Homophony and Phonological Contrasts in Novel Word Learning: A Visual World Eye-Tracking Study with Adult Native and Non-Native Speakers

by

Wenfu Bao

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

Master of Science

Department of Linguistics
University of Alberta

Abstract

This thesis used visual world eye-tracking to examine how adults from native and non-native backgrounds learn novel words that contain homophones and non-homophones, and vary in different types of phonological contrasts: consonant contrasts, tone contrasts, and consonant & tone contrasts. It also investigated whether brief periods of unoccupied rest following learning affect the outcome.

Previous studies have reported that adult non-native learners have difficulties in acquiring Mandarin tones, but mostly based on speech perception and production findings. Despite some work using novel word learning tasks to explore Mandarin acquisition in non-native adults, few studies have explicitly used nonwords to directly compare learning of segmental and tonal contrasts by adult learners. Further, as a common phenomenon in human languages, homophony has been investigated in novel word learning tasks, but much less among adults when compared with infants and young children. In addition, evidence suggests that brief periods of unoccupied rest facilitate memory in a similar manner to sleep. However, little research has examined the role of rest in consolidation of novel word learning in both native and non-native learners.

Thus, in this thesis, we studied 34 native Mandarin speakers and 34 native English speakers in a novel word learning task, which contained 28 novel words (seven homophones, 21 non-homophones) that integrated Mandarin segments and tones. During the task, participants learned novel words in a pair with the minimal difference in consonants, tones, or both. Then they were tested in two phases separated by a 15-minute break, during which participants either had an eyes-closed rest or completed a distractor task. In order to explore whether there was a visual world competition effect in the test phase following learning, the target was presented in

the visual array either with or without its competitor. Participants' response accuracy and latencies in the two test phases, and eye movements in the second test phase were analyzed.

Regarding the effect of participants' language background, results suggest that in the test phase integrated with the learning, Mandarin speakers had significantly higher accuracy than English speakers. However, in the test phase following the training, different tendencies were shown in the two language groups for different types of phonological contrasts: English speakers were significantly more accurate on novel words involved with consonant contrasts than those involving tones, whereas Mandarin speakers did not show significant differences in accuracy between different phonological conditions. In terms of homophony learning, homophones had significantly higher accuracy than non-homophones across all participants. Moreover, Mandarin speakers processed homophones more slowly than non-homophones irrespective of phonological contrasts, though the difference was only significant for phonological conditions involving tones. In addition, only observations of English speakers lent support to the consolidation of novel word learning after brief periods of unoccupied rest. Lastly, our findings highlighted the competition effect in spoken word processing: when there was no competitor in the visual array, participants had significantly faster and more accurate responses, and more target looks.

These findings have contributed to our understanding about novel word learning in adult native and non-native learners from various aspects, such as homophone acquisition, competition effects and the benefit of rest. However, future work is needed to assess the homophony learning pattern in a scenario where a greater proportion of homophones among novel words is used, and to explore novel word learning of Mandarin and the rest-induced memory enhancement in non-native learners from language backgrounds other than English.

Preface

This thesis is an original work by Wenfu Bao. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board 2, Project Name "Learning of Mandarin Tones by Heritage and Non-Native Speakers", No. Pro00089301, June 6, 2019.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisors Dr. Anja Arnhold and Dr. Juhani Järvikivi, who have helped me in so many aspects over the past two years. Their unwavering guidance and support has made my study at the University of Alberta beyond delightful and enjoyable. I appreciate the freedom that they have given to me to pursue the research in which I am interested, and the suggestions and comments that they have provided during all the time of the research and writing of this thesis. They have said that I always look happy, and that's because I am so fortunate to have them in my life.

I would also like to thank Dr. Anne-Michelle Tessier for being the external examiner of my thesis and for providing insightful questions and discussion during the defense; and I thank Dr. Herbert Colston for chairing the examining committee and for making everything run very smoothly.

I gratefully acknowledge the help that I have received throughout the duration of this study. Dr. Liam Blything has extended a great amount of assistance in building the experiment. Dr. Johanne Paradis and Dr. Adriana Soto-Corominas have kindly provided the language aptitude test materials. Dr. Yvonne Lam, Xiaoyun Wang and Cheer Wu have been instrumental in helping me find potential participants. Elizabeth French and Vanessa Ianson have facilitated the compensation of participants. Hannah Sysak and Lindsay Griener have assisted me in solving various problems during collecting data. I also appreciate every individual who participated in my study, especially those reaching out to me.

In addition, at the Department of Linguistics, I also had great pleasure of working with Dr. Stephanie Archer in her Little Magpies Lab, where I have the fondest memory of that beautiful summer. Among my fellow students, Regina Hert, Dalia Cristerna-Roman and Aisha Barise have brought so much joy into my life and I am grateful that they have been part of my journey.

I also extend my gratitude to the funding support offered by Social Sciences and Humanities Research Council of Canada (http://www.sshrc-crsh.gc.ca/) Partnership Grant (Words in the World, 895-2016-1008).

Last but not the least, I want to thank my family, who always give me motivation and support, which was especially meaningful during the COVID-19. I dedicate this thesis to my beloved parents and brother.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

A number of studies have reported challenges in acquiring Mandarin tones faced by adult non-native learners, but most of them have focused on speech perception and production findings (e.g., Wang, Spence, Jongman, & Sereno, 1999; Shen & Froud, 2016). Although some work has used novel word learning tasks to examine acquisition of Mandarin in non-native adults (e.g., Gullberg, Roberts, Dimroth, Veroude, & Indefrey, 2010; Chang & Bowles, 2015), few studies have explicitly used nonwords to directly compare learning of segmental and tonal contrasts by adult learners. Further, as a common occurrence in human languages, homophones have been investigated in novel word learning tasks often among infants and young children (e.g., Storkel & Maekawa, 2005; Dautriche, Fibla, Fievet, & Christophe, 2018), but much less is known about the role of homophones in novel word learning among adults. In addition, evidence suggests that brief periods of unoccupied rest induce memory enhancement in a similar manner to sleep (e.g., Dewar, Alber, Butler, Cowan, & Della Sala, 2012; Brokaw et al., 2016; Wamsley, 2019). However, little research has examined the function of rest in consolidation of novel word learning among both native and non-native learners.

The present study attempts to use visual world eye-tracking to examine how adults from different language backgrounds learn novel words which involve homophones and non-homophones, and vary in types of phonological contrasts (consonant, tone, consonant & tone), and whether brief periods of rest following learning affect the outcome. For this aim, we recruited native Mandarin speakers and native English speakers to carry out a novel word learning task, in which they learned nonwords that incorporated Mandarin segments and tones and then performed two tests separated by a short break during which they either had an eyes-closed rest or completed a distractor task.

This chapter begins by introducing the Mandarin phonological system, and then reviews previous studies on the novel word learning of Mandarin. The third and fourth sections outline prior research on the effects of homophone and competition in lexical processing, respectively. A discussion of memory consolidation following learning is presented in the final section.

1.1 Mandarin phonology

Mandarin Chinese (hereafter Mandarin), also referred to as Standard Chinese, is widely spoken within mainland China, and considered as one of the most frequently used varieties of Chinese around the globe. It is mainly based on the pronunciation of the Beijing dialect, with some lexical and syntactic influences from other dialects. Despite wide differences across dialects, there are common properties that characterize Mandarin phonology. For example, the Mandarin syllable usually has less than four segments, which are combinations of consonants, glides and vowels (Wee & Li, 2015). Moreover, as a tone language, Mandarin exploits pitch variations as a contrastive feature to differentiate between words. The following discusses Mandarin segments and tones individually.

In Mandarin, there are a total of 19 consonants, which are transcribed using the International Phonetic Alphabet (IPA) in Table 1.1 below. The distinction between pairs of stops or affricates is in aspiration, whereas the fricative pair /ʂ, ʐ/ differs in voicing (Lin, 2007). A major feature that sets apart Mandarin speakers from Beijing and from other places is the retroflexes /ʈʂ, $ʈʂ^h$, ʂ, ʐ/, which often cannot be found in non-Beijing dialects (Duanmu, 2007).

Table 1.1 Mandarin consonant phonemes in IPA (adapted from Duanmu, 2007, p. 24)

	Labial	Dental	Retroflex	Velar
Stop	p, pʰ	t, tʰ		k, kʰ
Affricate		ts, tsʰ	ʈʂ, ʈʂʰ	
Fricative	f	s	ʂ, ʐ	x
Nasal	m	n		ŋ
Liquid		l		

Apart from these consonants, there are three glides /j, w, ɥ/ and three palatals /tɕ, tɕʰ, ɕ/ in Mandarin. The glides are generally analyzed as consonantal allophones of the high vowels /i, u, y/ (Duanmu, 2007). The palatals are treated as allophones of the dentals, the retroflexes or the velars, since they can only appear before high front vowels whereas the latter three series can never occur in this position (Norman, 1988).

With respect to vowels, there are five of them in Mandarin: the high vowels /i, u, y/, the mid vowel /ə/, and the low vowel /a/. Among the three high vowels, /i/ and /u/ are similar to those in English, and can be used to end a dipthong, as in /mai/ 'buy' and /mau/ 'cat'; whereas /y/ cannot serve as the second component of a diphthong (Duanmu, 2007). Further, vowels function as the nucleus in Mandarin syllables, which may also include an initial consonant, a medial glide and a coda. However, many of the possible combinations do not exist. As a result, the limited syllable inventory leads to an abundance of homophones in Mandarin.

In addition, vowels are the only Mandarin segments that can carry a tone to distinguish different words. For example, the Chinese word /da/ can mean 'match', 'arrive', 'hit' and 'big' while carrying the four different tones respectively. A tone is typically described in terms of its pitch height and shape. In Figure 1.1 below, the left panel illustrates the fundamental frequency tracings of the four Mandarin tones, produced by a Beijingese radio announcer; the right panel uses the five-point scale proposed by Chao (1930) to denote the pitch onset and offset for each individual tone (Zhu & Wang, 2015). Specifically, tone 1 (T1) is a high level contour transcribed as [55], tone 2 (T2) is a low rising contour as [35], tone 3 (T3) a high falling contour as [214], and tone 4 (T4) a low falling contour as [51]. Further, what is not shown in the figure is a fifth neutral tone, which is a mid level contour represented as [33] and often associated with weak syllables (Chen, 1984). The four lexical tones T1 to T4 will be symbolized with superscript digits in Mandarin syllables below, for example, /ma^{2}/ represents a syllable /ma/ with T2.

