ELECTROPHYSIOLOGICAL EVIDENCE FOR THE INTEGRAL NATURE OF TONE IN MANDARIN SPOKEN WORD RECOGNITION

ELECTROPHYSIOLOGICAL EVIDENCE FOR THE INTEGRAL NATURE OF TONE IN MANDARIN SPOKEN WORD RECOGNITION

By AMANDA HO, B.A. (Hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University MASTER OF SCIENCE (2015) Hamilton, Ontario (Cognitive Science of Language)

TITLE: Electrophysiological evidence for the integral nature of tone in Mandarin spoken word recognition

AUTHOR: Amanda Ho, B.A. (Hons.) (McMaster University)

SUPERVISORS: Dr. John F. Connolly and Dr. Anna L. Moro

NUMBER OF PAGES: ix, 72

Abstract

Current models of spoken word recognition have been predominantly based on studies of Indo-European languages. As a result, little is known about the recognition processes involved in the perception of tonal languages (e.g., Mandarin Chinese), and the role of lexical tone in speech perception. One view is that tonal languages are processed phonologically through individual segments, while another view is that they are processed lexically as a whole. Moreover, a recent study claimed to be the first to discover an early phonological processing stage in Mandarin (Huang et al., 2014). There seems to be a lack of investigations concerning tonal languages, as no clear conclusions have been made about the nature of tonal processes, or a model of spoken word recognition that best incorporates lexical tone. The current study addressed these issues by presenting 18 native Mandarin speakers with aural sentences with medial target words, which either matched or mismatched the preceding visually presented sentences with medial target words (e.g, 家 /jia1/ "home"). Violation conditions involved target words that differed in the following ways: *tone violation*, where only the tone was different (e.g., 价 /jia4/ "price"), *onset violation*, where only the onset was different (e.g., 虾 /xia1/ "shrimp"), and *syllable violation*, where both the tone and the onset were different (e.g., 糖 /tang2/ "candy"). We did not find evidence for an early phonological processing stage in Mandarin. Instead, our findings indicate that Mandarin syllables are processed incrementally through phonological segments and that lexical tone is strongly associated with semantic access. These results are discussed with respect to modifications for

existing models in spoken word recognition to incorporate the processes involved with tonal language recognition.

Acknowledgements

This thesis would not have been possible without the efforts of many. I am deeply indebted to my principal supervisor, Dr. John F. Connolly for his invaluable guidance and thoughtful perspectives on various aspects of this research. My appreciation for his contributions is beyond my words. I am truly grateful for my second supervisor, Dr. Anna L. Moro, for her warm support and valued insight on the formation of my research question and stimuli. My Mandarin instructor, Ms. Jun Wu (吴老师), played a significant role in creating sentential contexts for my target words, and I am incredibly thankful for her help. I specifically thank Daniel Schmidtke and Dr. Elisabet Service for their involvement with my statistical analyses. I would also like to thank Christine Zhang, Jing Wen, and Lisa Lin for recording my stimuli, and Kai Fan for reviewing my stimuli for native accuracy.

I would like to extend my gratitude to the Language Memory and Brain Lab (LMBLab) for dedicating their time and effort in assisting me with collecting data and running participants. Of the LMBLab, Rober Boshra deserves a special mention for his much-appreciated assistance in programming my experiment and sincere reassurances. Further, I thank Richard Mah for kindly offering his wisdom in electroencephalography (EEG).

Diane Doran, Zoë Wälchli, Heather Stephens, Narcisse Torshizi, Kyle Ruiter, Edalat Shekari, Cassandra Chapman, Jitka Bartosova, Connie Imbault, and Samantha Kramer proved to be the most endearing friends and colleagues during the time I spent at McMaster University. I would also like to recognize Kristi Kwan for her unconditional

friendship and continued emotional support, and Jeffrey Choi for his ingenious humour when I endured personal difficulties during the final span of this project. Last but not least, I would like to thank my father (爸爸) for his warm encouragements throughout his ongoing battle with epilepsy.

To conclude, I would like to dedicate this thesis to my grandma (嫲嫲), aunt (二姑媽), cousin (莊表姐), and cousin-in-law (Allan 哥哥) for always believing that I can surmount every obstacle I am faced with.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1.0 Spoken word recognition models: An overview

Human speech perception is a fast and transient cognitive process. Spoken language proceeds at a rate of two to three words per second, where meaning unfolds as acoustic information is continuously mapped in the brain (Zhao, Guo, Zhou, & Shu, 2011). Lexical representations are thus activated through the integration of phonological and semantic information provided by the acoustic cues in the speech signal (Zou, Desroches, Liu, Xia, & Shu, 2012). It has been suggested that the recognition time for polysyllabic content words is related to the point in the speech stream where phonological information is congruent with a single lexical candidate (Tyler, 1984). Others have suggested that the recognition time for spoken language is affected by single phoneme differences in word recognition (Goldinger, Luce, & Pisoni, 1989; Luce, Pisoni, & Goldinger, 1990). Generally, most models propose that lexical candidates compete for recognition during speech perception, which involves multiple underlying cognitive processes. Various models have been proposed to account for this acoustic-phonetic interaction in spoken words, but different claims have been made about the level of phonological involvement in the activation of potential lexical representations. Speech recognition is commonly explained through top-down and/or bottom-up information processing. Top-down processing involves conceptually driven perceptions, while bottom-up processing involves information based on the incoming signal. Spoken word recognition models explore the extent to which recognition is processed in a context dependent (top-down) or unidirectional (bottom-up) manner.

One description of spoken word recognition is the Cohort model, which emphasizes the "left-to-right" nature of speech recognition over time (Marslen-Wilson & Tyler, 1980). This model assumes that listeners eliminate word forms from a set of potential lexical candidates that are defined on the basis of word-initial bottom-up input (Marslen-Wilson & Tyler, 1980). Listeners compile potential *cohorts* based on initial phoneme perception to further activate syntactic and semantic information from the speech stream (Marslen-Wilson, 1987). This suggests that lexical candidates compete for activation as a function of phonological similarity, which is perceived sequentially. According to this account, as the acoustic signal unfolds and more information becomes available, initially activated alternatives are suppressed until only one candidate remains (Marslen-Wilson, 1987). Thus, words sharing word-initial information (e.g., *cap, cat, cab, catch, captain*) compete for phonological recognition in lexical decision tasks. However, because this model is dependent on word-initial information only, there is no phonological competition among rhyming words (e.g., *bad, lad, had, mad*) due to their differences in onset information, and as a result, the recognition of rhyming words would be facilitated as compared to words sharing word-initial information (Marslen-Wilson & Tyler, 1980). Because recognition is dependent on bottom-up input, context plays no role in the process of form-based access and selection. Instead, contextual information is only integrated at higher-levels of representations concerning the syntactic and semantic properties of selected cohorts (Marslen-Wilson, 1987).

Although it is important to consider bottom-up input, top-down influences of speech perception have been incorporated into subsequent models of spoken word

recognition. It is important to note that the Cohort model was initially influenced only by bottom-up information, but over the years, it had developed to also incorporate top-down influences. For instance, the evolved version of the Cohort model (distributed connectionist model) employed a representation of speech perception that combined lexical information with abstract phonological information, in which the output of phonological representations were also a product of stored semantic knowledge (Gaskell & Marslen-Wilson, 1997). This has been demonstrated in priming studies concerning semantic and referential aspects of linguistic communication associated with morphemes (Marslen-Wilson & Tyler, 2007). Likewise, the Neighbourhood Activation Model (NAM) (Luce & Pisoni, 1998), Shortlist/MERGE (Norris, 1994; Norris & McQueen, 2008; Norris, McQueen & Cutler, 2000), and TRACE (McClelland & Elman, 1986) incorporate top-down influences. These models allow for phonological competition among rimes due to the contextual influences that supply activation to word units from higher levels of processing beyond word-initial segmental recognition. The NAM, established by Luce & Pisoni (1998), explored the structural relations among sound patterns of words in memory and their effects on speech perception. They examined words belonging to "similarity neighbourhoods," defined as a collection of words that are phonetically similar to a given stimulus word (Luce & Pisoni, 1998). Similarity neighbourhoods incorporate *neighbourhood density*, which refers to the total number of words occurring in the neighbourhood, and *neighbourhood frequency*, which refers to how often the word appear in the neighbourhood (Luce & Pisoni, 1998). They found that the number and nature of words in a similarity neighbourhood affected the speed and

accuracy of word recognition (Luce & Pisoni, 1998). Reaction time data demonstrates that words occurring in low density neighbourhoods were processed faster than words in high density neighbourhoods, indicating that the neighbourhood structure of words in the mental lexicon strongly affects spoken word recognition (Luce & Pisoni, 1998). In addition, the NAM describes the effects of similarity neighbourhood structures on the process of discriminating acoustic-phonetic lexical representations, where phonological neighbours or words compete for activation (Luce & Pisoni, 1998). Acoustic-phonetic forms of novel words and non-words are activated at different perceptual levels that correspond to phonetically relevant acoustic differences among word decision units (Luce & Pisoni, 1998). Higher-level lexical information is subsequently integrated with acoustic-phonetic information for word recognition (Luce & Pisoni, 1998). Nonetheless, the NAM does not account for the temporal dynamics of spoken word recognition as competition among words is based on the overall phonological similarity of potential lexical candidates. According to the NAM, words that differ by only one phoneme (e.g., *rap, tap, sap, lap*) compete for recognition. Similarity neighbourhoods are primarily based on global similarity since evidence supporting the NAM is derived from recognition time based on the frequency and sum of its neighbours (Luce & Pisoni, 1998). Thus, the more neighbours a word has, the more frequently those neighbours occur, and the harder it is to recognize the word. This only reveals the impact of overall recognition, but does not provide information about how neighbourhood effects change as the word unfolds over time.

Both Shortlist/Merge (Norris, McQueen & Cutler, 2000) and TRACE (McClelland & Elman, 1986) account for the continuous mapping of acoustic-phonetic information as the stimulus unfolds over time, as competition results from the lateral inhibition between lexical candidates. According to Shortlist/Merge, information flows from pre-lexical processes to the lexicon without feedback, as information is processed in one direction from sounds to words (Norris et al., 2000). Top-down influences from lexical knowledge to phonological processing occur subsequently at the post-lexical decision stage (Norris et al., 2000). Because decision processes require both lexical and phonemic sources of information, Shortlist/Merge is able to explain the varying task demands on facilitatory and inhibitory effects in word and non-word recognition. For example, Shortlist/Merge is able to explain inhibitory effects of competition in non-words with subcategorical mismatches (Marslen-Wilson & Warren, 1994), facilitatory effects in non-words that are more like real words (McQueen et al., 1999), and the lack of inhibitory effects in non-words that deviate from real words (Connine et al., 1997). While phonemic decisions in both words and non-words are based on serial inputs from bottom-up connections, competition effects occur due to phonetic similarity at any point in a word (Norris et al., 2000). In contrast to the Cohort model, Shortlist/Merge allows for competition among non-cohorts, which allows for a broader competitor set of phonemes. Even though Shorlist/Merge also only has bottom-up connections between phonemes and lexical levels of representation, it emphasizes the influence of lexical knowledge on phonological processing at the decision stage, providing an account for competition among rimes and other neighbours.

Similarly, TRACE assumes that recognizing a word involves activating its distinctive word-form representation based on serial acoustic-phonetic inputs, but it also emphasizes the temporal and dynamic nature of speech (McClelland & Elman, 1986). TRACE is based on the principles of interactive activation, where information is processed through units that are excitatory or inhibitory depending on the activations of other connecting units (McClelland & Elman, 1986). This model allows for different types of lexical competition as it integrates multiple sources of information or constraints in speech perception. According to this interactive-activation approach, information is processed through *units* (McClelland & Elman, 1986). Throughout the course of processing, activation is updated on the basis of the input while each unit is continually receiving information from other units (McClelland & Elman, 1986). Consequently, a "trace" is formed by the system's working memory that processes these perceptual cues by segmenting continuous speech streams into words (McClelland & Elman, 1986). As a result, a continuous two-way flow of lexical and phonemic information is processed through feedback connections, where acoustic information is processed bottom-up and lexical information is processed top-down simultaneously during word recognition. Due to the bottom-up and top-down interactions between lexical candidates in lexical selection, TRACE predicts that cohorts are processed earlier than rimes during competition of spoken word recognition (McClelland & Elman, 1986).

1.2.0 Studies on spoken word recognition

Earlier studies have examined behavioural responses in spoken word recognition. In a gating paradigm, Grosjean (1980) found that words sharing initial segments were activated together. The gating paradigm involved the repeated presentation of spoken stimuli such that the duration of its fragments was successively increased until the entire word had been presented (Grosjean, 1980). After each presentation, subjects were asked to propose the word being presented and to give a confidence rating of their proposed word (Grosjean, 1980). In addition, Marslen-Wilson & Zwitserlood (1989) used prime words that differed only in their first segment from the semantically related visual probe in a cross-modal priming paradigm. The results showed that irrespective of the amount of overlap between onset information with the original words or non-words, priming with rime information did not facilitate word recognition (Marslen-Wilson & Zwitserlood, 1989). This evidence supports the claim that words sharing initial segments are activated together during spoken word recognition. Grosjean (1980) found similar results in a gating task, where listeners were presented with successively longer fragments of words. Subjects were asked to propose the word being presented and to give a confidence rating after each segment (Grosjean, 1980). Based on accuracy and confidence judgments during the recognition of polysyllabic content words, multiple candidates were activated until speech input was consistent with only one lexical candidate (Grosjean, 1980). In addition, Zwitserlood (1989) examined the spoken word recognition of German in a visual priming paradigm, where a "priming word" was presented to the subject followed by a "target word." The subject subsequently judged whether the target word was a word

or a non-word. Zwitserlood (1989) found that a short fragment (e.g., *kapi*) of an auditory prime (e.g., *kapitein*, Dutch for *captain*) facilitates lexical decision to targets with the actual prime (e.g., *boot*, Dutch for *boat*), and to targets that only share onset information (e.g., *kapitaal*, Dutch for *capital*). This suggests that regardless of its context of occurrence, an initial input produces early multiple activation in the lexicon (Zwitserlood, 1989).

These studies support the Cohort model of spoken word recognition, but this model, like all other models, originates from behavioural experimentation, which fails to capture the dynamic nature of speech perception. In addition, criticisms of this model have been offered. It has been pointed out that it is difficult to identify which part of the word is considered the cohort, since word onsets in continuous speech are not always clearly segmented (Norris, 1994). Moreover, partial onset matches are not accounted for in this model, which is problematic when perceiving words in noisy environments (McClelland & Elman, 1986). To gain a detailed understanding of the nature of language processing, Swinney (1981) suggested that we examine the microstructure of the entire process as it occurs in real time. Outcomes from behavioural experimentation are only end-state characteristics of underlying language structures. For that reason, Zwitserlood, (1989) proposed that we employ an experimental paradigm that can provide an on-line and continuous measure of language performance.

