Neuroscience Association, Guelph, ON.

Malins, J. G. & Joanisse, M. F. (2011, April). An ERP investigation of tonal versus phonemic influences on spoken word recognition in Mandarin Chinese. Poster presented at the 18th Annual Meeting of the Cognitive Neuroscience Society, San Francisco, CA.

Malins, J. G. & Joanisse, M. F. (2009, November). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. Poster presented at the 50th Annual Meeting of the Psychonomic Society, Boston, MA.

Desroches, A. S., Robertson, E. K., **Malins, J. G.**, Newman, R. L., Archibald, L. M. D., & Joanisse, M. F. (2009, April). ERPs reveal the temporal dynamics of phonological processing deficits in language impairment. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Denver, CO.

Malins, J. G. & Joanisse, M. F. (2008, October). Lexical tone processing in Mandarin Chinese: An eyetracking investigation. Talk given at the 3rd Annual Toronto Workshop on East Asian Languages, Toronto, ON.

RELEVANT WORK EXPERIENCE

2007-2013	Graduate Teaching Assistant The University of Western Ontario, London, Canada
2005	Research Assistant McMaster University, Hamilton, Canada

SOCIETY MEMBERSHIPS

Cognitive Neuroscience Society

REVIEWER

Journal of Memory and Language
Language and Cognitive Processes
Developmental Neuropsychology