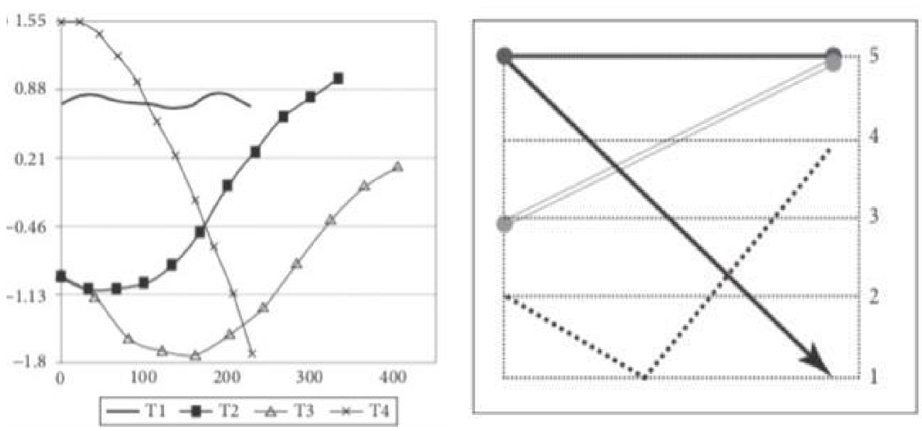


Figure 1.1 The four Mandarin tones. Left: Mean fundamental frequency contours. Right: The five-point scale representation (Zhu & Wang, 2015, p. 507; axis labels are not specified in the source).

Besides the primary cue of pitch, temporal and amplitude properties provide secondary cues to tone perception (Liu & Samuel, 2004; Kong & Zeng, 2006). For example, T3 has significantly longer duration than the other three tones, while T4 is significantly shorter (Chang & Bowles, 2015).

1.2 Novel word learning of Mandarin

A considerable number of studies have investigated the acquisition of Mandarin phonology by adult non-native learners from various language backgrounds, such as native speakers of English (e.g., Lin, 1985; Tsao, Liu, & Kuhl, 2006; Hao, 2018), Japanese (e.g., Tsukada, Kondo, & Sunaoka, 2016), Thai (e.g., Wu, Munro, & Wang, 2014) and Dutch (e.g., Sadakata & McQueen, 2014; Zou, Chen, & Caspers, 2017). A majority of them have reported the difficulty in acquiring Mandarin tones faced by adult non-native learners, especially based on speech perception and production findings (e.g., Wang et al., 1999; Ning, Shih, & Loucks, 2014; Tsukada, Xu, & Rattanasone, 2015; Shen & Froud, 2016). Empirical evidence suggests that prior experience with or knowledge of tone languages may bring subsequent tone learning advantages (e.g., Wayland & Li, 2008; Caldwell-Harris, Lancaster, Ladd, Dediu, & Chirstiansen, 2015; Shittu, 2019), and that adults are capable of learning novel tone contours in a foreign tone language (e.g., Kann, Barkley, Bao, & Wayland, 2008; Chandrasekaran, Krishnan, & Gandour, 2009).

Nevertheless, relatively much less research has been conducted to examine lexical learning of Mandarin through the lens of novel word learning, a task that is more representative of the language acquisition process (Chang & Bowles, 2015). Although there have been some studies that used the novel word learning task, most of them have primarily focused on infants and young children (e.g., Chan et al., 2011; Chen & Liu, 2014), especially comparing monolingual and bilingual infants (e.g., Byers-Heinlein & Werker, 2013; Singh, Poh, & Fu, 2016; Burnham, Singh, Mattock, Woo, & Kalashnikova, 2018; Wewalaarachchi & Singh, 2020). It appears that fewer researchers have addressed novel word learning in adult non-native learners, although several investigations have explored the learning of Mandarin by native Dutch speakers after minimal exposure (Gullberg et al., 2010), and native English speakers after receiving tone-word training (e.g., Showalter & Hayes-Harb, 2013; Chang & Bowles, 2015; Bowles, Chang, & Karuzis, 2016).

For instance, Chang and Bowles (2015) investigated context effects on tone learning in native English speakers by means of a word learning task, which involved Mandarin disyllabic pseudowords associated with line-drawing pictures. A pervasive influence of phonological context on non-native tone learning was found. For example, tones were significantly harder to acquire in disyllables than in monosyllables. Using a similar tonal word learning paradigm in a follow-up study, Bowles and colleagues (2016) examined aspects of an aptitude for tone learning among native English speakers, such as pitch ability, musicality, general cognitive ability and second language aptitude. They discovered that pitch ability measures, especially perceptual acuity in tone discrimination and tone identification, helped predict successful tone learning beyond other measurements. In addition, Showalter and Hayes-Harb (2013) trained native English speakers to associate Mandarin nonwords varying in lexical tones with their written forms and line-drawings of novel objects. Results showed that participants who used tone marks as the orthographic information in learning outperformed those without the availability of orthographic symbols. Thus, it is attested that various factors influence adult native English speakers' ability of tonal learning.

As for novel word learning studies that directly compare learning of tonal and segmental contrasts, very few researchers have explicitly used nonwords to address this issue among adult learners. Therefore, the present work attempts to fill this gap by investigating novel word learning of Mandarin segments and tones among adults from both native and non-native backgrounds. That being said, there are Cantonese studies that used word learning tasks to examine the acquisition of tones (e.g., Cooper & Wang, 2012, 2013), consonants and vowels (e.g., Poltrock, Chen, Kwok, Cheung, & Nazzi, 2018), which may inform our understanding of Mandarin lexical learning and thereby are reviewed as follows.

Cooper and Wang (2013) explored the effect of short-term pitch experience on lexical learning of Cantonese by native English speakers and found that listeners who underwent training obtained a significantly higher proficiency in word identification than those without training. In another study, Cooper and Wang (2012) investigated the influence of linguistic and musical experience on non-native word identification and lexical tone perception. They found that either musical experience or a tone language background (here, Thai) contributed to significantly better word learning proficiency, in comparison to those who had neither musical nor prior tonal experience. Moreover, Poltrock et al. (2018) examined the effect of native phonological knowledge on adult learning of novel words in Cantonese. In a novel word learning task, native Cantonese, Mandarin, and French speakers were trained to learn Cantonese pseudowords that differed minimally by a consonant, a vowel, or a tone. Results indicated that both native Mandarin and French speakers performed worse than native Cantonese speakers on all three contrasts, but Mandarin adults were better at tones than French adults.

1.3 Homophone effects in lexical processing

Homophones, a common occurrence in many human languages, are same-sounding words that differ in meaning. For instance, the English syllable /meɪd/ can denote *maid* or *made*, which have different meanings. According to Ke, Wang, and Coupe (2002), in the Brown Corpus of American English (Kucera & Francis, 1967), 19.9% of the 5010 most frequent words have homophones; in the *Dictionary of Current Chinese* (Chinese Academy of Social Sciences Institute of Linguistics, 1985), 80% of the monosyllabic words have homophones and 55% of them have five or more meanings. An extreme case is the Mandarin syllable /yi^4/[1], which has more than 90 homophones (Ke et al., 2002).

Previous studies on lexical processing have often used lexical decision tasks and reported homophone effects in both alphabetical languages like English and logographic languages like Chinese and Japanese. A typical example of the English research is Rubenstein, Lewis, and Rubenstein (1971), who used lexical decision tasks to compare homophones to non-homophones with different frequencies in visual word recognition. They found that homophones had a significantly longer latency and lower accuracy than non-homophones, particularly in the group of low-frequency words. Rubenstein et al. (1971) attributed this homophony processing disadvantage to less certainty about the orthography of a word when it belongs to a homophone pair, which is especially true of less frequent words.

Another example of English homophone studies is Pexman, Lupker, and Jared (2001). Through a series of experiments, Pexman et al. (2001) concluded that the size of the homophone effects is determined by the frequency of the homophone and of its mate, and by the orthographic and phonological characteristics of the homophone. Moreover, they proposed a feedback mechanism to explain the homophone effects. Specifically, when a homophone is presented, feedback from phonology to orthography activates the orthographic representations for both members of the homophone pair. Thereby, it takes more time for participants to identify the appropriate orthographic representation for the presented word, potentially producing a homophony effect.

---

[1] As stated above in section 1.1, /yi^4/ means a syllable /yi/ carrying T4.

In contrast to the inhibitory effects in English, a processing advantage for homophones is found in research on Chinese. For example, using lexical decision tasks, Ziegler, Tan, Perry, and Montant (2000) observed that response times (RTs) were significantly shorter for Chinese characters with homophonic mates than for characters without homophone mates (also see Chen, Vaid, & Wu, 2009 for similar results). According to them, this facilitative homophone effect is due to a large number of homophonic mates in Chinese, which increase the phonological familiarity of a homophone and aid its processing. From the perspective of the feedback account, homophony is more likely to create competition in English than in Chinese, because homophonic mates in English tend to be more visually similar than their Chinese counterparts.

In order to explore the reason for contradictory homophone effects in previous English and Chinese studies, Hino, Kusunose, Lupker, and Jared (2013) used Japanese kanji words, which are logographic characters like those in Chinese, to examine the relationship between the patterns of homophone effects and script type. They suggested that regardless of script type, the direction of homophony effects is influenced by the number of homophonic mates that the target homophones have. Specifically, Hino et al. (2013) found that Japanese homophones with only a single homophonic mate were processed slowly, resulting in an inhibitory homophone effect as found in English. However, homophones with multiple homophonic mates were processed more rapidly than non-homophones, which led to a facilitatory homophone effect.

In addition to lexical decision tasks, the novel word learning task has also been used in homophone studies. However, previous research has primarily focused on homophony learning in infants and young children (e.g., Storkel & Maekawa, 2005; Dautriche, Chemla, & Christophe, 2016; Ramachers, Brouwer, & Fikkert, 2017; Dautriche et al., 2018). To our knowledge, no study has so far systematically investigated how adults acquire homophones through novel word learning tasks. Thus, the present work will try to fill this literature gap by addressing adult homophony learning in the novel word learning task.