The "visual world paradigm" in eye tracking is an on-line method in monitoring spoken language processing in real time, where participants are presented with both visual and auditory stimuli. In a study by Magnuson, & Tanenhaus (1998), participants

followed spoken instructions about the objects presented on the screen during eye tracking. The objects consisted of a referent (e.g., beaker), a cohort (e.g., beetle), a rhyme (e.g., speaker), and an unrelated object (e.g., carriage). Contrary to the predictions of the Cohort model, competition was observed for both cohort and rhyme competitors as participants fixated on rhyme competitors significantly more than unrelated items. This demonstrates that regardless of whether potential lexical candidates share the same onset, both cohorts and rhymes compete for lexical activation. Participants were also more likely to fixate on cohort competitors over rhyme competitors, suggesting that speech input is mapped onto potential lexical representations as it unfolds over time (Allopenna et al., 1998). Magnuson, Tanenhaus, Aslin, & Aslin (2003) replicated this study with pseudowords that differed in their final phoneme (e.g., /pibo/ and /pibu/) or their initial phoneme (e.g., /pibo/ and dibo/). Similar cohort and rime competition patterns were found, showing that the time course of processing novel words was closely related to that of real words.

1.2.1 Electrophysiological responses to spoken word recognition

Although eye tracking has been successful in revealing that both cohorts and rhymes compete during recognition, electroencephalography (EEG) can provide further insights into our understanding of the underlying processes involved in spoken word recognition. EEG has high temporal accuracy in measuring speech processes as they unfold in real time. Evoked potentials (EPs) and event-related potentials (ERPs) are derived from EEG activity that is sensitive to a range of sensory and cognitive processes,

respectively (Kutas & Federmeier, 2011). ERPs reflect the sum of postsynaptic neuronal activity recorded at the scalp as small voltage fluctuations in the electroencephalogram in response to a stimulus event that elicits a broadly-defined cognitive process (Friederici, Pfeifer & Hahne, 1993). These responses reflect millisecond-level processing in the brain, as EEG is a continuous measure of neural activity (Kutas & Federmeier, 2011). In addition, ERP components are associated with particular cognitive functions (e.g., word recognition, semantic memory) consisting of a series of positive and negative voltage deflections. Typically, ERP components are referred to by their positive- or negative-going polarity followed by a number indicating the latency in milliseconds post-stimulus onset. For instance, a negative-going waveform that peaks at about 100 ms post-stimulus onset is called the N100.

A central focus in the ERP literature on language is the N400, which is a negative-going component that typically peaks at 400 ms post-stimulus onset (Kutas & Hillyard, 1980). EEG studies on word recognition have examined the nature and influence of phonological levels of structure, such as syllables, onsets and rimes, and phonemes, on the representation and processing of morphology (Kutas & Federmeier, 2011). For instance, words, pseudowords (e.g., GORP), and illegal strings (e.g., NLK) unrelated to expected words were found to elicit an N400 response when presented in a sentence comprehension task (Laszlo & Federmeier, 2009). With regard to contextual information, Kutas & Hillyard (1980) observed an N400 when sentences ended in semantically inappropriate words. Similarly, Holcomb (1993) examined electrophysiological responses to words and pseudowords that were either semantically related or unrelated to

prime words in a lexical decision task. A semantic priming effect was found, as subjects responded faster and more accurately to semantically related than to unrelated prime words. Thus, the N400 reflected the ease with which semantic meaning is integrated into the broader sentential context, where larger N400 amplitudes were elicited when the information was incongruous with the discourse representation (Osterhaut & Holcomb, 1992). The overall evidence suggests that the N400 is sensitive to contextual constraints and semantic associations in language processing.

ERP studies on spoken word processing have also explored the integration of early semantic processing at a level where lexical and contextual information interact (Kutas & Van Petten, 1998, 1994; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Initially, Connolly, Stewart, & Phillips (1990) found an "N200-like" response that differed in topographical distributions as the N400 response when subjects listened to sentences that varied in the degree to which the context predicted the terminal word of the sentence. Later, Connolly, Phillips, Stewart, & Brake (1992) explored both the "N200" and N400 component in high constraint (high Cloze probability) and low constraint (low Cloze probability) sentences. Cloze probabilities determine whether sentence contexts are predictive of the last word of the sentence (Taylor, 1953). Connolly et al. (1992) found a functional separation between the two components, where the "N200" was sensitive to acoustic/phonological information, and the N400 was sensitive to cognitive/linguistic information. It was hypothesized that this "N200" reflected an earlier processing stage that preceded the semantic integration of the stimulus (Connolly & Phillips, 1994). To test this ERP component preceding the N400, Connolly & Phillips

(1994) conducted an experiment using highly constrained sentences (high Cloze probability) that ended with a semantically congruent or anomalous word. Terminal words of sentences were categorized into one of four experimental conditions: (1) Phoneme Match-Semantic Match condition, which consisted of a highest Cloze probability word (e.g., "The piano was out of *tune*"), (2) Phoneme Mismatch-Semantic Match condition, which consisted of an initial phoneme that differed from that of the highest Cloze probability word, but was semantically appropriate (e.g., "Don caught the ball with his *glove*"), (3) Phoneme Match-Semantic Mismatch condition, which consisted of an initial phoneme with the same initial phoneme as the highest Cloze probability word, but was semantically inappropriate (e.g., "The gambler had a bad stream of bad *luggage*"), and (4) Phoneme Mismatch-Semantic Mismatch condition, which consisted of semantically anomalous words with an unexpected initial phoneme (e.g., "The dog chased our cat up the *queen*"). The Phoneme Mismatch-Semantic Match condition did not elicit an N400, but instead, only an early negativity (peaking between 270-300 ms) was elicited. The Phoneme Match-Semantic Mismatch condition did not elicit an early negativity, but instead, only the N400 was elicited. However, both the early negativity and the N400 were elicited in the Phoneme Mismatch-Semantic Mismatch condition, whereas neither the early negativity nor the N400 was elicited in the Phoneme Match-Semantic Match condition. These results demonstrated that the early response and the N400 were indeed two components that have a level of independence from each other (Connolly & Phillips, 1994). Therefore, Connolly & Phillips (1994) proposed that this early negativity (later labelled the Phonological Mapping Negativity or PMN, see

Newman & Connolly, 2009) was sensitive to the phonological features of words that reflected the phonological discrepancy in the speech signal between initial phonemes of the expected word and that of the perceived word, independent of semantic appropriateness. Further, their results seemed to support the notion of matching the incoming signal with a phonemic template of an expected word. In a later study, Connolly, Service, D'Arcy, Kujala, & Alho (2001) explored the PMN component by presenting participants with a visual word/non-word (e.g., House/Telk), followed by a prime letter (e.g., M), and were instructed to anticipate the word/non-word by replacing the first letter of the word with the prime letter, which in this case, would be "Mouse" for the word and "Melk" for the non-word. Participants were either presented with a matched (e.g., Mouse/Melk) or mismatched (e.g., Barn) stimulus (Connolly et al., 2001). The PMN was elicited for both words and non-words in the mismatch condition (Connolly et al., 2001). Because no significant amplitude difference in the PMN was observed between word and non-word conditions, it further supported the proposal that the PMN is pre-lexical and independent of contextual influences derived from top-down lexical selection processes (Connolly et al., 2001). Connolly & Phillips (1994) interpreted the PMN within the framework of the Cohort model, which is initially based on sensory input only, as the PMN was sensitive to conditions where contextual expectancies for the initial phoneme of target words were violated. In addition, Connolly & Phillips (1994) found that word selection occurred in the 175-225 ms latency range, which also corresponded with the word selection time found in behavioural studies supporting the Cohort model (Marslen-Wilson, 1987; Grosjean, 1980). However,

Desroches, Newman & Joanisse (2009) conducted a similar study that examined the temporal dynamics of different types of phonological competition, but interpreted the PMN within the framework of Shortlist/MERGE. The rationale behind this conclusion was that phonological expectations were established through visual pictures rather than auditory input, which induced bottom-up and top-down phonological expectations through connections between the lexical level and the phoneme level.

The findings of Connolly & Phillips (1994) were replicated by Van Den Brink, Brown, & Hagoort (2001). Van Den Brink et al. (2001) investigated the time course of contextual influences on spoken word recognition in Dutch. This study used semantically constraining Dutch sentences with sentence-final words that differed across three conditions to examine the moment at which context begins to have an effect on word recognition. The fully congruent (FC) condition ended with the highest-cloze probability word (e.g., "*De schilder kleurde de details in met een klein penseel*": "The painter coloured the details with a small paint *brush*"). The initially congruent (IC) condition ended with an anomalous word that began with the same initial phonemes as the highest-cloze probability word (e.g., *De schilder kleurde de details in met een klein pensioen*: "The painter coloured the details with a small *pension*"). Alternatively, the fully incongruent (FI) condition ended with an anomalous word that began with different initial phonemes as the highest-cloze probability word (e.g., *De schilder kleurde de details in met een klein doolhof*: "The painter coloured the details with a small *labyrinth*"). Although an N400 response was observed in the IC and FI conditions, the N200 (what these authors chose to call the PMN) was also elicited in the FI condition, which preceded

that of the FC and IC conditions. Because this N200 component was elicited in all conditions irrespective of whether the phonemes corresponded to the initial phonemes of the highest-cloze probability word, Van Den Brink et al. (2001) concluded that the N200 component was different from the PMN component found in Connolly & Phillips (1994). Van Den Brink (2001) argued that the N200 effect reflected post-lexical processing that was consistent with the semantic goodness-of-fit of when the initial phonological analysis and semantic information of a word interact. However, later studies on the spoken word recognition of non-words have indicated that the PMN component indeed reflects pre-lexical processing (Newman et al., 2003; Kujala, Alho, Service, Ilmoniemi, & Connolly, 2004). Both Newman et al. (2003) and Kujala et al. (2004) found that PMN amplitudes were not dependent on whether the heard item was a word or non-word in a phoneme-deletion task, suggesting that the PMN was elicited in the absence of lexical semantic access. Therefore, these results reject the proposal of Van Den Brink et al. (2001) that the negativity preceding the N400 reflects a post-lexical processing stage involving top-down contextual influences.

1.3 Spoken word recognition in Chinese

To date, spoken word recognition models have been highly influenced by studies on Indo-European languages, with very little work done using tonal languages concerning the influence of tone on word recognition processes. As a result, existing theories in spoken word recognition only account for a fraction of the world's languages. Tonal languages, such as Mandarin Chinese, differ from Indo-European languages in their

morphosyllabic and segmental structure. Indo-European languages can have multiple syllables and morphemes forming a meaningful word, whereas there is a one-to-one correspondence between meaningful syllables and morphemes in tonal languages (Zhao, Guo, Zhou, & Shu, 2011). For that reason, the syllable level for tonal speakers may be more central in word perception. Further, the pronunciation of a syllable in tonal languages requires segmental (onset and rime) and supra-segmental (tone) information (Zhao et al., 2011). Tone is characterized by differences in the fundamental frequency (F_0), which distinguishes between two lexical items with the same onset and rime (Ho & Bryant, 1997). For instance, Mandarin has four lexical tones that vary in pitch: high level (Tone 1), high rising (Tone 2), falling rising (Tone 3), and high falling (Tone 4) (Howie, 1976). Therefore, the syllable /ma/ can either mean mother (/ma1/ “妈”), hemp (/ma2/ "麻”), horse (/ma3/ “马”), or the verb to scold (/ma4/ “骂”) depending on the associated lexical tone (Wong, 2002). Therefore, tone conveys both lexical and phonological information. Current models of spoken word recognition do not sufficiently address how both segmental and supra-segmental information constrain word recognition in tonal speakers, as they mainly focus on the recognition of segmental information only.

1.3.1 Behavioural studies on the lexical access of Chinese

The lexical access of Chinese was first considered in reading studies. Because Chinese uses a logographic writing system instead of an alphabetic writing system, it was originally assumed that Chinese readers access word meanings without the use of phonology (Biederman & Tsao, 1979). However, Perfetti & Zhang (1995) found that

phonological information was activated as part of Chinese character identification. In their first experiment, Perfetti & Zhang (1995) designed a semantic judgment task (see below), where critical trials contained homophones (two words that sound the same but have different meanings and orthographic representations) and synonyms (two words with similar meanings but do not sound the same). A meaning-without-phonology model would predict semantic interference with homophone judgments but not phonological interference with semantic judgments. On the other hand, a meaning-with-phonology model would instead predict that interference would occur in both situations. The results showed an increase in reaction time and error rate for both homophonic and synonym pairs, which demonstrates a two-way interference in judgments of Chinese words (Perfetti & Zhang, 1995). Thus, meaning similarity interferes with the judgment that two characters have different pronunciations, while phonological similarity interferes with the judgment that two characters have different meanings (Perfetti & Zhang 1995). This study indicated that phonological processes are an automatic part of Chinese reading that accompanies lexical processing. The second experiment explored whether there was a weaker form of the meaning-without phonology hypothesis, where phonological information was not automatically part of word identification, but required time to accumulate. This would suggest that the results from the first experiment arose because of a long stimulus onset asynchrony (SOA), which provided adequate time for phonological activation to accumulate and interfere with semantic judgments. The SOA denotes the amount of time between the start of one stimulus and the start of the next stimulus (Perfetti & Zhang, 1995). SOA trials in this experiment were 90, 140, and 260

ms. The results showed that the name of a character was activated for Chinese readers within 90 ms of word processing, suggesting that readers activated both semantic and phonological information within the shortest SOA (Perfetti & Zhang, 1995). In other words, at the earliest processing point manipulation, Chinese readers were unable to bypass phonology to access the meaning of a word.

Similarly, Tan, Hoosain, & Siok (1996) examined the time course of phonological and semantic activation in Chinese in a backward visual masking task established by Perfetti, Bell, & Delaney (1988). In this paradigm, the target word is presented visually before a word or pseudoword mask and a pattern mask. By manipulating the presentation time of the masking stimuli, the time course of graphemic, phonemic, and semantic information can be revealed (Perfetti et al., 1988). In the first experiment, the mask type and target exposure of homophonic and semantically similar target-mask pairs were manipulated (Perfetti et al., 1988). Visually similar masks facilitated target identification at 43 ms, but significant effects were absent for homophonic and semantic masks (Perfetti et al., 1988). Nevertheless, when the target was presented 14 ms above threshold level, semantic masks did not affect target processing, but visually similar and homophonic masks produced significant facilitation effects (Perfetti et al., 1988). As phonological interference was observed in absence of semantic interference, these findings suggest that phonology is accessed earlier than semantics in Chinese visual word recognition.