1.4 Competition effects in spoken word processing

The visual world paradigm is an eye-tracking method to study online spoken language processing. It is frequently used to measure the time course of spoken word recognition in continuous speech, while participants' eye movements to objects or words in a visual workspace are monitored as the speech signal unfolds (Salverda & Tanenhaus, 2017). The structure of the visual world varies across studies, but usually consists of the target, one or more competitors, and unrelated distractors. Specifically, target is the object of interest in the visual scene, which is the referent of the linguistic stimuli. Competitors are objects that are related to the target along some specified dimension. For instance, the competitor may share some common features with the target phonologically or visually (Salverda & Tanenhaus, 2017). An object that is unrelated to the target is labelled as a distractor. Competition is often indicated when participants' looks to the target diverge to the other objects.

A major line of visual world research has focused on phonological processing on the word level. Earlier behavioral evidence has revealed a standard competition effect of phonological forms overlapping at word onset (e.g., Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989). For example, these studies indicate that on hearing the word "beaker", its cohort word "beetle" is also activated. Allopenna, Magnuson, and Tanenhaus (1998) used visual world eye-tracking to examine whether competition effects would be observed also for words that rhyme with the target but do not share an onset. In their study, the visual display was line drawings of four objects on a computer screen, which included a target (e.g., the referent "beaker"), an onset cohort competitor (e.g., "beetle"), a rhyme competitor (e.g., "speaker"), and an unrelated distractor (e.g., “dolphin”). Participants were instructed to move the objects to specified locations (e.g., “Pick up the beaker. Now put it below the square.”) while their gaze to the objects on the display was monitored. Allopenna et al. (1998) found that in comparison to the unrelated distractor, both cohort and rhyme objects competed for lexical activation, although cohorts competed earlier and more strongly for overt attention than rhymes. Thus, their results lend support to the hypothesis that speech input is continuously mapped to potential lexical representations as it unfolds over time.

In addition to segmental information, spoken word recognition also relies on suprasegmental information, which plays a significant role in tonal languages. However, there is no general agreement on the contribution of segmental and tonal information in Mandarin processing. Some studies suggest that tonal information might be a weaker cue for word recognition than segmental information (e.g., Taft & Chen, 1992; Tong, Francis, & Gandour, 2008; Li, Lin, Wang, & Jiang, 2013; Li, Wang, Davis, & Guan, 2019). For example, Taft and Chen (1992) asked native speakers of Mandarin to decide whether two presented Chinese characters were homophonic as they listened to pairs of Mandarin words. They found that participants’ responses were significantly slower and less accurate when the pronunciation of the two words differed only in tones than in vowels. Also using the homophone judgement task, Li et al. (2019) addressed the segmental and tonal representation and processing in visual Chinese words among both native and non-native Mandarin speakers. They found that both groups relied more on segmental features than on tonal aspects of lexical entries in Mandarin phonological judgement, but this reliance was greater in non-native speakers, as indicated by their more errors and longer RTs.

However, there is also research showing a parallel processing of tonal and segmental information in word recognition (e.g., Schirmer, Tang, Penney, Gunter, & Chen, 2005; Malins & Joanisse, 2010; Zhao, Guo, Zhou, & Shu, 2011). For instance, Malins and Joanisse (2010) used eye-tracking to monitor listeners' looks to pictures in a visual array while the auditory stimuli were unfolding. Participants were asked to listen to a monosyllabic Mandarin word and match it to one of four pictures in the display, including the target item (/ʈʂ^{h}ua^{2}ŋ/ 'bed'), a phonological competitor (segmental: /ʈʂ^{h}ua^{1}ŋ/ 'window'; cohort: /ʈʂ^{h}ua^{2}n/ 'ship'; rhyme: /xua^{2}ŋ/ 'yellow'; tonal: /niu^2/ 'cow') and two phonologically unrelated distractors. Eye movement data indicated a significant delay in looks to the target in the segmental and cohort conditions, suggesting competition between targets and competitors for both types of phonological relationships. A follow-up analysis of the time course of change in target looks revealed nearly identical trajectories for the two conditions. Thus, the authors concluded that tonal and segmental information are accessed concurrently and play comparable roles in constraining lexical access.

Furthermore, homophones have been investigated in a number of visual world paradigm studies (e.g., Meyer, Belke, Telling, & Humphreys, 2007; Wang, Wang, & Malins, 2017; Yip & Zhai, 2018). For example, Meyer et al. (2007) designed a visual search experiment, in which participants were asked to decide whether the target was present or absent in a four-object search display. There were two types of related competitors: objects with homophonous names (e.g., *buoy* for the target *boy*) and semantically related objects (e.g., *shirt* for the target *trousers*), but only one type occurred in the display at most. Participants' RTs and eye movements indicated that the homophone competitors had a similar level of competition effects as the semantic competitors. Moreover, Yip and Zhai (2018) investigated lexical ambiguity resolution using Chinese homophones embedded at the end of sentences with different preceding contexts. In their study, participants were told to listen to the sentence and look at different Chinese characters or line-drawing pictures on the display. Eye movement data demonstrated that on hearing a Chinese homophone, listeners automatically and rapidly used the context to select the appropriate meaning for the ongoing language processing.

1.5 Memory consolidation following learning

Memory consolidation is a process in which new memory traces become strengthened against interference and can be retrieved at a later point in time (Dewar et al., 2012). Previous studies have shown that sleep following learning facilitates consolidation of new memories, relative to an equivalent duration of wake (Stickgold, 2005). This sleep benefit is also found in lexical learning research (e.g., Fenn, Nusbaum, & Margoliash, 2003; Kurdziel & Spencer, 2016). For example, Fenn et al. (2003) used a speech learning task, where individuals listened to monosyllabic consonant-vowel-consonant words produced by a speech synthesizer while observing their printed version on screen and then were asked to recognize new words after the training. Results suggested that participants had significantly greater accuracy in the post-test after a 12-hour period including regular sleep than after a 12-hour period of wakefulness. Instead of using artificial language stimuli, Kurdziel and Spencer (2016) asked native English-speaking adults to learn very low-frequency English words in a novel word learning task and recall them after a 12-hour delay. They found that across a delay that included sleep, individuals were significantly better at recalling novel words than when the delay was spent awake. Therefore, sleep is also beneficial for retention of newly learned words within a native language.

Further, under the question of whether sleep is really required to obtain the memory benefits, a number of studies have demonstrated that brief periods of unoccupied waking rest induce memory enhancement in a similar manner to sleep (e.g., Dewar et al., 2012; Brokaw et al., 2016; Humiston & Wamsley, 2018; Wamsley, 2019). For instance, in the study of Brokaw et al. (2016), participants listened to a story and performed a delayed recall test following a 15-minute interval, during which they either had an eyes-closed rest or completed a distractor task. Observations suggested that a short period of waking rest can facilitate verbal declarative memory. Additionally, Humiston and Wamsley (2018) used the same procedures to test whether waking rest benefits procedural memory. Individuals were trained on a motor sequence task prior to either resting with their eyes closed or completing a distractor task. After a retention interval, participants were tested on motor sequence task performance either immediately or four hours later. Results confirmed the benefits of waking rest in facilitating motor-procedural memory at an immediately following test, but not when tested after a four-hour delay. However, none of the aforementioned studies address novel word learning.

In sum, a brief period of unoccupied rest after learning facilitates memory consolidation, relative to an equivalent period of time spent engaged in sensorimotor or cognitive tasks (Wamsley, 2019). This effect parallels the function of sleep in memory consolidation and reactivation. However, none research has examined the role of rest in consolidation of novel word learning in both native and non-native speakers. Therefore, the current study will investigate whether a short period of rest enhances performance in a post test of novel word learning among speakers from different language backgrounds, in comparison to completing a distractor task for an equivalent duration.

Chapter 2: Present Study

As mentioned in Chapter 1, many studies on the acquisition of Mandarin tend to focus on the speech perception and production of tones. However, few researchers have used the novel word learning task to directly compare learning of tonal and segmental contrasts among adults from different language backgrounds. Further, little is known about how adults learn homophones in the frame of novel word learning, and whether short periods of rest influence their outcome. Therefore, there is a need for studies exploring phonological contrasts and homophony in novel word learning among adults from both native and non-native backgrounds.

Due to its strengths that complement more traditional methods, the visual world eye-tracking paradigm has been extensively used to examine online spoken word processing. As the auditory stimuli unfold, participants' eye movements to potential referents in the display reflect their developing interpretation of the linguistic input (Dahan & Tanenhaus, 2005). Moreover, time course can be assessed without interrupting the spoken utterance, and the listener's interpretation can be inferred without requiring a metalinguistic decision (Dahan & Tanenhaus, 2005).

In the current study, we used the visual world eye-tracking paradigm to investigate how adults from native and non-native backgrounds learn novel words which include homophones and non-homophones and differ in types of phonological contrasts. We were further interested in the role of brief periods of rest in consolidation of novel word learning. To this end, we studied 34 native English speakers and 34 native Mandarin speakers through a novel word learning task, where participants learned novel words in a pair with consonant contrasts, tone contrasts, or consonant & tone contrasts. There were a total of 28 novel words that integrated Mandarin segments and tones, and seven of them were homophones (six had two meanings, one had three meanings).

The novel word learning task consisted of three sections: Learning Phase/Test Phase I, a break, and Test Phase II. In the integrated Learning Phase/Test Phase I, participants were asked to learn novel words through word-object associations and select the target object that matched the novel word they heard during the test. Then there was a short break, in which participants either rested with their eyes closed or played a computer game for 15 minutes. In Test Phase II, for each trial, the target was presented in the visual array either with a competitor (which differed from the target minimally by a consonant, a tone, or both) and a distractor (i.e. the competitor condition), or with two distractors (i.e. the no-competitor condition). Participants had to choose the target object that corresponded to the novel word they heard. Thereby, competition was considered as an indicator of participants' learning success. In addition to response accuracy and RTs recorded in Test Phases I and II, participants' eye movements in Test Phase II were analyzed as well.

The following research questions were addressed in the present study:

1. Does participants' language background modulate their learning outcome? In particular, does it predict their response accuracy and RTs in Test Phases I and II, as well as their eye movements in Test Phase II?
2. Is learning affected by whether the word is a homophone or a non-homophone? Does this depend on participants' language background?
3. Do phonological contrasts entailed in the novel word pairs predict participants' response accuracy and RTs in Test Phases I and II, as well as their eye movements in Test Phase II? Does this depend on their language background?

4. Do different types of break predict participants' response accuracy, RTs and eye movements in Test Phase II? Does this depend on their language background?

Based on previous literature, we hypothesized that:

1. Participants' language background will affect their learning outcome, and will significantly predict their response accuracy, RTs and eye movements. Native Mandarin speakers will have significantly faster and more accurate responses, and more target looks than native English speakers.
2. Given contradictory homophone effects found in previous Chinese and English research (as reviewed in section 1.3 above), it is likely that homophony will interact with participants' language background to predict their RTs, and it seems possible that the interaction will also affect their response accuracy and eye movements. The two language groups will probably show different tendencies.
3. Phonological contrasts will interact with participants' language background to significantly predict their response accuracy, RTs and eye movements. For native English speakers (as reviewed in sections 1.2 and 1.4 above), words learned in a pair with tone contrasts will probably receive the slowest and least accurate responses.
4. Different types of break will significantly predict participants' response accuracy, RTs and eye movements in Test Phase II. The rest group will have faster and more accurate responses, and more target looks than the game group.

Chapter 3: Method

3.1 Participants

Seventy adults who were students or staff at the University of Alberta, Edmonton, participated in this study. Sixty-eight of them were university students recruited through the Linguistics Sign-up System and received partial course credit in exchange for their participation. The other two participants were university staff and were paid 15 Canadian dollars as compensation. Informed consent was collected from all participants prior to their participation. Two student participants were excluded from the analysis because one of them was a native French speaker and the other one did poorly in the novel word learning task with the accuracy rate below the chance level. Therefore, a total of 68 participants (46 female; M_{age} = 22.0 years; range = 18-63; SD = 7.35) were included in the final analysis.

Thirty-four of these participants (21 female; M_{age} = 23.5 years; range = 18-63; SD = 10.08) were native English speakers who were from Canada (n=33) and the United States of America (n=1), and the other 34 participants (25 female; M_{age} = 20.5 years; range = 18-25; SD = 1.86) were native Mandarin speakers who were all born in mainland China. All participants had normal or corrected-to-normal vision and normal hearing based on self-report. The experiment protocol and consent procedures were reviewed and approved by the Research Ethics Board 2 of the University of Alberta (Study ID: Pro00089301).

3.2 Materials

3.2.1 Novel word stimuli

A total of 18 pairs of disyllabic novel words carrying Mandarin tones were used in the novel word learning task (see Table 3.1). The first syllables of these novel words were created by integrating the six consonants /p^h, t^h, k^h, m, n, l/, the three vowels /a, u, i/ and the four lexical tones in Mandarin, while the second syllables were constantly /sa^1/ with T1. Phonological combinations that are real words in Mandarin were avoided. Thus, a total of 28 novel words were generated, of which seven were homophonic. Among these homophones, six had two meanings: /$t^ha^1sa^1$/, /$k^hu^1sa^1$/, /mu^2sa^1/, /ni^2sa^1/, /li^3sa^1/ and /li^4sa^1/, and one, /$p^ha^1sa^1$/, had three meanings.

The 18 pairs consisted of three different groups with respect to phonological contrasts. First, the tone contrasts group contained novel word pairs that contrasted in the tone of the first syllable only and presented all six possible tonal comparisons in Mandarin: T1-T2, T1-T3, T1-T4, T2-T3, T2-T4 and T3-T4. Second, in the consonant contrasts group, novel word pairs differed in the consonant of the first syllable only: /p^h/ vs. /k^h/, /t^h/ vs. /m/, /k^h/ vs. /n/, /m/ vs. /l/, /n/ vs. /t^h/ and /l/ vs. /p^h/. Additionally, T1 and T4 were used more than once in this group because previous studies (e.g., Kiriloff, 1969; Shen & Lin, 1991; Hao, 2012) suggested that these two tones tend to be less difficult for non-native speakers to learn, and were also served to balance the number of pairs compared to other groups. Lastly, the consonant & tone contrasts group incorporated both consonantal and tonal differences between the first syllables within the pair. The same six possible tonal comparisons but a different set of consonant contrasts were presented, so the novel word pairs contrasted in a way distinct from those in the previous two groups. The experimental words are listed in Table 3.1 below.

Table 3.1 Novel words used in the novel word learning task (homophones are in bold).

<table>
<tr><th>Phonological Contrasts</th><th colspan="3">Novel Word Pairs</th></tr>
<tr><td rowspan="6">Tone Contrasts</td><td>1</td><td>/pʰa¹sa¹/</td><td>/pʰa²sa¹/</td></tr>
<tr><td>2</td><td>/tʰa¹sa¹/</td><td>/tʰa³sa¹/</td></tr>
<tr><td>3</td><td>/kʰu¹sa¹/</td><td>/kʰu⁴sa¹/</td></tr>
<tr><td>4</td><td>/mu²sa¹/</td><td>/mu³sa¹/</td></tr>
<tr><td>5</td><td>/ni²sa¹/</td><td>/ni⁴sa¹/</td></tr>
<tr><td>6</td><td>/li³sa¹/</td><td>/li⁴sa¹/</td></tr>
<tr><td rowspan="6">Consonant Contrasts</td><td>1</td><td>/pʰa¹sa¹/</td><td>/kʰa¹sa¹/</td></tr>
<tr><td>2</td><td>/tʰa²sa¹/</td><td>/ma²sa¹/</td></tr>
<tr><td>3</td><td>/kʰu³sa¹/</td><td>/nu³sa¹/</td></tr>
<tr><td>4</td><td>/mu⁴sa¹/</td><td>/lu⁴sa¹/</td></tr>
<tr><td>5</td><td>/ni¹sa¹/</td><td>/tʰi¹sa¹/</td></tr>
<tr><td>6</td><td>/li⁴sa¹/</td><td>/pʰi⁴sa¹/</td></tr>
<tr><td rowspan="6">Consonant & Tone Contrasts</td><td>1</td><td>/pʰa¹sa¹/</td><td>/na²sa¹/</td></tr>
<tr><td>2</td><td>/tʰa¹sa¹/</td><td>/la³sa¹/</td></tr>
<tr><td>3</td><td>/kʰu¹sa¹/</td><td>/pʰu⁴sa¹/</td></tr>
<tr><td>4</td><td>/mu²sa¹/</td><td>/tʰu³sa¹/</td></tr>
<tr><td>5</td><td>/ni²sa¹/</td><td>/mi⁴sa¹/</td></tr>
<tr><td>6</td><td>/li³sa¹/</td><td>/kʰi⁴sa¹/</td></tr>
</table>

In order to familiarize the participants with the task, two further novel words /musa/ and /t^husa/ with the Mandarin neutral tone in both syllables were used in a practice trial. All of the novel words were pronounced by the author, a 24-year-old male native Mandarin speaker who was born and raised in Northern China. They were recorded through a Fostex field recorder (Model FR-2LE, 16-bit resolution, sampling frequency 44,100 Hz) in a WhisperRoom sound booth at the *Prosody Lab* at the University of Alberta, and then manually split into separate sound files using Praat (Boersma & Weenink, 2020).

3.2.2 Novel object stimuli

A total of 47 novel objects were selected from the Novel Object and Unusual Name (NOUN) Database (Horst & Hout, 2016). The NOUN database is a collection of novel object pictures that are hard to define because of their complicated color, shape and material configurations. The novelty of objects was confirmed by self-report and a lack of consensus on questions asking participants to name them in a previous study by Horst and Hout (2016). In the present study, the above-mentioned novel words in section 3.2.1 were randomly assigned as names of the selected novel objects for participants to learn the novel word-object mappings. In addition to the two novel objects used in the practice trial, the remaining 45 objects were used in the experiment trials, 36 as target and competitor objects and nine as distractor objects (see Appendix A for a complete list of the novel word-object mappings, and Appendix B for the novel objects used as distractors). A few examples are given in Figure 3.1 below.

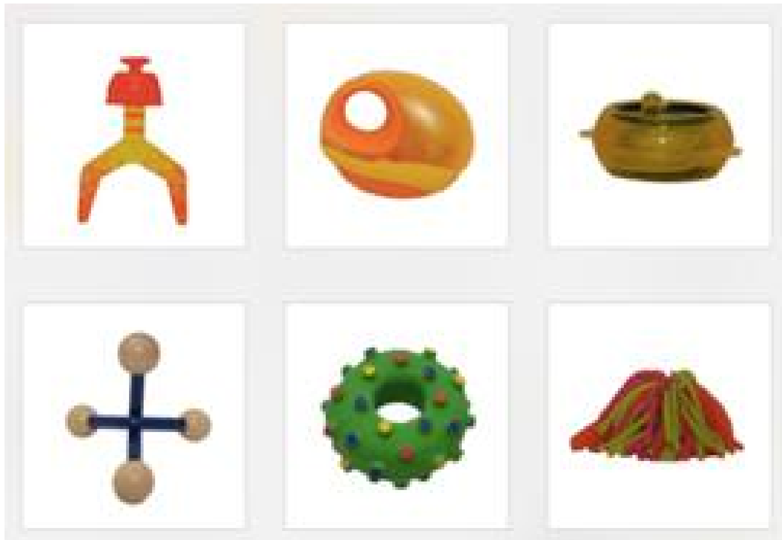


Figure 3.1 Examples of novel objects used in the novel word learning task (selected from the NOUN database, Horst & Hout, 2016).