Moreover, to investigate whether phonological information facilitates lexical access in Chinese, Tan & Perfetti (1997) conducted an experiment using the "phonologically mediated priming" (PMP) (Lesch & Pollatsek, 1993) paradigm. In a

PMP paradigm, a target word (e.g., /shang3/ 尚 "even") is visually presented after a synonym prime that is synonymous with the target word (e.g., /you2/ 犹 "still"), a homophone of the synonym prime (e.g, /you2/ 邮 "post"), and an unrelated prime (e.g., /gu1/ 古 "ancient"). The SOA was varied at 129, 243, and 500 ms. In the 129 and 243 ms conditions, response times were affected by the homophone density of the prime words, where homophone density was defined as the amount of competing homophones for the prime word (Tan & Perfetti, 1997). Word identification was facilitated for low-density homophones but not for high-density homophones. However, at the 500 ms condition, only synonyms facilitated word identification (Tan & Perfetti, 1997). According to this study, there was a two-way activation from orthographic and phonological information to semantic interpretation, where phonological information was accessed in character meaning at 243 ms (Tan & Perfetti, 1997). It is possible that phonological activation was more difficult with high-density homophone primes because lexical tone was a predominant competitor in word recognition.

1.3.2 On-line studies of spoken word recognition in Chinese

Recently, neurolinguistics research has moved from an English/Indo-European language focus to include other language systems, most notably tonal languages. For instance, Brown-Schmidt & Canseco-Gonzalez (2004) employed ERPs to study Mandarin speakers while they listened to normal and anomalous sentences in Mandarin – a language with four (4) tones. The tone, syllable, or both tone and syllable of sentence-final monosyllabic words were manipulated to create semantically implausible sentences

in a sentence-comprehension task (Brown-Schmidt & Canseco-Gonzalez, 2004). Control sentences ended with an expected word with a plausible tone (e.g., "At the theatre, I ate popcorn and *candy*" – /*tang2*/). In the tone condition, sentences ended with an expected syllable, but an implausible tone (e.g., "At the theatre, I ate popcorn and *government office*" – /tang1/). In the syllable condition, sentences ended with an unexpected syllable, but a plausible tone (e.g., "At the theatre, I ate popcorn and *people*" – /*ren2*/). The final condition contained a double-anomaly, where sentences ended with an implausible syllable and tone (e.g., "At the theatre, I ate popcorn and *political party*" – /*dong3*/). Tone, syllable and double-anomalous sentences elicited an N400 response, however, the syllable condition elicited the earliest and strongest negativities between 200 and 250 ms, which were interpreted as an "early N400 effect" (Brown-Schmidt & Canseco-Gonzalez, 2004). Nonetheless, these effects are too early for an N400 response. Negativities in this time window more likely indicate the presence of a PMN. It was concluded that the syllable unit was more anticipated in production and comprehension in Chinese than the tone unit, when identifying word meaning and inconsistencies in sentential semantic contexts (Brown-Schmidt & Canseco-Gonzalez, 2004). This research failed to consider other aspects of the syllable, such as the onset and rime. Thus, further research considering the differential nature of syllables as compared to tones in Chinese is needed to better understand how each is processed.

Other work has investigated tonal and segmental information in Cantonese spoken word processing – a tonal language with six (6) tones (Schirmer, Tang, Penney, Gunter & Chen, 2005). They combined each sentence (e.g., Ah Ming doesn't feel well; therefore

he sees the doctor to treat his __ and then applies for sick leave to recover) with either a semantically congruous word (e.g., /beng6/ 病 "illness") or a semantically incongruous word. In the semantically incongruous conditions, the word either differed in tone (/beng2/ 餅 "biscuit"), segmental (/bou6/ 步 "step"), or both tonal and segmental (e.g., /gwai3/ 季 "season") information. Incongruous conditions elicited an increased frontal negativity at 300 ms post-stimulus onset, an effect interpreted as an "N400-like effect" (Schirmer et al., 2005). When the word was in complete violation of both tonal and segmental information, an ERP effect occurred 100 ms earlier than the other semantically incongruous conditions (Schirmer et al., 2005). The authors concluded that the time course and amplitude of the N400-like negativity were comparable between tonal and segmental violations, suggesting that segmental and supra-segmental information were accessed simultaneously in Cantonese (Schirmer et al., 2005). Consequently, the perception of segmental information in tonal languages may be more strongly connected with tonal information as compared to Indo-European languages (Schirmer et al., 2005). Nevertheless, since the rime was the only segmental manipulation in the study, the results do not adequately explain the constraints on the activation of phonologically similar words.

Using eye tracking, Malins & Joanisse (2010) made similar conclusions about the time course of segmental and supra-segmental activation. In a visual world paradigm, where participants were instructed to identify the perceived auditory word from an array of pictures on a computer screen, Malins & Joanisse (2010) examined how tonal versus segmental information influenced spoken word recognition in Mandarin. The target word

(/chuang2/ “bed”) had a segmental (/chuang1/ “window”), cohort (/chuan2/ “ship”), rime (/huang2/ “yellow”) and tonal (/niu2/ “cow”) phonological competitor. The results revealed that both the cohort and segmental conditions showed slower looks to the target word as compared to baseline, suggesting that segmental and tonal information were accessed in parallel (Malins & Joanisse, 2010). In addition, the rime condition showed no effect on the time course of looks to the target word, which may indicate that initial phonemes play a more significant role in constraining word recognition in Mandarin (Malins & Joanisse, 2010). These results fit with TRACE (McClelland & Elman, 1986), as the target word provided bottom-up phonological information, which activated possible lexical candidates for phonological competition. Malins & Joanisse (2010) suggested modifying TRACE to include inhibitory connections from morphemes to tonal cues, creating a more diverse model of spoken word recognition (Malins & Joanisse, 2010).

The effect of contextual constraints in Mandarin through the manipulation of syllables in disyllabic words was studied to investigate the effect of phonological mismatch between syllables (experiment 1) and onsets within a syllable (experiment 2) (Liu, Shu, & Wei, 2006). In the first experiment, a spoken sentence frame (e.g., “*The sound in the radio became weaker and weaker. It seems that I must buy several new sets of* ...”) ended with a word that formed a semantically congruous (e.g, /*dian4*/-/*chi2*/ 电池 “*battery*”) or incongruous ending. There were three types of incongruous endings: (1) the “cohort incongruous condition,” in which only the cohort syllable was the same (e.g., /*dian4*/-/*lu2*/ 电脑 “*computer*”); (2) the “rhyme incongruous condition,” in which only the

rhyme syllable was the same (e.g., /*shui3*/-/*chi2*/, 水池 "*water pool*"); and, (3) the "plain incongruous condition," in which both syllables were different (e.g., /*bing4*/-/*tai4*/ 疾病 "*illness*"). Relative to the plain incongruous condition, where both cohort and rhyme syllables were mismatched, the N400 effect appeared earlier when only the cohort syllable was anomalous. Since the final syllable was preserved when an earlier response was observed, these results were consistent with TRACE (McClelland & Elman, 1986). The final syllable may have provided a context that permitted lexical candidates that do not begin with the same segments to become activated. This suggested that there was also feedback from the lexical level to the morphosyllabic level. In the second experiment, high or low constraint sentences ended with a minimal-onset-mismatch non-word, a maximal-onset-mismatch non-word, or a first-syllable-mismatch non-word (Liu et al., 2006). The results showed that the "N400 effect" was elicited earlier in the high constraint condition (200-300 ms window) as compared to the low constraint condition (300-400 ms) (Liu et al., 2006). The "N400" was also elicited earlier in the maximal-onset-mismatch and first-syllable mismatch condition (200-300 ms) than in the minimal-onset-mismatch condition (300-400 ms). These findings replicate those of Connolly & Phillips (1994), possibly suggesting that instead of an "N400 effect," a PMN response was elicited in the maximal-onset-mismatch and first-syllable mismatch conditions within the 200-300 ms time window. Liu et al. (2006) concluded that at least in highly constrained contexts, phonetic and semantic processes are accessed together.

Instead of disyllabic words, Zhao, Guo, Zhou, & Shu (2011) studied phonological competition in Mandarin monosyllabic words using a novel picture/spoken-word/picture

task. This is a method of priming, which is defined as a facilitation of the response to a test item (i.e., a word or a picture) by a preceding item (i.e., prime) (Ratcliff & McKoon, 1981). Other works have used a visual/picture/spoken-word matching paradigm to reveal interactions between top-down and bottom-up processes during the time course of auditory word recognition (see Connolly, Byrne, & Dywan, 1995; Deroches, Newman, & Joanisse, 2008). In Zhao et al's (2011) novel picture/spoken-word/picture paradigm, participants were instructed to compare whether two pictures belonged to the same semantic category. The spoken word in between the two pictures either matched (e.g., picture: /bi2/ 'nose'; spoken: /bi2/ 'nose') or mismatched the initial picture. There were four different mismatch conditions: onset mismatch (e.g., picture: /bi2/ 'nose'; spoken: /li2/ 'pear'), rime mismatch (e.g., picture: /bi2/ 'nose'; spoken: /bo2/ 'shoulder'), tone mismatch (e.g., picture: /bi2/ 'nose'; spoken: /bi3/ 'pen'), or syllable mismatch (e.g., picture: /bi2/ 'nose'; spoken: /ge1/ 'brother'). The results showed that between 400 and 500 ms, all mismatch conditions elicited larger N400 effects than the match condition (Zhao et al., 2011). The syllable mismatch condition elicited an earlier and stronger N400 effect than the match condition (300-400 ms). However, the other three partial mismatch conditions (onset, rime, and tone) did not elicit larger N400 effects than the match condition in this early time window. In fact, the onset, rime, and tone mismatch conditions elicited comparable N400 amplitudes that did not significantly differ from each other (Zhao et al., 2011). These results were in line with the NAM since words that differ from each other by only one phonological feature or unit compete for activation (Luce & Pisoni, 1998). However, these results also extended beyond the NAM, as

competition was also observed at the morphosyllabic level. The authors concluded that syllable level processing is central in tonal language processing because whole-syllable violations elicited earlier and stronger N400 effects as compared to partial-syllable violations (Zhao et al., 2011). These results may be due to the one-to-one correspondence in the morphosyllabic structure of Chinese, where Chinese speakers perceive syllables as one holistic unit rather than multiple segmented units. Similar to Malins & Joanisse (2010), Zhao et al. (2011) suggested a modified spoken word recognition model based on TRACE that incorporates an added syllable morpheme level and "toneme" nodes to account for lexical tones perception.

All of the studies on the spoken word recognition of Chinese thus far, have only observed N400 effects. However, these studies have not examined components earlier than the N400, so it remains unknown whether tonal information also influences earlier pre-lexical processes. To date, only two studies have examined both the N400 and PMN components in the spoken word recognition of tonal languages. Most recently, this issue was examined using an auditory word-matching paradigm, where both disyllabic prime and target words were presented aurally (Huang, Yang, Zhang, & Guo, 2014). The experimental stimuli consisted of a cohort (/ge2*bi*4/–/ge2shi4/; "the next door"–"format") condition, an identical (/ge2shi4/–/ge2shi4/; "format'–"format") condition, and an unrelated (/*ran*2*hou*4/–/ge2shi4/; "then"–"format") condition. A significant P200 effect was found for the unrelated condition, which is described as an earlier phonological processing stage, indicating that tonal and segmental representations for target words were accessed from the beginning (Huang et al., 2014). Because of the overlap with the

preceding P200 effect, the unrelated condition also had attenuated PMN amplitudes as compared to the cohort and identical conditions (Huang et al., 2014). In addition, a larger early-N400 effect was detected for the cohort and identical conditions. Huang et al. (2014) speculated the reason why a smaller early-N400 component was observed for the unrelated condition was because disambiguation was unnecessary based on the second syllable since the first syllable of the target word already differed from the prime. Finally, a late-N400 effect was observed in the cohort and unrelated conditions, suggesting that phonology had influenced later semantic processing (Huang et al., 2014). Nevertheless, Huang et al. (2014) did not relate these results to current spoken word recognition models. Although there was evidence for cohort competition, rhyme competition was not a control in this experiment, so it is difficult to confirm evidence for models outside of the Cohort model.

In a study that controlled for rhyme competitors, Malins & Joanisse (2012) manipulated phonological information of Chinese monosyllables in a picture/spoken word paradigm (see Connolly et al., 1995). Picture cues primed phonological expectations, which were subsequently confirmed or violated by a spoken word. There were five possible competitors: (1) segmental (e.g. picture: /hua1/ "flower"; sound: /hua*4*/ "painting"); (2) cohort (e.g., picture: /hua1/ 'flower'; sound: /h*ui*1/ 'gray'); (3) rhyme (e.g., picture: /hua1/ 'flower'; sound: /*g*ua1/ 'melon'); (4) tonal (e.g., picture: /hua1/ 'flower'; sound: /*jing*1/ 'whale'); (5) unrelated (e.g., picture: /hua1/ 'flower'; sound: /*lang2*/ 'wolf'). The results indicated that the rhyme (onset mismatch) condition and the unrelated (total mismatch) condition elicited a larger PMN than the cohort condition

(rime mismatch). Yet, the cohort condition elicited a later N400 component that was more negative as compared to the rhyme (cohort mismatch) condition, meaning that word-initial overlap creates the greatest competition among lexical items (Malins & Joanisse, 2012). These findings are analogous to the phonological and semantic processing of English (c.f., Connolly & Phillips, 1994). However, the segmental (tone mismatch) condition elicited larger PMN amplitudes than the cohort condition, suggesting that tonal information is accessed as soon as it becomes available. In addition, the degree of overlap with expectations did not produce gradations of the N400 component, suggesting that total mismatches were perceived similarly to onset, rime, and tonal mismatches (Malins & Joanisse, 2012). It has been demonstrated that the N400 is sensitive to certain semantic approximations. For example, a reduced N400 effect was observed when the phrase "The pizza was too hot to…" ended with the word '*drink*,' as compared to the word '*cry,*' when the expected word was '*eat'* (Kutas & Van Petten, 1994). This reflected the lexical priming of '*drink*' by the semantic context of food consumption. These results contradict those found by Zhao et al. (2011), as whole-syllable mismatch effects did not differ from the effects of individual components of syllables. Malins & Joanisse (2012) deduced that tonal and phonemic information were accessed immediately, and that Mandarin syllables were processed incrementally. These findings support the Cohort and TRACE models, but along with Zhao et al. (2011), Malins & Joanisse (2012) also suggested that tonal feature detectors be included in a working model of spoken word recognition to account for tonal perception.

These studies seem to suggest that the N400 response in Chinese follows the all-or-none principle, where a response is independent of the strength of the stimulus. An all-or-none response was found with the PMN, which was a result of overall phonological relatedness (Newman, Connolly, Service, & McIvor, 2003). Newman et al. (2003) employed a phoneme-deletion task requiring participants to segment an auditory prime into its constituent phonemes (e.g., "*clap*" without the /*k*/ sound), where incorrect answers varied in the degree to which they matched the correct answer. Equal PMN amplitudes were elicited for both mismatches involving an initial phoneme and mismatches involving several subsequent phonemes. Thus, it was suggested that the PMN was an all-or-none response independent of the degree of phonological similarity between expectations and heard stimuli (Newman et al., 2003). Since Kutas & Van Petten (1994) found varying degrees of the N400 amplitude to violations that varied in semantic relatedness, perhaps a model that explains lower and higher levels of word processing is necessary to understand spoken word recognition. The lower level of word processing would be sensitive to phonological violations that respond in an all-or-none manner, whereas the higher level of word processing would be sensitive to more nuanced responses related to semantics.

2.0 The present study

Evidently, there is a lack of consensus concerning the time course of spoken word recognition in Chinese. Some studies indicate that Chinese recognition is incremental and support TRACE with an added tonal component (e.g., Malins & Joanisse, 2010; 2012; Liu et al., 2006), whereas other studies believe that Chinese recognition is based on

whole-syllable perception (e.g., Brown-Schmidt & Canseco-Gonzalez, 2004; Schirmer et al., 2005; Zhao et al., 2011). There is also a lack of consistency with segmental manipulations, as some studies manipulated segments within monosyllabic words (e.g., Brown-Schmidt & Canseco-Gonzalez, 2004; Schirmer et al., 2005; Malins & Joanisse, 2010; 2012; Zhao et al., 2011), while others manipulated syllables within disyllabic words (e.g., Liu et al., 2006; Huang et al., 2014). Moreover, some studies incorporated whole sentences (e.g., Brown-Schmidt & Canseco-Gonzalez, 2004; Shirmer et al., 2005; Liu et al., 2006), while others looked at target words without a sentential context (Zhao et al., 2011; Malins & Joanisse, 2010; 2012; Huang et al., 2014). The amount of variability in the results on the temporal processes involved in the spoken word recognition of Chinese warrants further research, as there are still no clear conclusions on tonal language perception.

The present study was designed to investigate the following issues: (1) the effect of the P200 and its interaction with the PMN; (2) the relationship between the "early N400" and the PMN in spoken word recognition in Mandarin; (3) whether tonal languages are processed lexically as a whole, or phonologically through individual segments; and (4) how tonal languages fit into spoken word recognition models. To do this, we examined monosyllabic target words in sentence-medial position since sentence-final words induce "wrap-up" effects, which reflect global processing in regards to the overall decision and response requirements, causing large positive ERP amplitudes (Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984; Van Petten & Kutas, 1990; Connolly & Phillips, 1994; Kutas & Federmeier, 2011; Swaab, Ledoux, Camblin,

& Boudewyn, 2011). Sentence-final words that were deemed unacceptable have been found to elicit an enhanced N400-like effect regardless of whether the violation was semantic (negative-going response) or syntactic (positive-going response) in nature (Hagoort, Wassenaar, & Brown, 2003). This N400-like effect was presumably related the overall integration of sentential information into one complete message. According to Malins & Joanisse (2012), an increasingly larger PMN response when comparing onset and tone violation conditions would provide evidence for incremental processing. However, as indicated by Zhao et al. (2011), if the syllable violation condition elicit the largest and earliest N400 amplitudes as compared to the other individual segmental violations, then the results would support holistic processing. Thus far, the study by Huang et al. (2014) is the only study that has reported the P200 phonological mismatch effect during spoken word recognition in Chinese. The current study will explore the robustness of this component and whether it affects the PMN. Our findings should help us gain more insight on how tonal languages fit into spoken word recognition models.

3.0 Methods

3.1 Participants

Eighteen right-handed university students (7 males) between 18 and 23 years of age (mean = 20.4; SD = 1.33) were recruited through the McMaster Linguistics Research Participation System and through advertisement posters. All participants were native speakers of Mandarin Chinese from China and had been living in Canada for an average of 2.44 years (SD = 1.50). They were screened for proficiency in Mandarin to establish

eligibility in the study (See Appendix B; Mandarin screening form). The purpose of this screening was to ensure all participants were literate in Chinese and to determine their amount of exposure to other tonal languages, as these factors could affect the results. All participants were typically developed with no reported speech or hearing deficits, and were not taking any medication at the time of testing (See Appendix C; Demographic screening form). Participants were granted experimental credit for their participation. An informed consent approved by the McMaster Research Ethics Board (See Appendix A; Consent form) was completed prior to experimentation. See Appendix D for the participant debriefing form.

3.2 Stimuli and experimental conditions

Two hundred and forty-eight experimental sentences in Mandarin Chinese were composed with the help of a female native speaker of Mandarin from the Beijing Language and Culture University. Target words were selected from the Modern Chinese Frequency dictionary (Wang, Chang, & Li, 1985) with a mean cumulative frequency in percentile[1] 90.09 (SD = 2.0). Since sentence-final words are often strong attractors of global processing factors and wrap-up effects (Hagoort & Brown, 1999), target words were positioned within the subject phrase of a predicational sentence, which consisted of a subject phrase followed by a verb phrase. In the *baseline condition*, the tone and word-initial segment were identical to the target word. Only the tone of the target word was mismatched in the *tone violation condition*, whereas only the word-initial segment of the

[1] Column 4 of the Modern Chinese Frequency dictionary (Wang et al., 1985).

target word was mismatched in the *onset violation condition*. Both the tone and word-initial segment of the target word were mismatched in the *syllable violation condition*. Sample sentences of each condition are illustrated in Table 1, where the number indicates the tone with which target words are pronounced.

Condition	**Target word**	**Sentence**	**Translation**
Baseline (no violation)	家 /jia1/ 'home'	我朋友的**家**很明亮。	My friend's **home** is very bright.
Tone Violation	价 /jia4/ 'price'	我朋友的**价**很明亮	My friend's **price** is very bright.
Onset violation	虾 /xia1/ 'shrimp'	我朋友的**虾**很明亮。	My friend's **shrimp** is very bright.
Syllable violation	糖 /tang2/ 'candy'	我朋友的**糖**很明亮。	My friend's **candy** is very bright

Table 1: Examples of target stimuli

Each experimental condition consisted of 62 sentences with a mean character total of 10.13 (SD = 1.18) and a mean duration of 2795.2 ms (SD = 411.77) for target words. Care was taken to preserve the lexical category of mismatched target words to only monosyllabic noun phrases. In addition, 248 filler sentences of similar syntactic structure to the experimental sentences were constructed with a mean character total of 9.72 (SD = .99) and a mean duration of 2289.70 ms (SD = 264.67). Three native female speakers of Mandarin recorded the speech tokens to eliminate experimental effects attributed to individual speaker idiosyncrasies. These speech tokens were recorded in a sound-attenuated room using Audacity software (Mazzoni & Dannenberg, 2000) with an Audio-Technica ATM73A head-mounted microphone, attached to a desktop computer using a

TASCAM US-122MKII USB 2.0 Audio/MIDI interface. Speech was recorded at a sampling rate of 44.1 K Hz. For consistency across recordings, each sentence was repeated three times and the most natural sample was selected for further manipulation. All stimuli were volume normalized to -10 dB and noise reduced at -24 dB. In sum, 496 sentences (62 baseline, 62 tone violation, 62 onset violation, 62 syllable violation, and 248 filler) were presented to each participant, in which the stimulus order was randomized.

3.3 Procedure

Testing took place at the Language Memory and Brain Lab at McMaster University. The procedure lasted approximately 2 hours. Auditory stimuli were delivered binaurally using earphones (Etymotic Research) and an amplifier (ARTcessories HeadAmp4). Both visual and auditory stimuli were presented using Presentation (NeuroBehavoiuralSystems Presentation 14.7). Prime sentences (see below) were presented in Microsoft YaHei font (size 52) were displayed in white characters on a black background through a 24-inch Hewlett-Packard (HP) computer monitor positioned 1m away. Participants were instructed to decide whether the visually presented prime sentence matched or mismatched the following spoken sentence by indicating their decision using a response pad (Cedrus Model RB-830). Participants pressed the left green button for "match" with their left index finger and the right green button for "mismatch" with their right index finger. In order to control reading rate and ocular artifacts, a rapid serial visual presentation (RSVP) protocol was used in this study.

Earlier work has shown that RSVP reading is quite natural as sentences can be read accurately at rates up to 12 words per second (Potter, 1984). A trial began with the presentation of a fixation cross for 375 ms. The fixation cross was then replaced by characters each displayed for 325 ms. Immediately following the sentence prime, another fixation cross appeared and a spoken sentence was played through the earphones. The next trial began after the participant indicated a match or a mismatch (see Figure 1).

In order to create an expectation for the following spoken sentences, visually presented sentences were always presented with the appropriate tone, onset, and semantic characteristics as if they were in the baseline condition. Participants were instructed to fixate on the cross at the center of the screen at all times while minimizing head movements and eye blinks during the experiment. Subjects performed five practice trials before the test trials began to become familiar with the procedure; these trials were not included in the analysis. Twelve 10 s breaks (every 31 sentences) and 3 self-timed breaks (every 124 sentences) were provided to reduce effects of fatigue. As this experiment was lengthy, requiring considerable focus and attention, the self-timed breaks allowed participants to choose when they were ready to continue, helping them stay alert throughout the entire experiment.

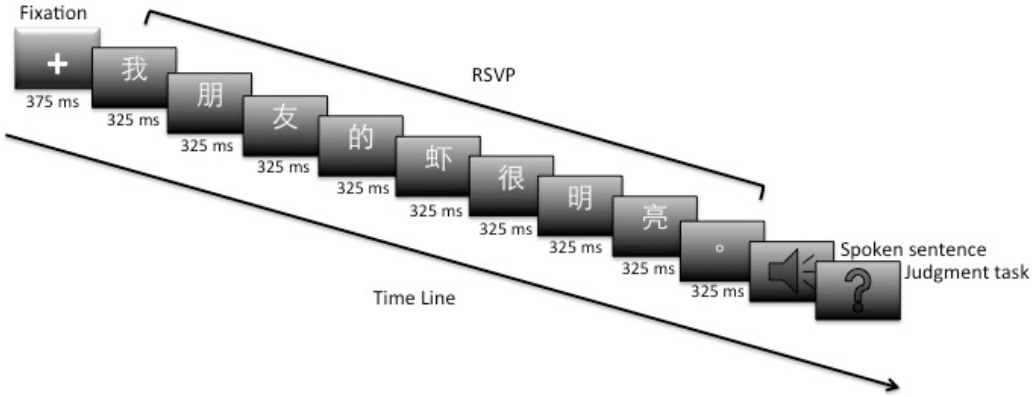


Figure 1: Illustration of experimental paradigm

3.4 Electrophysiological recording

Continuous EEG was recorded using BioSemi Active-Two system at a sampling rate of 512 Hz bandpass filtered at 0.01 Hz to 100 Hz using Ag/AgCl electrodes to record from 64 channels on a head cap labeled according to the International 10-20 system (see Figure 2). Online recordings were referenced to the nose tip and re-referenced offline to the mastoids. Electrooculographic (EOG) activity was recorded from electrodes placed above and over the outer canthus of the left eye. Data were processed using Brain Vision Analyzer (Brain Products, Version 2.0.4). EEG was segmented into epochs from 200 ms pre-stimulus to 750 post-stimulus onset time-locked to the onset of the auditory stimulus. Following this, data were baseline corrected to the mean voltage of the pre-stimulus interval. Trials containing blinks and other artifacts were removed using ocular correction independent component analysis (ICA) with a maximum voltage criterion of +/-100 uv for the portion of the waveform subjected to statistical analysis (Jung, Makeig, Humphries, Lee, Mckeown, Iragui, & Sejnowski, 2000; Makeig, Jung, Bell, Ghahremani,

& Sejnowski, 1997). After ICA, 61.11/62 (99%), 60.89/62 (98%), 61.11/62 (99%), and 60.5/62 (98%) of the trials were left for the baseline, tone violation, onset violation, and syllable violation condition respectively. In addition, incorrect trials (match trials identified as mismatch trials and mismatch trials identified as match trials) were also rejected. Therefore, the final number of trials included in the grand average analysis is 35.06/61.11 (57%), 54.94/60.89 (90%), 58.22/61.11 (95%), and 59.39/60.5 (98%) for the baseline, tone violation, onset violation, and syllable violation condition respectively. The large number of trials rejected from incorrect behavioural responses to the baseline condition may be due to the fact that we did not control for the type of dialect spoken by our participants. As a result, target words may have been interpreted as incorrect due to certain ambiguous pronunciations in Mandarin. For example, speakers from Taiwan pronounce 'week' as /xing1qi2/, whereas speakers from Beijing pronounce 'week' as /xing1qi1/, where both pronunciations are acceptable in Mandarin.

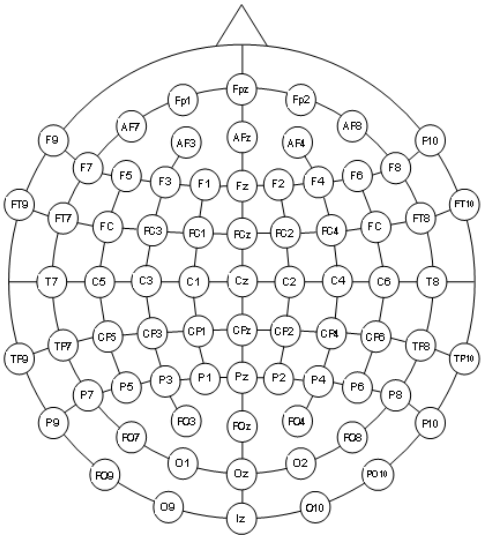


Figure 2: Layout of 64-channel setup for EEG recording

3.5 Data analysis

Sentences were grouped into four experimental conditions that were identified as having an expected or unexpected onset, and a right or wrong tone: expected onset and right tone (baseline), expected onset and wrong tone (tone violation), unexpected onset and right tone (onset violation), and unexpected onset and wrong tone (syllable violation) (see Table 2 for summary). Triggers were positioned at specific times within each stimulus to mark relevant events. Markers were placed at the beginning of the first segment and coda of the target word using Praat (Boersma & Weenink, 2015) to indicate when the onset and tone occurred within each sentence.