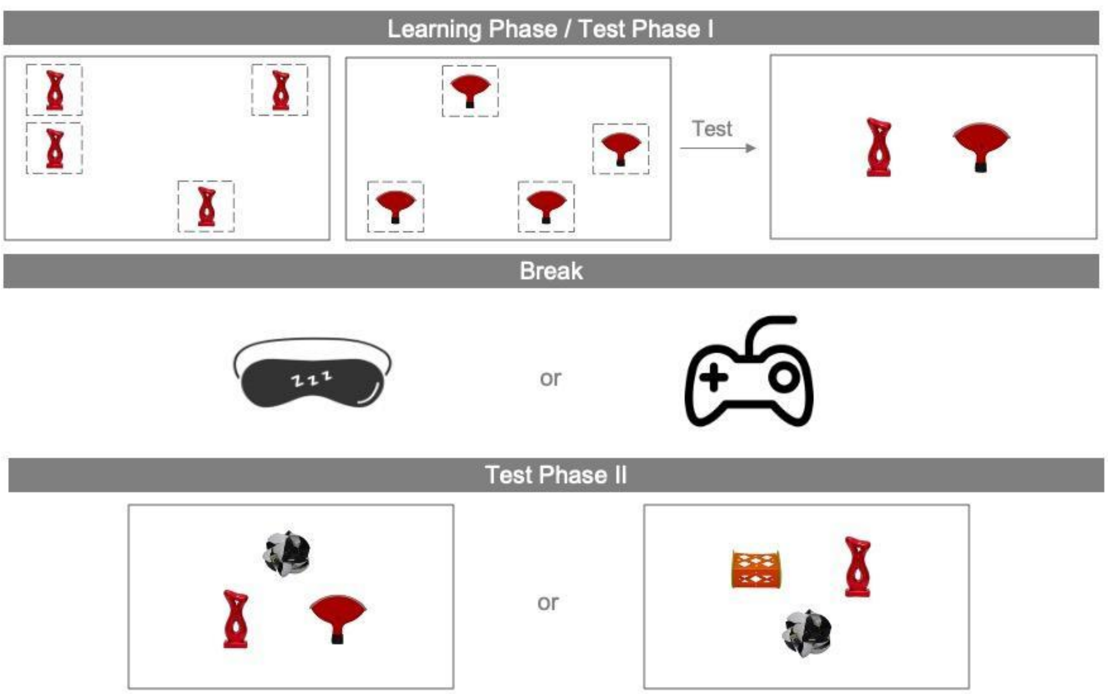


Figure 3.2 Structure of the novel word learning task

3.3 Novel word learning task

All participants were asked to complete a novel word learning task and were tested individually at the *Centre for Comparative Psycholinguistics* at the University of Alberta. Built with Experiment Builder (Version 1.10.1630; SR Research, 2015), the novel word learning task used the visual world paradigm to investigate participants' eye movements to objects in the visual scene as they listened to related novel words. As illustrated in Figure 3.2 above, the task consisted of the following three sections: Learning Phase/Test Phase I (see section 3.3.1), Break (see section 3.3.2), and Test Phase II (see section 3.3.3). It took approximately 35 minutes for each participant to complete the task.

3.3.1 Learning Phase/Test Phase I

The first section included a learning phase and an integrated test phase. Participants were seated in a sound-attenuated booth, resting their heads on a chin rest and looking at the visual stimuli presented on a display monitor (24 inches, 1920 x 1080 pixel resolution). Their eye movements were recorded using an SR-Research EyeLink 1000 desktop mount eye tracker, which was run on a host DELL PC with a sampling rate of either 500 Hz or 1000 Hz. At the beginning, participants were told that they would see an object and hear its name as it moved on the display, and that after observing a pair of objects and learning their names, they would be tested on selecting the correct object (i.e. the target object) that matched the name they heard in the end (as depicted at the top section of Figure 3.2 above). Participants were also instructed to pay close attention to the auditory stimuli because they sounded similar, and to select the target object by pulling the corresponding trigger on a Microsoft Sidewinder gamepad as quickly as possible.

Subsequently, a five-point calibration and cross-validation was performed on each participant. Within each trial, participants listened to a pair of novel words in the learning phase and were tested on only one word in the integrated test phase. Specifically, the learning phase started with a novel object's picture appearing at one location on the screen, which was equally divided into 12 areas, and the object's name was played auditorily on a speaker with a 200ms offset. At each location, the novel word audio was played twice with a 1000ms interval in between and a 2000ms pause in the end. Then the novel object moved to another spot. In total, the novel object appeared randomly at four different locations on the display. Afterwards, in a similar manner, another novel object's image appeared at four locations, with its name being auditorily played on the speaker twice at each position. Following the learning phase, there was a "+" crosshair presented in the screen center for 2000ms. Then in Test Phase I, pictures of the previous two novel objects appeared simultaneously at the two central locations on the screen, with their order (left/right) randomly selected and balanced across all trials. However, only one novel word was played auditorily with a 200ms offset, and there was a 1000ms pause before responses could be made using the gamepad. Participants pulled the left trigger of the gamepad if they considered the novel word matching the left object on the screen, and pulled the right trigger if they believed that the novel word matched the right object. The tested novel word was labelled as the target word (referring to the target object) in this study, and the other paired novel word as the competitor word (referring to the competitor object).

Prior to the experimental trials, participants completed a practice trial to familiarize themselves with the task. Then they performed a total of 18 study trials (see Appendix C for a complete list of the novel word pairs and target novel words used in the 18 trials). Learning

Phase/Test Phase I took around ten minutes in total to complete. Participants' responses and RTs were recorded, and their eye movement data were collected with the eye tracker.

3.3.2 Break

The second section was a 15-minute break, in which participants were randomly assigned into either a game group or a rest group as described in the middle part of Figure 3.2 above. In the game group, participants remained in the sound booth and played a computer game called "Snood" (Dobson, 1996). Snood is a puzzle game where participants clear blocks of colors by joining three or more icons of the same color. This visuospatial task was chosen because it is engaging and involves only minimal hand and eye movement. Participants were instructed how to play the game and asked to continue playing for the entire 15 minutes. In the rest group, participants lay down or sat on a comfortable reclining chair outside of the sound booth and were asked to put on an eye-mask provided to them. They were told to keep their eyes closed for the entire 15 minutes and that they could fall asleep if they would. In order to minimize the distraction to participants in the rest group, the experimenter took away their mobile phones and left the room for the duration of the entire break.

3.3.3 Test Phase II

The last section was Test Phase II, which is illustrated in the bottom part of Figure 3.2 above. Participants went back to or stayed in the sound booth and redid the calibration and cross-validation before the test. Then they were told that they would see three novel objects on the screen, some of which were the ones they had learned earlier, but would hear only one name of them. Participants were instructed to click the target object that matched the name they heard with a mouse as quickly as possible.

At this time, six new positions in the screen center were generated, three of which built an upright triangle (see the left panel of Test Phase II in Figure 3.2 above) while the other three constituted an upside-down one (see the right panel of Test Phase II in Figure 3.2). Each trial started with three novel objects appearing simultaneously, and one novel word was played auditorily on the speaker with a 200ms offset. There was a 1000ms interval before the participant could respond. In order to explore whether there was a visual world competition effect, the target object was presented either with or without its competitor, with which it was paired during the learning because they were phonologically contrastive in consonants, tones, or both. Specifically, in the presence of a competitor, the target object appeared with the competitor as well as an unrelated distractor object. However, if there was no competitor, the target object appeared with two unrelated distractor objects that they had not encountered during the learning phase. Each novel word of the 18 pairs in Table 3.1 above served as the target once and as the competitor once.

Since each novel word occurred within two visual settings, with and without the competitor, each participant completed a total of 72 study trials (see Appendix D for a complete list of the target novel words used in the 72 trials). There were no practice trials in this phase. Test Phase II took around ten minutes in total to complete. Participants' responses and RTs were recorded, and their eye movement data were collected with the eye tracker.

Chapter 4: Results

For each participant, we analyzed their response accuracy and response latencies in Test Phases I and II, as well as their eye movements in Test Phase II, which are reported in the following sections 4.1, 4.2 and 4.3, respectively.

4.1 Response accuracy

In terms of participants' response accuracy, there were 1224 (18 trials x 68 participants) and 4896 (72 trials x 68 participants) data points in Test Phases I and II, respectively. Response accuracy was coded as a binary variable (correct vs. incorrect), and analyzed as the dependent variable in a binomial generalized linear mixed-effects model constructed using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in R (Version 3.6.1; R Core Team, 2019). The estimated *p*-values were obtained using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). The input variables of the model consisted of the fixed-effect structure and the random-effect structure. Regarding the fixed-effect structure, it was: *Language Background* (native Mandarin vs. native English), *Homophone* (yes vs. no) and *Phonological Contrast* (consonant contrasts vs. tone contrasts vs. consonant & tone contrasts) in Test Phase I; it included two additional factors: *Competition* (competitor vs. no-competitor in the visual stimuli) and *Break Type* (game vs. rest) in Test Phase II. Regarding the random-effect structure, it was the same in Test Phases I and II, which is described in detail below.

In order to arrive at the optimal model, a backward fitting model comparison approach was employed. Using the *anova* function, a complex model was compared to its simpler version which had one component removed, by inspecting the estimated *p*-value and Akaike Information Criterion (AIC) values. The complex model was favored only when the difference was significant as indicated by the *p*-value (smaller than the conventional alpha level of 0.05) and when it provided a better fit for the data as indicated by the AIC value. Otherwise, we selected the simpler model. The fixed-effect and random-effect structures were fitted separately. With respect to the fixed effects, we started with a full model where all factors and their interactions were included, and progressed in the stepwise backward way by means of the *anova* model comparison method. As for the random effects, we settled on a random intercept for *Subject* and a random intercept for *Word* in Test Phases I and II, because the by-subject random slopes for *Homophone* and *Phonological Contrast* did not improve the model significantly in Test Phase I and models with these random slopes failed to converge in Test Phase II (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). The final models for Test Phases I and II are reported in the following subsections 4.1.1 and 4.1.2, respectively.

Moreover, to better visualize the accuracy data, accuracy proportions were calculated for each participant as the proportion of trials in which the participant correctly selected the target object out of the total number of experimental trials for this participant in Test Phases I and II, respectively. The effects of different predictors and interactions on accuracy proportions were plotted using the *ggplot2* package (Wickham, 2016).

4.1.1 Test Phase I

In Test Phase I, model comparisons indicated that *Language Background* (AIC = 352.61 vs. 368.30) was the only significant predictor affecting response accuracy. The final model is reported in Table 4.1 below. According to its fixed-effects summary, native English speakers were significantly less accurate than native Mandarin speakers (p = 0.000104). See Figure 4.1 below.

Table 4.1 Binomial generalized linear mixed-effects model with *Language Background* fitted to response accuracy in Test Phase I.

Model formula: Accuracy ~ Language Background + (1 | Subject) + (1 | Word)

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	5.6876	0.7368	7.719	1.17e-14 ***
Language Background_Native English	-1.7469	0.4501	-3.881	0.000104 ***

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

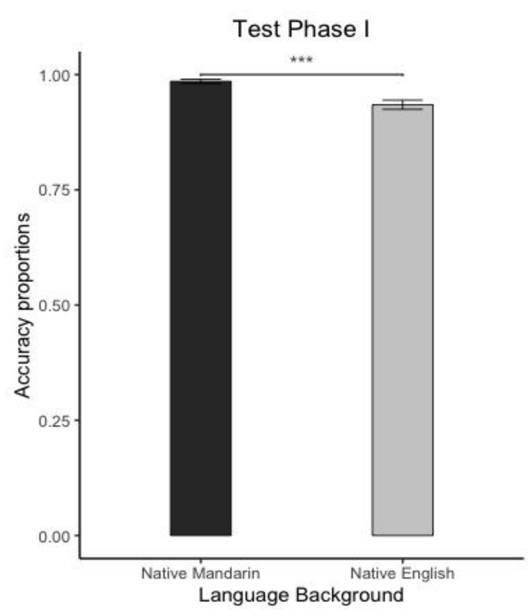


Figure 4.1 Effect of *Language Background* (*x*-axis) on accuracy proportions (*y*-axis) in Test Phase I. Error bars represent standard error of mean, and asterisks indicate a significant difference.

4.1.2 Test Phase II

The final model, along with a summary of its fixed effects, is reported in Table 4.2 below. In Test Phase II, the best fit model contained significant predictors: *Language Background*, *Break Type* and *Competition* (AIC = 5497.5 vs. 5577.1). In addition, there were significant interactions between *Language Background* and *Phonological Contrast* (AIC = 5505.8 vs. 5512.3), *Homophone* and *Language Background* (AIC = 5505.8 vs. 5508.9), and *Break Type* and *Language Background* (AIC = 5497.5 vs. 5501.3). A model containing a three-way interaction between *Language Background*, *Homophone* and *Phonological Contrast* failed to converge, thus a simpler model without it was selected.

Table 4.2 Binomial generalized linear mixed-effects model with *Language Background*, *Homophone*, *Phonological Contrast*, *Competition* and *Break Type* fitted to response accuracy in Test Phase II.

Model formula: Accuracy ~ Homophone*Language Background + Language Background*Phonological Contrast + Language Background*Break Type + Competition + (1 | Subject) + (1 | Word)

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	0.99240	0.21272	4.665	3.08e-06 ***
Homophone_Yes	0.25774	0.20566	1.253	0.21013
Language Background_Native English	-0.99183	0.24603	-4.031	5.55e-05 ***
Phonological Contrast_Cons Contrast	0.20159	0.18581	1.085	0.27796
Phonological Contrast_Ton Contrast	-0.09888	0.13990	-0.707	0.47970
Break Type_Rest	-0.43895	0.21658	-2.027	0.04269 *
Competition_No	0.60361	0.06733	8.965	< 2e-16 ***
Homophone_Yes:Language Background_Native English	0.32164	0.14576	2.207	0.02734 *
Language Background_Native English:Phonological Contrast_Cons Contrast	0.47892	0.17147	2.793	0.00522 **
Language Background_Native English:Phonological Contrast_Ton Contrast	-0.01171	0.16393	-0.071	0.94306
Language Background_Native English:Break Type_Rest	0.74622	0.30346	2.459	0.01393 *

Note: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The effects of significant predictors and interactions based on the model are further discussed below.

4.1.2.1 Interaction between language background and phonological contrast

First of all, we found a significant interaction between *Language Background* and *Phonological Contrast* affecting response accuracy. As presented in Figure 4.2 below, the two language groups showed different patterns. Multiple comparisons using the Tukey's test in the *emmeans* package (Lenth, 2019) suggested that in the English group, novel words learned in a pair with consonant contrasts had significantly higher accuracy than those learned in a pair with consonant & tone contrasts (p = 0.0028), as well as than those learned in a pair with tone contrasts (p = 0.0001). However, there were no significant differences within the Mandarin group, or between the two language groups.

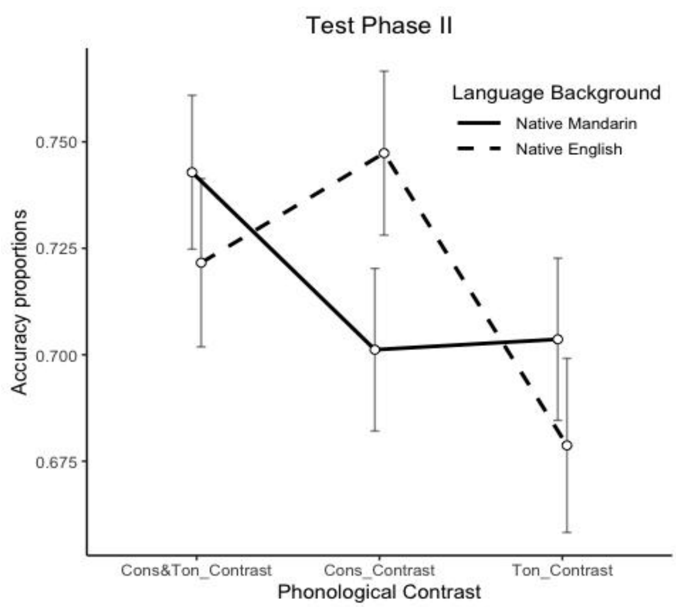


Figure 4.2 Interaction between *Language Background* and *Phonological Contrast* (x-axis) predicting accuracy proportions (y-axis) in Test Phase II. Error bars represent standard error of mean.

Thus, we concluded that native English speakers learned novel word pairs with consonant contrasts well, but their outcome was impaired when tones were involved.

4.1.2.2 Interaction between homophone and language background

We also found a significant interaction between *Homophone* and *Language Background* predicting response accuracy, which is visualized in Figure 4.3 below. Homophones had higher accuracy than non-homophones in both language groups, but subsequent multiple comparisons demonstrated that the difference was only significant in native English speakers (p = 0.0225). With the less-accurate non-homophones, native Mandarin speakers were significantly more accurate than native English speakers (p = 0.0225).

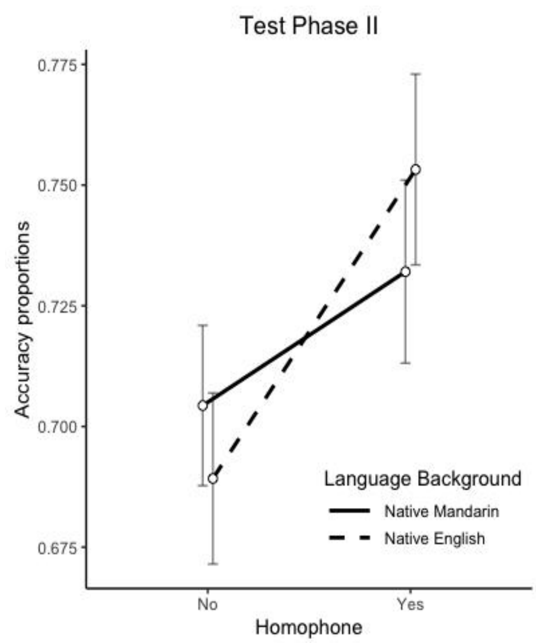


Figure 4.3 Interaction between *Homophone* (x-axis) and *Language Background* predicting accuracy proportions (y-axis) in Test Phase II. Error bars represent standard error of mean.

4.1.2.3 Interaction between break type and language background

In addition, there was a significant interaction between *Break Type* and *Language Background* affecting response accuracy. As can be seen from Figure 4.4 below, a strong contrast is shown between the two language groups: native English speakers had higher accuracy after they rested with their eyes closed than when they played the computer game, though the difference was not significant; whereas native Mandarin speakers who played the game performed better than those who had a rest, despite an insignificant difference in accuracy as well. Multiple comparisons indicated that in the game condition, native Mandarin speakers were significantly more accurate than native English speakers (p = 0.0096), but there was no significant difference between the two language groups in the rest condition.

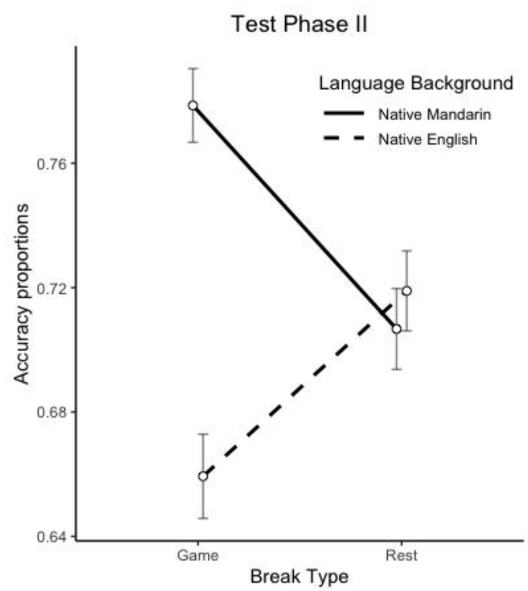


Figure 4.4 Interaction between *Break Type* (x-axis) and *Language Background* predicting accuracy proportions (y-axis) in Test Phase II. Error bars represent standard error of mean.

Therefore, we concluded that only English speakers benefited from a short period of unoccupied rest, while Mandarin speakers took advantage of the distractor task following learning.

4.1.2.4 Competition effects

Lastly, we found a significant effect of *Competition* on response accuracy. Pairwise comparisons showed that novel words presented without a competitor in the visual array received significantly more accurate responses than those tested in the setting with a competitor ($p < 0.0001$). See Figure 4.5 below. No significant interactions involving *Competition* were found.

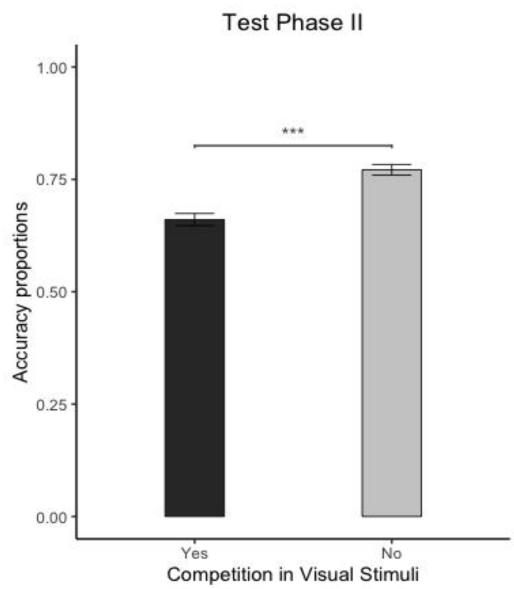


Figure 4.5 Effect of *Competition* (*x*-axis) on accuracy proportions (*y*-axis) in Test Phase II. Error bars represent standard error of mean, and asterisks indicate a significant difference.