		Tone	
		Right	**Wrong**
Onset	**Expected**	*Baseline* (no violation)	*Tone* violation
	Not expected	*Onset* violation	*Syllable* violation

Table 2: Condition summary

Electrodes were clustered into regions of interest (ROIs) according to Neural ElectroMagnetic Ontologies technical reports (ElectroMagnetic Ontologies [NEMO], 2013; See Frishkoff, Sydes, Mueller, Frank, Curran, Connolly, Kilborn, Molfese, Perfetti,

& Malony, 2011), where electrode sites were divided evenly into 4 levels of *caudality* (frontal, central, parietal, and occipital), and 5 levels of *laterality* (mid, left, right, left ventral/temporal, and right ventral/temporal). This yielded 20 ROIs: mid frontal (Fz, Af1, Afz, Af2, F1, F2), left frontal (F3, Af3, Af5, F5), right frontal (F4, Af4, Af6, F6), left frontotemporal (F7, Af7, F9), right frontotemporal (F8, Af8, F10), mid central (Cz, C1, C2, Fcz, Fc1, Fc2), left central (C3, C5, Fc3, FC5), right central (C4, C6, Fc4, FC6), left centrotemporal (T7, T9, FT7, FT9), right centrotemporal (T8, T10, FT8, FT10), mid parietal (Pz, P1, P2, Cpz, Cp1, Cp2), left parietal (P3, P5, Cp3, Cp5), right parietal (P4, P6, Cp4, Cp6), left posterotemporal (P7, P9, TP7, TP9), right posterotemporal (P8, P10, TP8, TP10), mid occipital (Oz, POz, PO1, PO2), left occipital (O1, PO3, PO5), right occipital (O2, PO4, PO6), left occipitotemporal (PO7, PO9, I1), and right occipitotemporal (PO8, PO10, I2) (see Table 3 and Figure 3 respectively for a summary and visual representation of ROIs).

There were three ERP components of interest in this experiment: the PMN, early N400, and N400. Based on the visual inspection of grand average waveforms (Figure 7), mean amplitudes were measured for the following three time windows: from 200-300 ms (PMN), from 300-400 ms (early N400), and from 400-500 ms (N400). A two-way repeated-measures analysis of variance (ANOVA) was conducted for each component's mean amplitude with *Onset* (expected and not expected) and *Tone* (right or wrong) as factors. A subsequent three-way repeated-measures ANOVA was conducted for each component's mean amplitude with an added *ROIs* (20 regions) factor (*Onset* x *Tone* x *ROIs*). In the situation where there was a significant *Onset* x *Tone* interaction, a post-hoc

Tukey's honestly significant difference test was conducted in order to further investigate which levels were significantly different from each other. Compared to the Bonferroni method, the Tukey's honestly significant difference test has greater power to detect significant differences for levels with equal numbers of observations (Baayen, 2008). In the situation where there was a significant *Onset* x *Tone* x *ROIs* interaction, a two-way ANOVA (*Onset* x *Tone*) was conducted for each ROI in order to investigate the component's scalp topography.

Regions of Interest (ROI)		**Electrodes**
Frontal	Mid frontal	Fz, Af1, Afz, Af2, F1, F2
	Left frontal	F3, Af3, Af5, F5
	Right frontal	F4, Af4, Af6, F6
	Left frontotemporal	F7, Af7, F9
	Right frontotemporal	F8, Af8, F10
Central	Mid central	Cz, C1, C2, Fcz, Fc1, Fc2
	Left central	C3, C5, Fc3, FC5
	Right central	C4, C6, Fc4, FC6
	Left centrotemporal	T7, T9, FT7, FT9
	Right centrotemporal	T8, T10, FT8, FT10
Parietal	Mid parietal	Pz, P1, P2, Cpz, Cp1, Cp2
	Left parietal	P3, P5, Cp3, Cp5
	Right parietal	P4, P6, Cp4, Cp6
	Left posterotemporal	P7, P9, TP7, TP9
	Right posterotemporal	P8, P10, TP8, TP10
Occipital	Mid occipital	Oz, POz, PO1, PO2
	Left occipital	O1, PO3, PO5
	Right occipital	O2, PO4, PO6
	Left occipitotemporal	PO7, PO9, I1
	Right occipitotemporal	PO8, PO10, I2

Table 3: Summary of electrode clusters adapted from NEMO (2013)

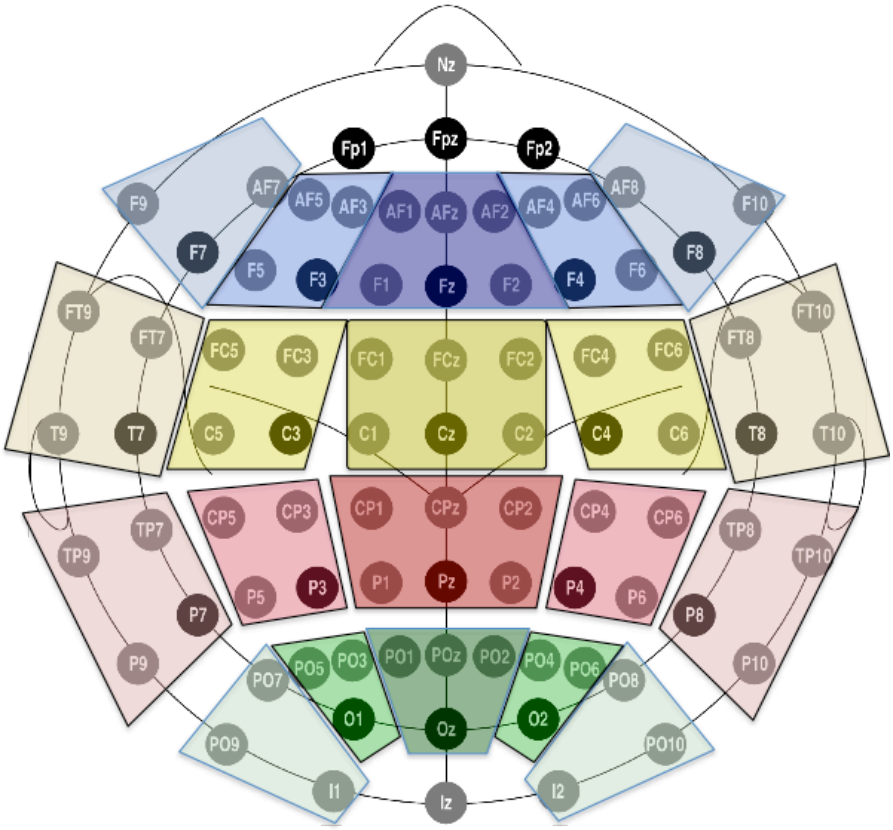


Figure 3:Visual representation of ROIs (NEMO, 2013)

4.0 Results

4.1 Electrophysiological findings

Upon visual inspection of grand average waveforms, all conditions that were characterized by unexpected stimulus features of either *Tone*, *Onset*, or both appeared to differ from the *Baseline* condition (see Table 2, above) in which both *Tone* and *Onset* met contextual expectations. Both grand average ERPs (Figure 4) and scalp maps (Figure 5) show that violation conditions elicited large negative-going ERPs with primarily parietal

distributions. Overall, the *Tone* violation condition elicited the largest and most prolonged negativity, whereas the *Syllable* violation condition (i.e., the condition in which both *Tone* and *Onset* violated expectations) elicited the smallest with the *Onset* violation condition eliciting a negative-going response complex having an amplitude that fell between the *Tone* and *Syllable* violation conditions.

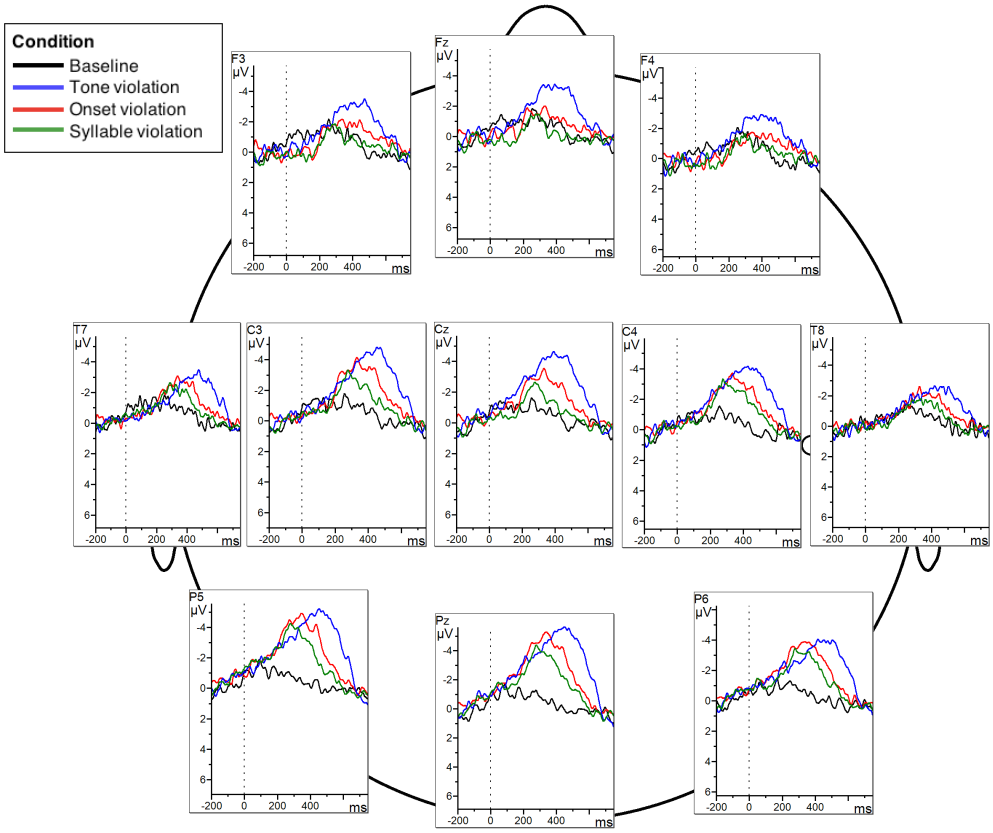


Figure 4: Grand average ERPs (N=18) to experimental conditions: baseline (black), tone violation (blue), onset violation (red), and syllable violation (green)

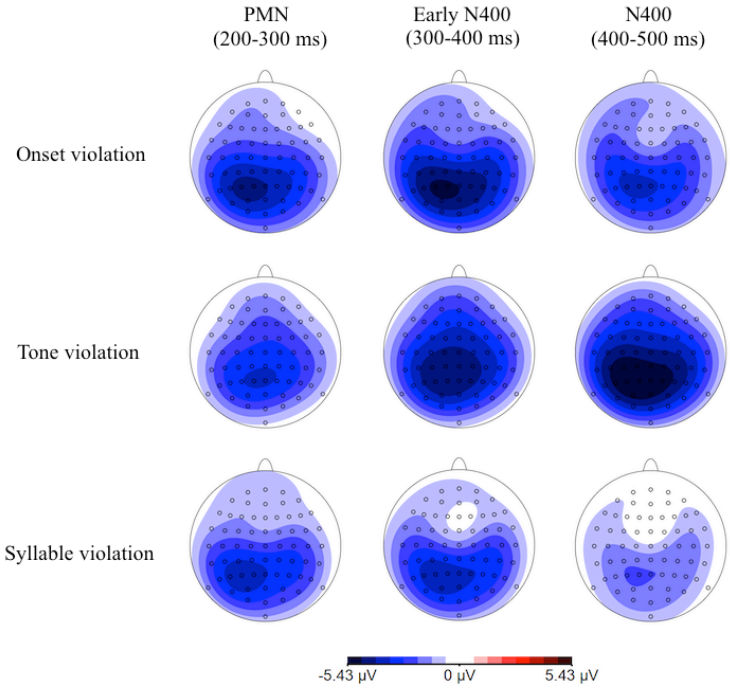


Figure 5: Topographical maps of grand average ERPs (N=18) for the PMN, early N400, and N400

4.2 PMN (200-300 ms)

The ANOVA evaluated the effects on the PMN of two factors, onset and tone, each with two levels related to either violating or meeting contextual expectations. This 2x2 design led to four (4) conditions: *Baseline*, *Onset*, *Tone*, and *Syllable* (see above). The ANOVA revealed a significant main effect of *Onset* ($F(1,17) = 5.14$, $p < 0.04$). However, this result cannot be interpreted without consideration of the *Onset* x *Tone* interaction ($F(1,17) = 13.88$, $p<0.002$) (see Figure 6). Post hoc analyses of this interaction revealed that violations of tonal expectations resulted in large PMN amplitudes regardless of whether onsets met or violated expectations; an effect supported

statistically by the failure of PMN responses to tone violations (red line, Figure 6) to differ whether onsets were expected or not (p=0.87), the *Syllable* v. *Tone* comparison. All other comparisons proved to be significant. The larger PMN response to tonal violations regardless of onset characteristics was reinforced by the large PMN amplitude difference between tonal conditions when onset expectations were met (p<0.0001) (i.e., the *Baseline* v. *Tone* comparison). That is, the PMN was small when tone and onset expectations were met (*Baseline*) compared to the significantly larger PMN when tonal violations occurred but onset expectations were met (*Tone*) (Figure 6). It was clear, however, that onset violations exhibited a powerful effect on the PMN, which proved significantly larger when tonal expectations were met but onset expectations were not (p<0.0001) (i.e., the *Baseline* v. *Onset* comparison) (Figure 6). In fact, the *Onset* condition produced the largest PMN in the experiment being significantly larger than the PMN seen in the *Syllable* condition (p<0.003) and in the *Tone* condition (p<0.04) (Figure 6). The *Baseline* v. *Syllable* comparison represented theoretically the greatest contrast between conditions as it compared the condition where both onset and tone occurred in accordance with expectations compared with the condition where neither tone nor onset met expectations. The post-hoc comparisons supported the contrast between conditions (p<0.0001) with the larger PMN occurring in the *Syllable* condition (Figure 6).

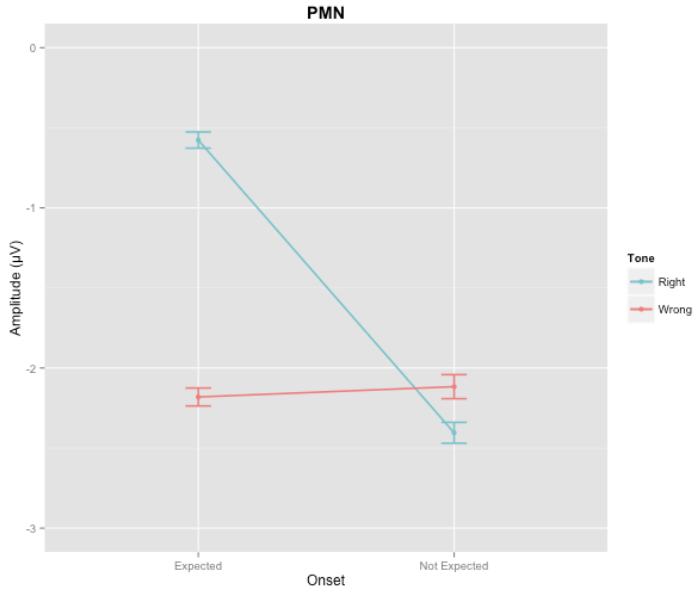


Figure 6: PMN component – mean amplitudes in the Baseline (top left data point), Tone (bottom left data point), Onset (bottom right data point), and Syllable (top right data point) conditions.

The most salient aspect of the PMN's behaviour in the interaction between tone and onset is that the PMN amplitude was large if the tone was unexpected regardless of whether the onset was expected or unexpected. Thus, tone information took priority over phonemic features when the tone information was recognized as unexpected or "wrong." However, when tone information was identified as correct then phonemic violations manifested here as unexpected onsets became important and led to the largest PMN statistically seen in this interaction as the *Onset* effect (Figure 6).

There was a significant interaction of *Onset* x *Tone* x *ROIs* for the PMN ($F(20, 340) = 8.41$, $p<0.001$). A secondary ANOVA conducted on each ROI revealed the most significant *Onset* x *Tone* interaction in the mid parietal region ($F(1,428) = 66.36$,

$p<0.001$) followed by the mid central region ($F(1,428) = 49.59$, $p < 0.001$). See Figure 9 for a heat map of the *Tone* x *Onset* interaction for each ROI in the PMN time window.

4.3 Early N400 (300-400 ms)

Analysis of the 300-400 ms interval revealed a significant *Tone* effect ($F(1,17) = 4.75$, $p < 0.05$) with larger N400 amplitudes observed for unexpected (wrong) tones compared with expected (right) tones. This main effect is best understood by considering the *Onset* x *Tone* interaction ($F(1,17) = 21.98$, $p < 0.0001$). Post hoc analyses of the *Onset* x *Tone* interaction confirmed the clear interaction seen in Figure 7. Like the PMN, the early N400 exhibited large amplitudes to tone violations. However, unlike the PMN, the early N400 amplitudes to tone violations were sensitive to whether onset expectations were met or not with larger amplitudes being observed when onset expectations were *met* than when they were violated ($p<0.001$) (red line, Figure 7) (the *Syllable* v. *Tone* comparison). Again, like the PMN, early N400 amplitudes were seen to tone violations when tone onset expectations were met ($p<0.0001$) (the *Baseline* v. *Tone* comparison). This effect is reflected by the reduced early N400 when tone and onset expectations were met (*Baseline*) compared to the significantly larger response when tone violations occurred when onset expectations were met (*Tone*) (Figure 7). The complementary comparison examining onset effects when tone expectations were met (the *Baseline* v. *Onset* comparison) revealed a major effect of the violation of onset expectations on the early N400 amplitude, which was larger when onset expectations were violated than met ($p<0.0001$) (blue line, Figure 7). The *Onset* condition did not produce the largest early

N400 amplitude in the study as it did for the PMN. While the early N400 was larger in *Onset* condition than in the combined onset and tone violation condition (*Syllable*, p<0.0001) there was no difference in amplitudes between the *Onset* and *Tone* conditions (ns). The *Baseline* v. *Syllable* comparison represented the greatest contrast theoretically amongst the four conditions – a contrast that resulted in the co-largest amplitude differences seen (p<0.0001).

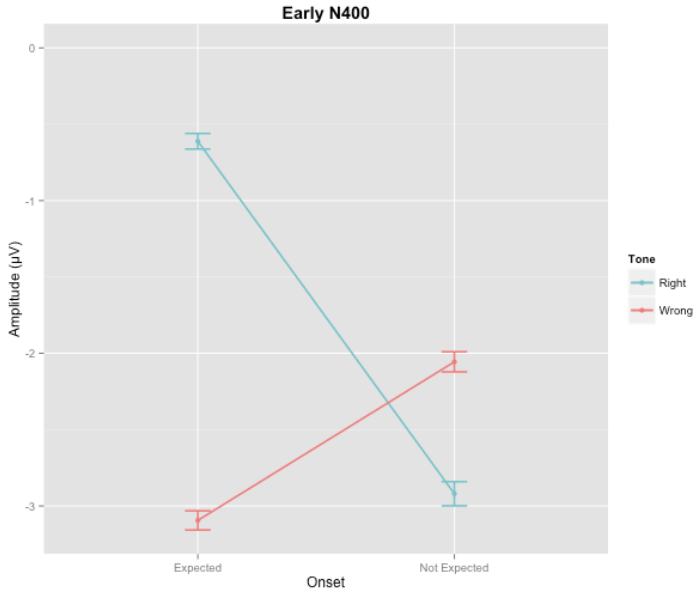


Figure 7: Early N400 component – mean amplitudes in the *Baseline* (top left data point), *Tone* (bottom left data point), *Onset* (bottom right data point), and *Syllable* (top right data point) conditions.

There was also a significant *Onset* x *Tone* x *ROIs* interaction ($F(20, 340) = 9.85$, $p < 0.001$), suggesting that different regions of the scalp were sensitive to onset expectancy and tone appropriateness. The ANOVA conducted on each ROI revealed the most significant *Onset* x *Tone* interaction in the mid parietal region ($F(1,428) = 139.39$, $p < 0.0001$) followed by the mid central region ($F(1,428) = 122.51$, $p < 0.0001$). See Figure 9

for a heat map of the *Tone* x *Onset* interaction for each ROI in the early N400 time window.

4.4 N400 (400-500 ms)

An ANOVA of the peak latency of the N400 response revealed a significant main effect for *Tone* ($F(1,17) = 13.56$, $p < 0.002$), suggesting that the N400 component was sensitive to tone appropriateness. However, the N400 sensitivity to tone violations is best interpreted in light of the *Onset* x *Tone* interaction ($F(1,17) = 25.30$, $p < 0.0001$). Post hoc

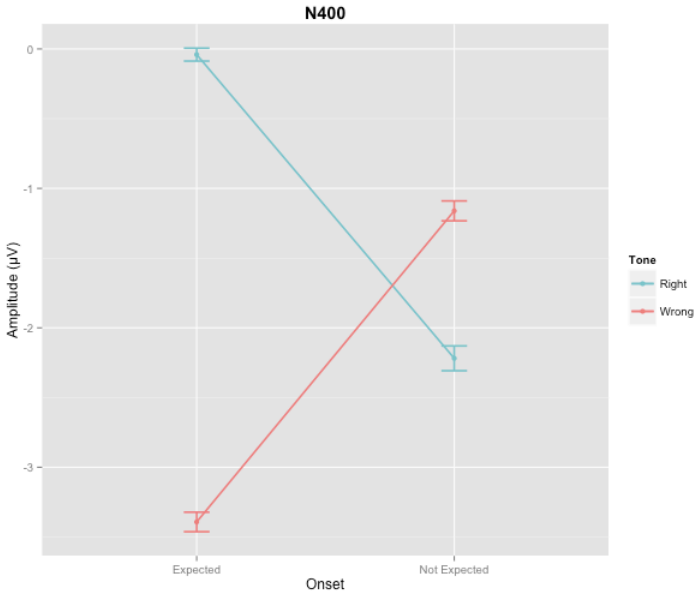


Figure 8: N400 component – mean amplitudes in the *Baseline* (top left data point), *Tone* (bottom left data point), *Onset* (bottom right data point), and *Syllable* (top right data point) conditions.

analyses revealed what might be considered a continuation of a trend across the three components being investigated and the timeframe from 200-500 ms of tonal precedence combined with a complex relationship between tone and onset importance in word

perception. In this 400-500 ms latency window, all conditions proved to be significantly different from each other for the first time. As has been seen in the early timeframes, the importance of tonal violations is demonstrated by the significant difference in the N400 amplitude in the *Tone* condition compared to that seen in the *Syllable* condition ($p < 0.0001$) (red line, Figure 8). This robust difference emphasizes that tone violations in the presence of onset expectations being met represent a highly salient cue for speech perception; particularly for the N400 where for the first time the response produced is significantly different from all other conditions including a tonal violation combined with an onset violation (the *Syllable* condition). Consistent with both the PMN and the early N400, amplitudes for the N400 to tone violations when onset expectations were met proved to be significantly larger than to amplitudes when both tone and onset expectations were met ($p < 0.0001$) (the *Baseline* v. *Tone* comparison, Figure 8). The comparison of onset effects when tone expectations were met (*Baseline* v. *Onset*) revealed a major effect of the violation of onset expectations on the N400 amplitude, which was larger when onset expectations were violated than met ($p<0.0001$) (blue line, Figure 8). The N400 was larger in *Onset* condition than in the combined onset and tone violation condition (*Syllable*, $p<0.0001$). Unlike results for the early N400, N400 amplitudes differed significantly between the *Onset* and *Tone* conditions ($p< 0.0001$). The *Baseline* v. *Syllable* comparison found, as was observed for the PMN and early N400 component, a significant difference in N400 amplitudes ($p<0.0001$) (Figure 8).

The ANOVA on the N400 also revealed a significant three-way interaction of *Onset* x *Tone* x *ROIs* ($F(20, 340) = 10.49$, $p < 0.001$). The *Onset* x *Tone* ANOVA

conducted on each ROI revealed the most significant interaction in the mid parietal region ($F(1,428) = 168.15$, $p < 0.001$) followed by the left parietal region ($F(1,284) = 129.82$, $p < 0.001$). See Figure 9 for a heat map of the *Tone* x *Onset* interaction for each ROI in the N400 time window.

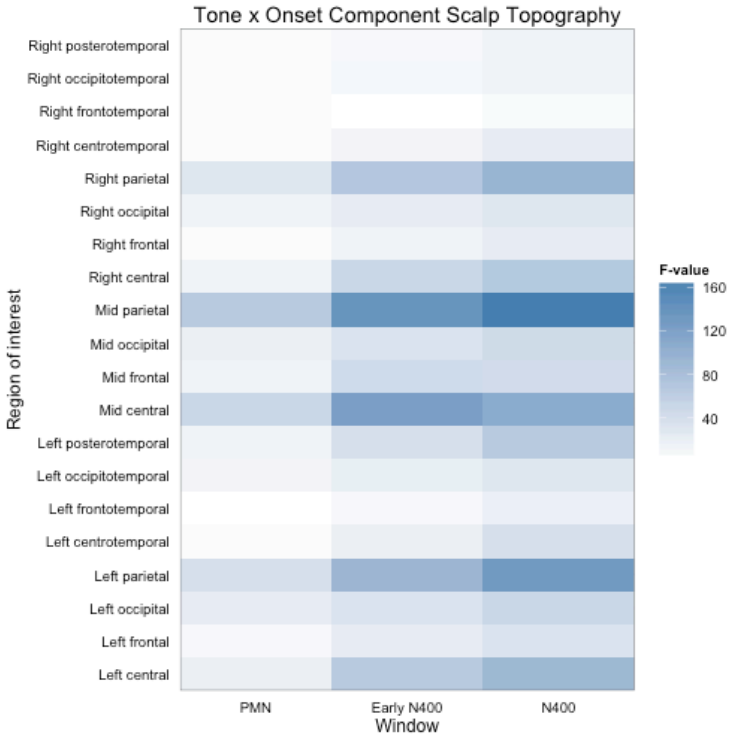


Figure 9: Heat map of the Tone x Onset interaction in 20 regions of interest (ROIs)

4.5 Topographical t-test comparisons

Paired t-tests on grand average waveforms were conducted for each point in time to compare ERPs between each violation condition. Each violation condition was compared with baseline to further establish significant differences in component scalp topography (see Figure 10). Differences were obtained within subjects and averaged and

tested against zero on the grand average level. The following distributions based on the highest t-test values can be seen. Onset violations: The PMN exhibited a predominantly left parietal distribution, which was similar but stronger than that seen for both the early N400 and N400. Tone violations: Although widely distributed the PMN's primary focus was right parietal. The early N400 was distributed broadly across the parietal region with a noticeable right parietal strength. In contrast the N400 exhibited a left parietal-midline parietal distribution. Syllable violations: The PMN showed a similar but weaker distribution similar to that seen to Onset violations. A similar description could be said to apply to the early N400. The N400 showed a weak predominantly left sided distribution with a focus in the right temporal area.

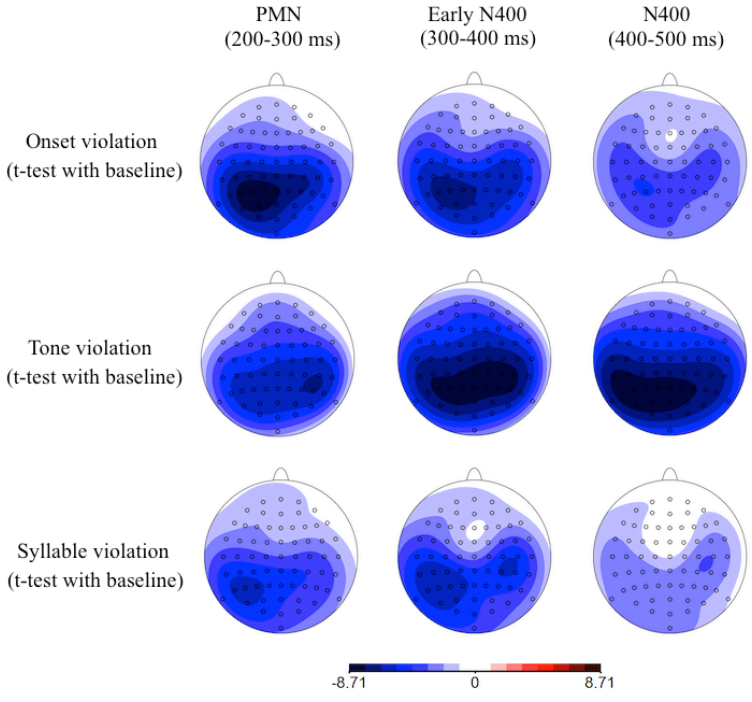


Figure 10: Topographical t-test comparisons (t-values) for the PMN, early N400 and N400

4.6 Difference waveforms

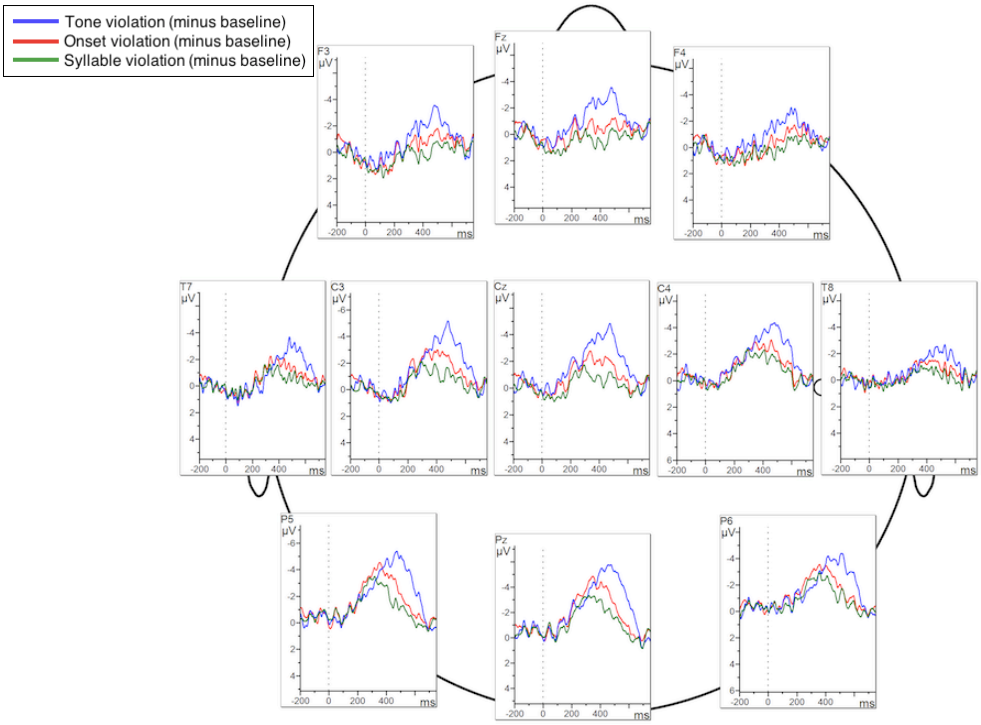


Figure 11: Difference waveforms of each violation condition minus the baseline condition

In order to better understand the relative amplitudes and distributions of each ERP component to the experimental conditions, we examined difference waveforms to draw more precise conclusions about the nature of Mandarin spoken word recognition. Difference waveforms were derived from the subtraction of the baseline condition from each violation condition. As illustrated in Figure 11, response amplitudes for each violation condition maintained a constant relationship with each other across the head with tonal violations consistently resulting in the largest amplitudes in the 350-600 ms period; an effect due primarily to the N400. This effect is seen most notably at parietal sites. There was some indication of a left-hemisphere dominance in this same latency

window (Figure 12). This suggests that tonal information is essential in Mandarin spoken word recognition. However, between 200-300 ms, the onset violation condition elicited the most negative ERPs compared to the other violation conditions. Even though tone is a prominent aspect in Mandarin, the onset of a spoken word, like English, still seems to be the first linguistic element to be recognized.