4.2 Response times

Only RTs from correct trials were analyzed in this study. RTs deviating beyond the 1.5 interquartile range from the mean RT were treated as outliers, and hence 5.79% and 4.56% of the data were excluded from Test Phases I and II, respectively. As a consequence, there were a total of 1107 and 3345 data points involved in the final analysis of Test Phases I and II, respectively. In order to more closely approximate a normal distribution, log-transformed RTs were used as the response variable in a linear mixed-effects model constructed using the *lme4* package (Bates et al., 2015). The estimated *p*-values of the model were obtained using the *lmerTest* package (Kuznetsova et al., 2017). For the fixed effects, the input variables of the model in Test Phases I and II were the same as those in analyzing the accuracy data, and the same backward stepwise model fitting approach was used to attain the best fit model. For the random effects, we also settled on a random intercept for *Subject* and a random intercept for *Word* in Test Phases I and II, as models with by-subject random slopes for *Homophone* and *Phonological Contrast* failed to converge in both phases (Matuschek et al., 2017). The final models for Test Phases I and II are reported in the following subsections 4.2.1 and 4.2.2, respectively.

4.2.1 Test Phase I

In Test Phase I, model comparisons demonstrated that neither *Language Background*, *Homophone* nor *Phonological Contrast* was a significant predictor affecting log-transformed RTs, so we did not perform further analysis for response latencies in Test Phase I.

4.2.2 Test Phase II

The final model, along with a summary of its fixed effects, is reported in Table 4.3 below. In Test Phase II, the best fit model contained significant predictors: *Homophone*, *Phonological Contrast* and *Competition* (AIC = 3052.5 vs. 3069.8). In addition, significant two-way interactions were found between *Homophone* and *Language Background*, *Homophone* and *Phonological Contrast*, and *Language Background* and *Phonological Contrast*. These were qualified by a three-way interaction between *Language Background*, *Homophone* and *Phonological Contrast* (AIC = 3052.5 vs. 3056.3).

Table 4.3 Linear mixed-effects model with *Language Background*, *Homophone*, *Phonological Contrast*, *Competition* and *Break Type* fitted to log-transformed RTs in Test Phase II.

Model formula: log(RT) ~ Homophone * Language_Background * Phonological_Contrast + Competition + (1 | Subject) + (1 | Word)

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(>|t|)
Intercept	7.83429	0.04903	98.21751	159.800	< 2e-16 ***
Homophone_Yes	0.23636	0.04612	41.07049	5.125	7.47e-06 ***
Language Background_Native English	0.07902	0.05927	112.61651	1.333	0.18516
Phonological Contrast_Cons Contrast	0.11742	0.04227	39.33815	2.778	0.00835 **
Phonological Contrast_Ton Contrast	0.00337	0.04922	38.24506	0.068	0.94577
Competition_No	-0.05652	0.01285	3249.26764	-4.397	1.13e-05 ***
Homophone_Yes:Language Background_Native English	-0.21343	0.04412	3251.87372	-4.838	1.37e-06 ***
Homophone_Yes:Phonological Contrast_Cons Contrast	-0.17314	0.06381	165.61815	-2.713	0.00736 **
Homophone_Yes:Phonological Contrast_Ton Contrast	-0.02068	0.05754	70.94097	-0.359	0.72034
Language Background_Native English:Phonological Contrast_Cons Contrast	-0.11185	0.03991	3255.25446	-2.803	0.00510 **
Language Background_Native English:Phonological Contrast_Ton Contrast	0.09771	0.04802	3255.25312	2.035	0.04197 *
Homophone_Yes:Language Background_Native English:Phonological Contrast_Cons Contrast	0.09103	0.07333	3249.77543	1.241	0.21454
Homophone_Yes:Language Background_Native English:Phonological Contrast_Ton Contrast	-0.11178	0.06388	3251.84118	-1.750	0.08025 .

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Based on the model in Table 4.3, the following subsubsections address the significant predictors and interactions affecting log-transformed RTs in Test Phase II. The *ggplot2* package (Wickham, 2016) was used again to visualize the effects of different predictors and interactions.

4.2.2.1 Three-way interaction between language background, homophone and phonological contrast

We first found a significant three-way interaction between *Language Background*, *Homophone* and *Phonological Contrast* predicting response latencies, which is illustrated in Figure 4.6 below. The two language groups showed different processing patterns for homophones. Specifically, native Mandarin speakers responded more slowly to homophones than to non-homophones across all three types of phonological contrasts, though the difference was only significant for novel words learned in a pair with consonant & tone contrasts ($p < 0.0001$) and with tone contrasts ($p = 0.0004$). While none of the pairwise comparisons were significant in the native English group, homophones and non-homophones had almost the same RTs for consonant contrasts and consonant & tone contrasts; for tone contrasts only, non-homophones had longer RTs than homophones.

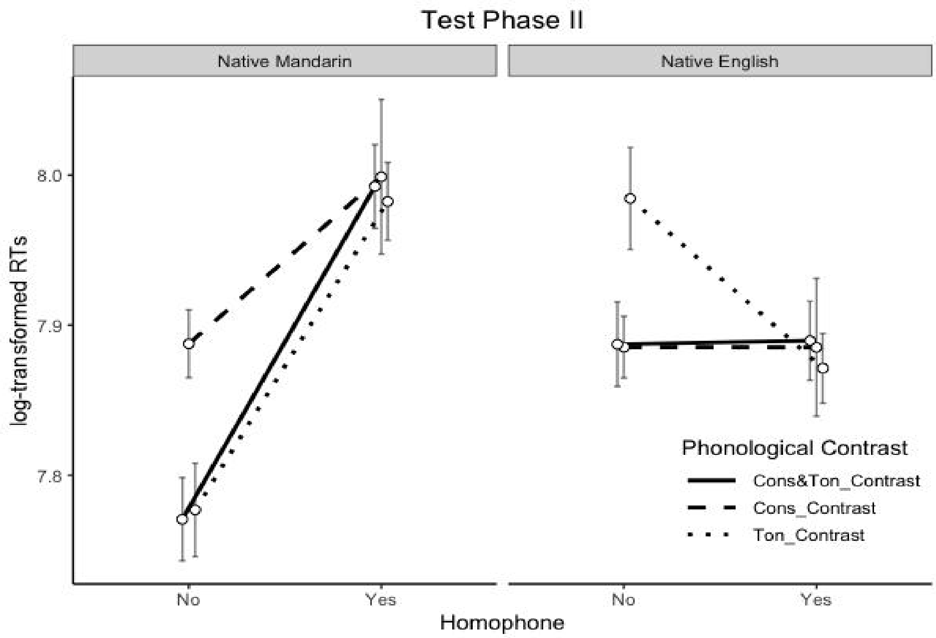


Figure 4.6 Three-way interaction between *Language Background*, *Homophone* (*x*-axis) and *Phonological Contrast* predicting log-transformed RTs (*y*-axis) in Test Phase II. Error bars represent standard error of mean.

In summary, homophones tended to be more ambiguous to process for Mandarin speakers than for English speakers as indicated by longer response latencies, but they were learned better than non-homophones by participants regardless of their language background as shown in previous section 4.1.2.2.

4.2.2.2 Competition effects

In addition, there was a significant effect of *Competition* on response latencies, which is presented in Figure 4.7 below. Pairwise comparisons showed that novel words tested in a visual setting without the competitor were processed significantly faster than those examined in the competitor context ($p < 0.0001$).

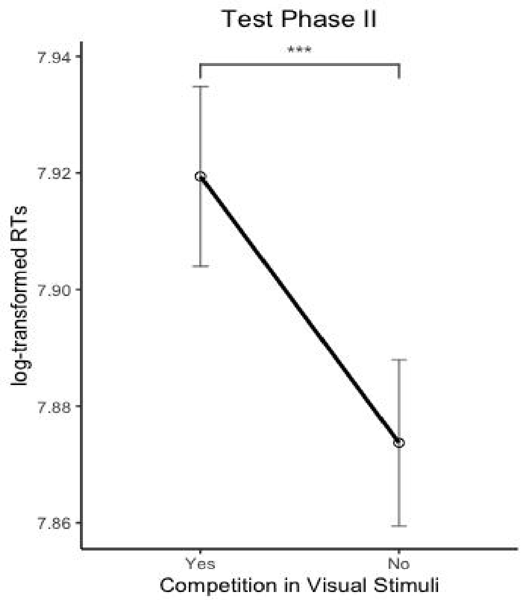


Figure 4.7 Effect of *Competition* (*x*-axis) on log-transformed RTs (*y*-axis) in Test Phase II. Error bars represent standard error of mean, and asterisks indicate a significant difference.

This result is consistent with what we found about the *Competition* effects on response accuracy in Test Phase II. To sum up, novel words tested in the no-competitor condition received significantly faster and more accurate responses than those presented in the competitor condition.

4.3 Eye movements in Test Phase II

Eye movements in Test Phase II only were analyzed in this study. For all participants, there were 4896 trials (68 participants x 72 trials) in total. These data were output as a sample report from the SR Research EyeLink Data Viewer, and then preprocessed using the *VWPre* package (Porretta, Kyröläinen, van Rij, & Järvikivi, 2016). Visual fixations that occurred in the target interest area were labelled as the target looks. Since 44 trials were empty, a total of 4852 trials were recorded, including 3916 and 936 trials using gaze data from the right and left eye, respectively. Among them, 3.83% of the data were marked as trackloss, thus 171 trials with less than 75% data present were removed. Accordingly, there were 4681 trials in total used for the following proportion of target looks analyses. As the eye tracker sampled at the frequency of either 500Hz or 1000Hz due to experimenter error, we subsetted the data and binned both data sets using a bin size of 20ms. Each bin was coded as *Target* or *Non-Target*. Then empirical logits (Barr, 2008) for the binned target proportion data within the time window from 0ms to 3000ms after target word onset were generated. In the end, after filtering out incorrect trials, there were a total of 429,072 data points of empirical logits of target looks, which were used as the response variable in the model.