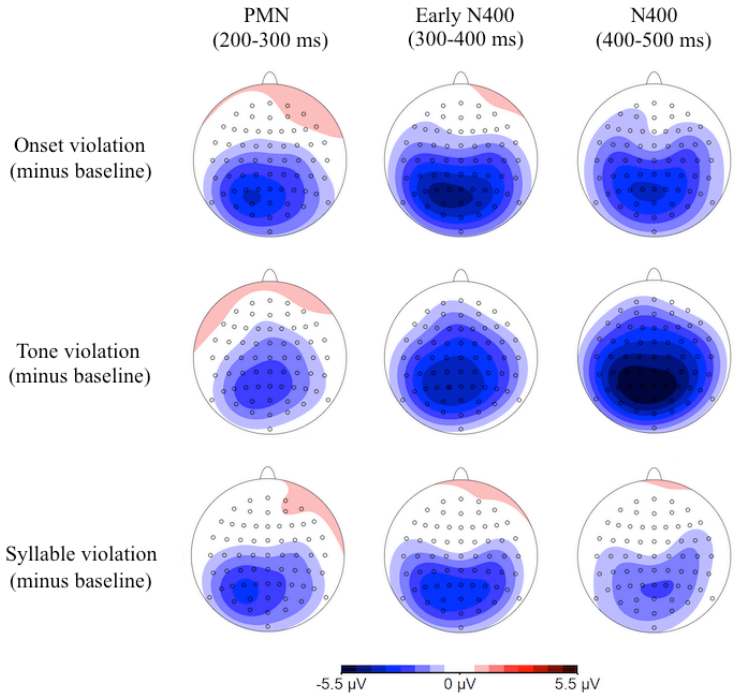


Figure 9: Topographical maps for the subtractions of each violation condition minus the baseline condition for the PMN, early N400 and N400

5.0 Discussion

The present study used ERPs to investigate Mandarin spoken word recognition through the manipulation of tonal and segmental information in Mandarin. Recall, the purpose of this study was to explore (1) the P200 component's sensitivity to phonological

mismatch in Mandarin; (2) the relationship between the "early N400" and the PMN; (3) whether Mandarin is processed incrementally or holistically; and (4) whether current spoken word recognition models adequately explain tonal language processing. Sentence-medial target words were presented in RSVP, which were confirmed or violated by the subsequent spoken sentence. Mismatch words varied in onset expectancy and tone appropriateness, allowing us to address the proposed questions about the time course of Mandarin spoken word recognition.

5.1 Indeterminate early phonological processing stage

Huang et al. (2014) found a P200 effect when disyllabic target words differed in segmental but not tonal information in their unrelated condition (e.g., /r*an*2*hou*4/–/ge2shi4/; "then"–"format"). Thus, they proposed a pre-lexical phonological processing stage in Mandarin that affected PMN amplitudes. Our results did not support these claims, as we did not observe any P200 effects in our 200-300 ms time window. Thus, we have no evidence confirming the P200 component and its influence on the attenuation of PMN amplitudes in processing phonological mismatches in Mandarin. Result discrepancies may be due to the fact that we employed a cross-modal paradigm on sentence-medial monosyllabic target words, whereas Huang et al. (2014) employed a unimodal auditory paradigm on disyllabic target words. Although cross modality designs may pose more phonological links between prime and target words (Gaskell & Marslen-Wilson, 2002), our analyses revealed that all violation conditions elicited significantly more negative responses than baseline condition in the 200-300 ms time window. This

demonstrates the robust effect of the PMN, at least in regards to our phonological manipulations in a cross-modal priming paradigm. Huang et al. (2014) claimed they were the first to report a P200 phonological mismatch effect independent of semantic interactions that attenuates PMN amplitudes during spoken word recognition. However, these claims warrant further examination.

5.2 Lexical tone in semantic access

As lexical tone is a supra-segmental feature that provides both lexical and phonological information, it is noteworthy to investigate how tone influences the PMN and N400 component. The PMN is associated with pre-lexical phonological processing (Connolly & Phillips, 1994; Newman & Connolly, 2009), whereas the N400 is more associated with the word-level access of meaning (Kutas & Hillyard, 1984). As illustrated in Figures 5, 6, 11, and 13, the tone violation condition elicited the largest negativity between 400 and 500 ms (N400) in the mid parietal region, demonstrating the overall importance of tone for semantic expectations. This does not necessarily imply that tone is by no means phonological in nature since the N400 is also sensitive to different levels of phonological structure (see Kutas & Federmeier, 2011). However, it is evident from our results that lexical tone is inseparable from semantic processing since listeners have the most difficulty resolving temporary ambiguities when tonal information was violated. Therefore, recognition in Mandarin is impossible without tonal information, which appears to be the most salient aspect of a word. To further explore the integral nature of tone in the recognition of Mandarin, it is essential to consider non-word

processing. Thus far, only one study has investigated non-words in Mandarin spoken word recognition, where the first syllable of disyllabic words was manipulated to create nonwords, which elicited "N400-like" responses (Liu et. al., 2005). The problem with these non-word manipulations, however, is the fact that lexical tone is still intact with the second syllable, allowing the listener partial advantage in resolving tonal ambiguities. If tone were central to semantic access, there would be uncharacteristic N400 responses to Mandarin monosyllabic non-words. This has yet to be investigated in Chinese spoken word recognition, and further research in this domain is crucial to unraveling the full implications of tonal perception.

5.3 Segmental phonological word recognition

Analogous to the results found in Malins & Joanisse (2012), a larger PMN response was observed for the onset violation condition compared to the tone violation condition, suggesting that the onset was recognized before the tone of a word (c.f., Malins & Joanisse, 2012, Figures 3a, c). Our onset violation condition elicited a response similar to semantically appropriate words beginning with an unexpected phoneme in English. Yet, our tone violation condition elicited a response similar to semantically inappropriate words beginning with an expected phoneme in English (c.f., Connolly et al., 1992; Connolly et al., 1994; Connolly et al., 1995; Connolly et al., 2001; Van Den Brink et al., 2001). The distribution of the PMN and N400 in Mandarin seem to be similar to that of English, except tone is the main linguistic component that determines semantic access. This provides evidence for incremental phonological processing in Mandarin speech

recognition. Contrary to the results found in Zhao et al. (2011), we did not find whole-syllable violations that appeared earlier with higher amplitudes than our partial violation conditions (onset and tone). In fact, we found that the syllable violation condition elicited overall attenuated negativities as compared to the onset and tone violation conditions. These results were similar to those found in Huang et al. (2014), as reduced PMN amplitudes were observed in their unrelated condition (e.g., /*ran*2*hou*4/–/ge2shi4/; "then"–"format") as compared to their cohort (e.g., /ge2*bi*4/–/ge2shi4/; "the next door"–"format") and identical (e.g., /ge2shi4/–/ge2shi4/; "format'–"format") conditions. This was because disambiguation was unnecessary in the unrelated condition, since there was an immediate absence of competition effects from the first syllable of the target word. The lack of segmental and tonal overlap between our prime and target words explained why our syllable violation condition elicited the overall smallest ERP amplitudes for the PMN, early N400, and N400. These findings therefore refute the claim that Chinese listeners perceive monosyllabic spoken words as a holistic unit (Zhao et al., 2011), rather, our findings support the claim that Chinese listeners perceive monosyllabic spoken words through individual phonological segments as the stimulus unfolds in time (Malins & Joanisse, 2012).

5.4 Implications for models of spoken word recognition

The findings from the current study provide a broader understanding of the linguistic measures involved in the phonological processing of Mandarin. This in turn provides implications for spoken word recognition models, such as the Neighbourhood

Activation Model (NAM) (Luce & Pisoni, 1998), Cohort model (Marslen-Wilson & Tyler, 1980), Shortlist/MERGE (Norris, McQueen & Cutler, 2000), and TRACE (McClelland & Elman, 1986). The NAM is based on the overall phonological similarity of potential lexical candidates, where the degree of phonological competition among words is equal for differences occurring in word-initial, word-medial, and word-final positions. Since we found discrete ERP modulations for phonological differences occurring word-initially (onset violation) and word-finally (tone violation), the NAM is unsuited for the perception of Mandarin. This conclusion is different from the one reached by Zhao et al. (2011), as they found comparable ERP effects to onset, rime, and tone mismatches. The nature of the tasks is a likely reason for this difference in conclusions. Instead of a passive listening task, our task required subjects to actively process the target words by subsequently making decisions about them, which facilitated the perception of subtle differences between visual and auditory stimuli. Since competition was found among cohorts and rimes, feedforward models, such as the Cohort model and Shortlist/MERGE are more appropriate for Mandarin. However, as we investigated spoken words as they unfolded in real time, our results were more in line with temporal processing models, such as TRACE, which allows for interactive bottom-up and top-down influences on speech perception. As suggested by Malins & Joanisse (2012) and Zhao et al. (2011), the incorporation of tonal feature detectors is necessary to fully accommodate for the processing of tonal languages. We suggest tonal feature detectors at the phoneme level (see Malins & Joanisse, 2012) instead of syllable level (see Zhao et al., 2011), since we did not observe syllable-mismatch effects over and above

individual segmental-mismatch effects. Recall studies on English spoken word recognition, where equal PMN amplitudes were found for initial phonological mismatches and for several subsequent phonological mismatches (Newman et al., 2003). Though, varying N400 amplitudes were observed for varying degrees of semantic relatedness (Kutas & Van Petten, 1994). Our findings suggest that Mandarin is processed similarly to languages like English, but with an added tonal element. This puts forward the proposal that Mandarin may also have two levels of processing, in which the lower level is an all-or-none response to phonological violations and the higher level is sensitive to gradations of semantic violations. This proposal requires further investigation in Chinese spoken word recognition, where manipulations include gradient levels of phonological violations and semantic relatedness.

6.0 Conclusion

We used ERPs to examine segmental and tonal violations and their effects on Mandarin spoken word recognition. To date, only a handful of studies have investigated the role of tone on word recognition processes, and even fewer studies have investigated word recognition processes of sentence-medial target words in Chinese. We did not find a pre-lexical phonological processing stage in Mandarin, however, a replication of our study using strictly auditory priming is necessary to dismiss this proposal completely. Our findings suggest that recognition in Chinese is incremental, where tonal information is still phonological but is realized when it becomes available, which evidently, is after the onset of the word. We also found that lexical tone in Mandarin is crucial in accessing

semantic word meaning. This sheds light on refining pedagogical methods in the second language acquisition of Mandarin by emphasizing tonal over individual segmental perception. Nevertheless, our results indicate that Mandarin syllables are processed incrementally like English albeit tone is more prominent than individual segments in word recognition. Lastly, these findings help us understand how online processing models of spoken word recognition can be modified to accommodate tonal language perception, allowing for a unified model in spoken word recognition that accounts for both tonal languages and Indo-European languages.

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.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Biederman, I., & Tsao, Y. C. (1979). On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology*, *11*(2), 125-132.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.3.84, retrieved 1 August 2015 from http://www.praat.org/

Brain Products GmbH. Brain Vision Analyzer (Version 2.0.4) (software). *Germany: Munich.*

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.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*(4), 463-480.

Connolly, J. F., Byrne, J. M., & Dywan, C. A. (1995). Assessing adult receptive vocabulary with event-related potentials: an investigation of cross-modal and cross-form priming. *Journal of Clinical and Experimental Neuropsychology*, *17*(4), 548-565.

Connolly, J. F., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, *12*(2), 237-243.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and language*, *43*(1), 1-18.

Connolly, J. F., Service, E., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, *12*(2), 237-243.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and language*, *39*(2), 302-318.

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

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.

Frishkoff, G., Sydes, J., Mueller, K., Frank, R., Curran, T., Connolly, J., Kilborn, K., Molfese, D., Perfetti, C., & Malony, A. (2011). Minimal Information for Neural Electromagnetic Ontologies (MINEMO): A standards-compliant method for analysis and integration of event-related potentials (ERP) data. Standards in Genomic Sciences, 5(2): 211–223. doi: 10.4056/sigs.2025347

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive brain research*, *1*(3), 183-192.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and cognitive Processes*, *12*(5-6), 613-656.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive psychology*, *45*(2), 220-266.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psycbopbysics, 28,* 267-283.

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*(6), 715-728.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, *16*(1), 38-50.

Ho, C. S. & Bryant, P. (1997). Development of phonological awareness of Chinese children in Hong Kong. *Journal of Psycholinguistic Research, 26*(1), 109–126.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47-61.

Howie, J.M., 1976. Acoustical Studies of Mandarin Vowels and Tones. Cambridge University Press, New York.

Huang, X., Yang, J. C., Zhang, Q., & Guo, C. (2014). The time course of spoken word recognition in Mandarin Chinese: A unimodal ERP study. *Neuropsychologia*, *63*, 165-174.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(02), 163-178.

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., & Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Cognitive brain research*, *21*(1), 106-113.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, *62*, 621.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological psychology*, *11*(2), 99-116.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. *Preparatory states and processes*, 217-237.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. *Advances in psychophysiology*, *3*, 139-187.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. *Handbook of psycholinguistics*, 83-143.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context.*Journal of Memory and Language*, *61*(3), 326-338.

Lesch, M. F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 285.

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 and hearing*, *19*(1), 1.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge, MA: MIT.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York: Wiley. 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.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, *94*(20), 10979-10984.

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.

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

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1-71.

Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 823-836.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological review, 101*(4), 653.

Mazzoni, D., & Dannenberg, R. (2000). Audacity (software). *Pittsburg, PA: Carnegie Mellon University*.

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

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1363.

Neural ElectroMagnetic Ontologies. (2013). *Electronic references*. Retrieved 8 July 2015 from http://nemo.nic.uoregon.edu/wiki/File:NEMO-TR-2012-008_ROIdefinitions.docx

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.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 24.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language*, *27*(1), 59-70.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. *New methods in reading comprehension research*, *118*, 91-118.

Ratcliff, R., & McKoon, G. (1981). Automatic and strategic priming in recognition. *Journal of verbal learning and verbal behavior*, *20*(2), 204-215.

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.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. (2011). Language-related ERP components, Chapter 15. In: Luck, S. J., Kappenman, E. S., (Eds.). *Oxford handbook of event-related potential components*. Oxford university press.

Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning?. *Journal of Memory and Language*, *37*(1), 41-57.

Tan, L. H., Hoosain, R., & Siok, W. W. (1996). Activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 865.

Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism quarterly*.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, *36*(5), 417-427.

Van Den Brink, D., Brown, C., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Cognitive Neuroscience, Journal of*, *13*(7), 967-985.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition*, *18*(4), 380-393.

Wang, H., Chang, R. B., & Li, Y. S. (1985). Modern Chinese frequency dictionary.

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

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.

Zou, L., Desroches, A. S., Liu, Y., Xia, Z., & Shu, H. (2012). Orthographic facilitation in Chinese spoken word recognition: An ERP study. *Brain and language*, *123*(3), 164-173.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*(1), 25-64.

Appendix

A: Letter of Information/Consent

DATE: _____

LETTER OF INFORMATION / CONSENT

A study about the online processing of Mandarin tones:
The implications of Mandarin monosyllabic spoken word recognition

Principal Investigator: Amanda Ho
Department of Linguistics and Languages
McMaster University
Hamilton, Ontario, Canada
E-mail: hoas@mcmaster.ca

Faculty Supervisors: Dr. John F. Connolly
Department of Linguistics and Languages
McMaster University
Hamilton, Ontario, Canada
(905) 525-9140 ext. 27095
E-mail: jconnol@mcmaster.ca

Dr. Anna L. Moro
Department of Linguistics and Languages
McMaster University
Hamilton, Ontario, Canada
(905) 525-9140 ext. 23762
E-mail: moroal@mcmaster.ca

What am I trying to find?
You are invited to take part in this study on how Mandarin tones are understood in the brain. I am doing this research as a graduate student in the Cognitive Science of Language program for my M.Sc. thesis. We are hoping to learn the relationship between sound patterns and word meanings in Mandarin speech. This research will help our understanding on whether Mandarin speakers process words as a whole or as individual sounds, contributing to the existing knowledge on how tonal languages fit into current spoken word recognition models.

What will happen?
I will also ask you for some general background information like your age, sex, and handedness for further analysis of the data. You will also complete a short questionnaire in Chinese to ensure you are proficient Mandarin.

For the experiment, you will be seated in front of a computer screen. In order to measure your neural activity, brain recordings will be made using caps that record brain electrical activity. This includes the use of gel on the surface of your scalp for electrical conductivity. This gel is harmless and water-soluble. A cap will be applied to your head with straps and additional sensors will be placed above and to the side of one or both your eyes to record eye movements.

During the experimental trials you will be presented with a sentence on screen, one character at a time. Next, you will hear a sentence through headphones. At the end of each sentence, you will be asked to indicate whether the sentence you read matched the sentence you heard. The entire study will be conducted in one session lasting approximately 2 hours, and you will be given 20 short breaks, lasting 15 seconds each, at regular intervals. Please note that there is no right or wrong answer to this experiment.

You will be rewarded with course credits through the Linguistics participation system. These credits will be applied to your account after you complete the experiment.

Are there any risks?
It is not likely that there will be much harm from this study. But, to improve brain signal recordings, gel will be applied to your scalp, which may cause some discomfort. After the session is complete, you may wash the gel from your hair. In order to obtain valid recordings for my analysis, many stimulus presentations are necessary so you may become tired during the session but stimulus presentation will be stopped at regular intervals in order for you to relax between parts of the experiment.

Due to sensitive questions on the pre-screen form, if there is a conflict of interest between you and the experimenter, another researcher will conduct the pre-screen procedure.

At the end of the experiment, you will need to walk down the hallway of TSH with gel in your hair to access the cleaning facilities. If you are uncomfortable doing so, a towel will be provided for you to wipe off extra gel beforehand. A researcher will walk you to the cleaning facility, where you will have full privacy.

Please note that you do not need to answer questions that you do not want to answer or that make you feel uncomfortable. You can withdraw from this experiment at any time. Described below are the steps I am taking to protect your privacy.

Are there any benefits?
The research will not benefit you directly. We hope to learn more about how the brain processes Mandarin words. I hope that what is learned as a result of this study will help us to better understand how the brain responds to tonal languages, which could help us further understand how tone fits into current spoken word recognition models, contributing to the scientific community at large.

Confidentiality
You are participating in this study confidentially. I will not use your name or any information that would allow you to be identified. No one but me will know whether you participated unless you choose to disclose this information.

All data will be secured in a password-protected computer in the research lab through secure doors at McMaster University for 7 years following the experiment, after which the data will be destroyed. This is in line with other retention policies for personal information, as these data are of high value and can be included in further analyses at a later date.

However, I will not need your pre-screen results in my analysis, so this information will be destroyed immediately after the experiment is completed.

Participation and Withdrawal
It is your choice to be part of the study or not and if you decide to be part of the study, you can stop (withdraw) for whatever reason, even after signing the consent form, or part-way through the study, or up until approximately January 2015, when I expect to be submitting my thesis. If you decide to withdraw, there will be no consequences to you and you will be rewarded partial credit for your participation. In cases of withdrawal, any data you have provided will be destroyed unless you indicate otherwise. Likewise, If you do not want to answer some of the questions you do not have to, but you can still be in the study. If you complete the pre-screen but are ineligible to participate, you will also be rewarded partial credit for your participation.

Information about the Study Results
I expect to have this study completed by approximately August 2015. If you would like a brief summary of the results, please let me know how you would like it sent to you.
Questions about the Study
If you have questions or need more information about the study itself, please contact me at: hoas@mcmaster.ca.

This study has been reviewed by the McMaster University Research Ethics Board and received ethics clearance.

If you have concerns or questions about your rights as a participant or about the way the study is conducted, please contact:

McMaster Research Ethics Secretariat
Telephone: (905) 525-9140 ext. 23142
c/o Research Office for Administrative Development and Support
E-mail: ethicsoffice@mcmaster.ca

CONSENT

- I have read the information presented in the information letter about a study being conducted by Amanda Ho of McMaster University.
- I have had the opportunity to ask questions about my involvement in this study and to receive additional details I requested.
- I understand that if I agree to participate in this study, I may withdraw from the study at any time or up until approximately January 2015.
- I understand that the Language Memory and Brain Lab (LMBLab) intends to use my data in future analyses.
- I have been given a copy of this form.
- I agree to participate in the study.

Signature: _____

Name of Participant (Printed) _____

[ ] Yes, I would like to receive a summary of the study's results.

Contact information:

[ ] No, I do not want to receive a summary of the study's results.

B: Mandarin Prescreening Form

Languages: _____

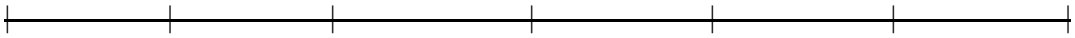
Age of Language Acquisition (If primary language acquired since birth, indicate 0):

Mandarin Chinese (Simplified): _____ English: _____

Other Languages Spoken (in order of fluency):

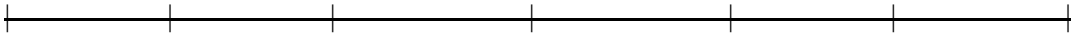
_____

How would you rate your level of reading proficiency in Mandarin Chinese (Simplified) and English?



Fluency in Mandarin Chinese only	Equal/Comparable Fluency in Mandarin and English	Fluency in English only

How would you rate your frequency of use of the languages on a daily basis?



Primary use of Mandarin Chinese	Equal use of both languages	Primary use of English

Reading Proficiency Exercise:
Please read out loud the following sentences in Mandarin to the researcher. Then, match the following sentences with their most appropriate counterpart:

简体中文：

天气太热, 别出去玩儿了， 好吗？	他们两个人在问路。
他还在教室里学习。	好，听您的。
他病了，上午没来公司，看医生去了。	他在哪儿呢？ 你看见他了吗？
我们要去北京大学，请问怎么走？	她很高，也很漂亮，我非常喜欢她。
她对我说，想去外面玩儿。	我不想出去，我想在家里看电视。
你觉得她怎么样？	我上午去医院了，没上班。

C: Demographic Prescreening Form

SCREENING FORM

Study # _____ Participant code: _____ Date of birth: _____ Test date: _____

Handedness: Right Left Ambidextrous Sex: Male Female

Highest level of education completed:

Languages in order of fluency: 1._____2._____3._____4._____

If English is not your first language: How old were you when you learned English?_____

If you were not born in Canada: How old were you when you moved to Canada?_____

History of using psychoactive drugs: _____

Is your hearing and vision normal? Yes No

If not, please describe: _____

Have you ever been diagnosed as having a condition affecting perception, learning, or language? Yes No

If yes, please describe (age, length, recovery): _____

Have you ever been diagnosed as having any neurological or psychological condition? Yes No

If yes, please describe (age, length, recovery):_____

Have you ever had a head injury, seizures, coordination problems or major surgeries? Yes No

If yes, please describe (age, length, recovery):_____

Have you ever lost consciousness, had any fainting spells, paralysis or dizziness? Yes No

If yes, when and for how long? _____

Are you presently taking any prescribed psychoactive drugs? Yes No

If yes, which one(s)?_____

Have you consumed any alcohol or drugs in the last 24 hours? Yes No

If yes, which one(s)?_____

Have you consumed any drugs in the last 7 days? Yes No

If yes, which one(s)? _____

Do you consume the following?

	How many times per day/week/month/year?
Alcohol	
Cigarettes	
Psychoactive Drugs	

Please rate your current state of alertness: - 1 2 3 4 5 +

How many hours did you sleep last night? :_____

EDINBURGH HANDEDNESS INVENTORY

Please indicate your preference in the use of hands in the following activities by listing the "+" in the appropriate columns. When the preference is so strong that you would never try to use the other hand unless absolutely forced to, list "++". If, in any case you really are indifferent, put "+" in both columns.

Some of the activities require both hands. In these cases, the part of the task or object, for which the preference is warranted is indicated in brackets.

Please try to answer all the questions, and only leave the column blank if you have no experience at all of the object of the task.

#	Task	Left	Right
1	Writing		
2.	Drawing		
3.	Throwing		
4.	Scissors		
5.	Toothbrush		
6.	Knife (without fork)		
7.	Spoon		
8.	Broom (upper hand)		
9.	Striking match (match)		
10.	Opening box (lid)		

Score = (Total left_____+ Total right_____)*100 = _____

D: Participant Debriefing Form

PARTICIPANT DEBRIEFING FORM

A study about the online processing of Mandarin tones
The implications of Mandarin monosyllabic spoken word recognition

Thank you for your participation in this study!

The general purpose of this research is to examine brain responses to words spoken in Mandarin Chinese. Mandarin Chinese is a tonal language comprised of syllables carrying different tones that differ in meaning (Ho & Bryant, 1997). "Lexical tone" in Mandarin refers to the variation in the pitch of a speaker's voice that is used to tell apart words that have the same sound patterns (Wang, 1973). In addition, Ho & Bryant state that Mandarin tone provides information about the sound patterning of the word and the word meaning (1997). So, I am most interested in the relationship between the N400 and the Phonological Mismatch Negativity (PMN) in Mandarin tones. The N400 is sensitive word meanings that peaks negatively at about 400 ms after the target word is presented (Connolly & Phillips, 1994). The PMN is a negative-going waveform that peaks between 250-350 ms post-stimulus onset and it is sensitive to sound patterns (Connolly & Phillips, 1994).

To understand whether Mandarin words are processed as a whole or as individual sounds, many researchers have observed Chinese words and its relation to the N400. For example, Schirmer, Tang, Penney, Gunter, & Chen (2005) used Event-Related Potentials (ERPs) to investigate Cantonese words and found that tonal and sound pattern violations prompted the N400. Zhao, Guo, Zhou, & Shu (2011) investigated the same phenomenon using a picture/spoken-word/picture task and found that words differing in tone produced the N400 in the same fashion as words that differed in individual sounds. Likewise, Brown-Schmidt & Canseco-Gonzalez (2004) explored the processing of tone and discovered that semantic mismatches with the tone, syllable, or both the tone and syllable manipulated in sentence-final words elicited N400 effects starting at approximately 150 ms post-stimulus and continuing until 1000 ms.

Conversely, the PMN has been scarcely researched. Joanisse & Malins (2012) also studied the processing of Mandarin words and revealed that when the onset (first sound) of a word was mismatched, a larger PMN was elicited as compared to the rhyme (end of a word) mismatch. However, the tone mismatch and rhyme mismatch conditions showed similar PMN levels, whereas the tone mismatch condition elicited a larger PMN than rhyme mismatch conditions.

If tone provides information about the sound patterning of the word and the word meaning, tonal information should elicit both the N400 and the PMN. To further examine this phenomenon, I manipulated the tone, onset, or syllable of some of the words in this experiment to help us understand some questions in literature regarding how tone is processed.

1. What is the relationship between the N400 and the PMN in Mandarin tones?
2. Do Mandarin speakers process words as a whole or as individual sounds?
3. How do tonal languages fit into spoken word recognition models?

You may obtain additional information about the results of the study by contacting the principle investigator (Amanda Ho) at hoas@mcmaster.ca. If you have any additional concerns or questions, you may contact the McMaster Research Ethics Board Secretariat at (905) 525-9140 Ext. 23142.