For statistical analyses, we employed Generalized Additive Mixed-Models (GAMMs; Wood, 2017), which were constructed using the *mgcv* package (Wood, 2017) and visualized using the *itsadug* package (van Rij, Wieling, Baayen, & van Rijn, 2017). GAMMs are well-designed for analysing visual world time-course data and allow for modeling non-linear interactions with continuous predictors (time) as well as guarding against autocorrelation. In contrast to the backward-fitting method used in previous accuracy and RTs analyses, a forward stepwise fitting procedure was applied to attain the optimal model. Specifically, the model structure consisted of the fixed-effect factors and the random-effect factors. We started with fitting the fixed-effect factors, which contained a linear pattern for the predictor or interaction and a non-linear pattern over time. For the random-effect factors, we included a random intercept for *Event* (a unique combination of item by-subject), a random smooth for *Subject* by time and a random smooth for *Word* by time step by step. Using the *compareML* function in the *itsadug* package (van Rij et al., 2017), we adopted the Maximum Likelihood (ML) estimation method for model comparisons to determine the inclusion of the smoothing parameter or parametric component. The contribution of each new component added to the model was evaluated by the estimated *p*-value. If smaller than the conventional alpha threshold of 0.05, the new component was considered for inclusion. Additionally, in order to take autocorrelation into account, an AR1 error model for the residuals was incorporated by specifying the rho parameter and the starting point for each time series.

The response data discussed in previous sections 4.1 and 4.2 have pointed out three significant predictors or interactions affecting accuracy and RTs in Test Phase II: a significant interaction between *Language Background* and *Phonological Contrast*, a significant interaction between *Homophone* and *Language Background*, as well as a significant effect of *Competition*. Using the visual world eye-tracking paradigm, we are particularly interested to address the time course of effects that influence participants' learning outcome, and to see whether predictors and interactions that have effects on accuracy and RTs would also affect eye movements. In what follows, we investigated the time course of the above predictors by modeling their interaction with time, and we fitted models specifically to check for these predictors and interactions.

4.3.1 Interaction between language background and phonological contrast

In order to investigate how *Language Background* interacted with *Phonological Contrast* to predict target looks over time in Test Phase II, we concatenated *Language Background* and *Phonological Contrast* as a new factor (i.e. having six levels, novel words learned in pairs with three types of phonological contrasts by native Mandarin and English speakers, respectively). The final model is shown in Table 4.4 below.

Table 4.4 Generalized additive mixed model with *Language Background* and *Phonological Contrast* fitted to the target looks over time in Test Phase II, reporting parametric coefficients and smooth terms.

Model formula: IA_Target_ELogit ~ LangPhono + s (Time, by = LangPhono) + s (Event, bs = "re") + s (Time, Subject, bs = "fs", m = 1) + s (Time, Word, bs = "fs", m = 1)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.58611	0.18095	-3.239	0.0012 **
LangPhonoNativeEnglish.Cons&Ton_Contrast	0.11007	0.14340	0.768	0.4428
LangPhonoNativeMandarin.Cons_Contrast	0.01343	0.14510	0.093	0.9263
LangPhonoNativeEnglish.Cons_Contrast	0.18385	0.18565	0.990	0.3220
LangPhonoNativeMandarin.Ton_Contrast	0.11307	0.09985	1.132	0.2575
LangPhonoNativeEnglish.Ton_Contrast	0.18241	0.15237	1.197	0.2313

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s (Time):LangPhonoNativeMandarin.Cons&Ton_Contrast	7.22289	8.048	5.830	2.50e-07 ***
s (Time):LangPhonoNativeEnglish.Cons&Ton_Contrast	6.62040	7.568	8.023	2.38e-10 ***
s (Time):LangPhonoNativeMandarin.Cons_Contrast	7.24918	8.009	5.730	2.54e-07 ***
s (Time):LangPhonoNativeEnglish.Cons_Contrast	7.44367	8.157	7.610	4.36e-10 ***
s (Time):LangPhonoNativeMandarin.Ton_Contrast	7.00032	7.890	7.329	2.08e-09 ***
s (Time):LangPhonoNativeEnglish.Ton_Contrast	6.67552	7.642	8.509	2.90e-11 ***
s (Event)	0.01131	3279.000	0.000	1
s (Time, Subject)	316.81009	602.000	1.560	< 2e-16 ***
s (Time, Word)	206.47857	252.000	5.865	< 2e-16 ***

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

In Table 4.4 above, the parametric coefficients can be interpreted as in previous linear mixed-effects models. The intercept here is the value of the empirical logit of target looks for the reference level, namely, novel word pairs learned with consonant & tone contrasts by native Mandarin speakers. The following five lines indicate the differences between the reference level and other levels within the interaction variable, which were all insignificant. Of more interest is the smooth terms summary. As shown above, a non-linear curve which changes over time was found for all six levels, with the *edf* values greater than 1 and the *p*-values lower than the threshold of 0.05. However, we cannot conclude what these curves look like without visualization. Moreover, this non-parametric part does not reveal any differences between these curves. Therefore, we plotted the non-linear smooths and compared these difference curves in Figure 4.8 below.

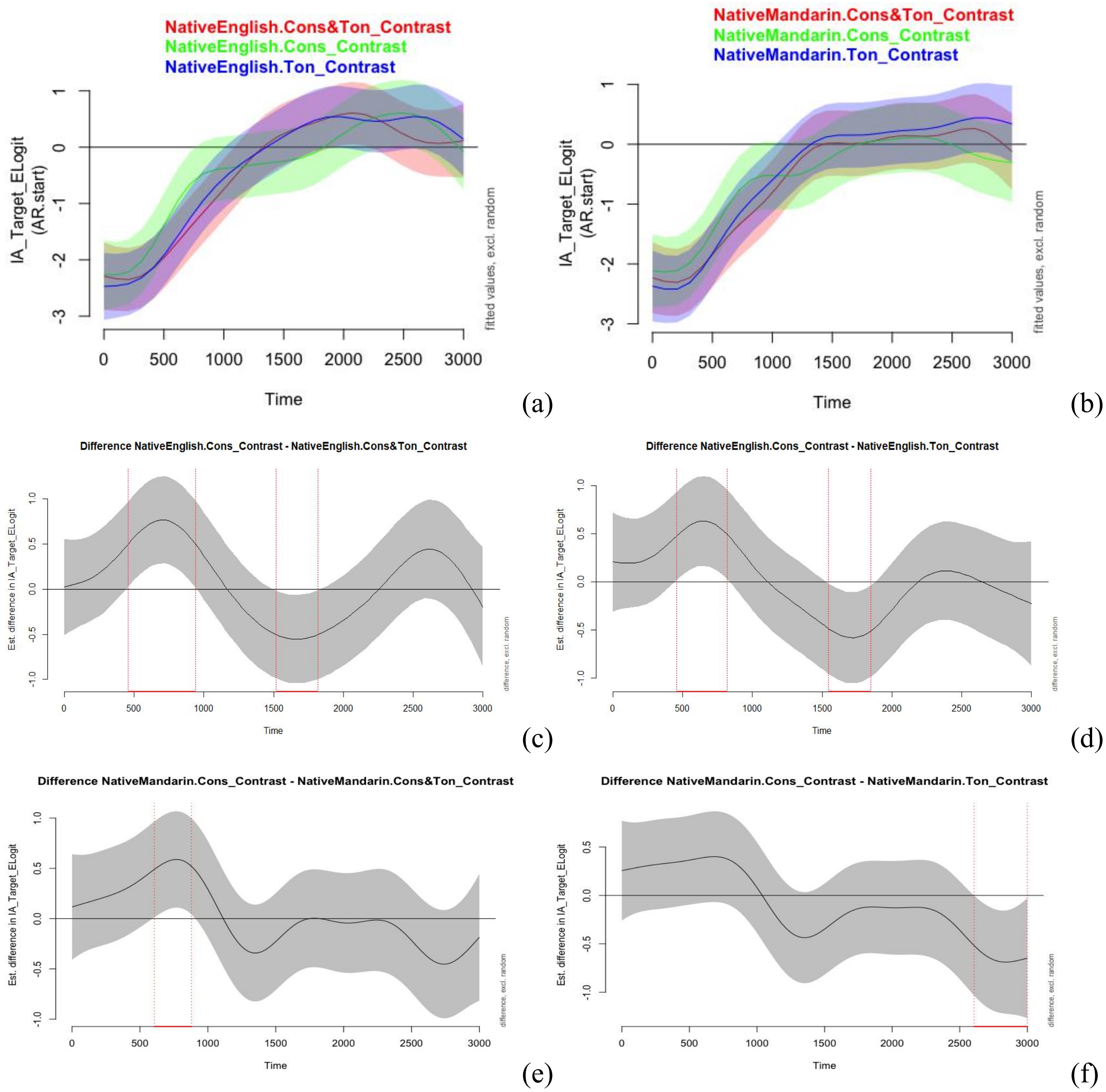


Figure 4.8 Panels (a) and (b) show non-linear smooths (fitted values) for the three types of phonological contrasts in native English speakers and native Mandarin speakers, respectively. Shaded bands represent the pointwise 95%-confidence interval. Panels (c) and (d) show the difference between the two non-linear smooths comparing novel word pairs with consonant contrasts to those with consonant & tone contrasts, as well as to those with tone contrasts in native English speakers, respectively. Panels (e) and (f) show the difference between the two non-linear smooths comparing novel word pairs with consonant contrasts to those with consonant & tone contrasts, as well as to those with tone contrasts in native Mandarin speakers, respectively. When the shaded pointwise 95%-confidence interval does not overlap with the *x*-axis (i.e. the value is significantly different from zero), this is indicated by a red line on the *x*-axis and vertical dotted lines.

As illustrated above, panels (a) and (b) show the non-linear smooths for the three types of phonological contrasts in native English speakers and native Mandarin speakers, respectively. An early effect of consonant contrasts on target looks was found in both language groups: participants tended to look more to the targets at the very beginning while they heard a novel word learned in a pair with consonant contrasts compared to the other two phonological contrasts. The remaining four panels illustrate the non-linear differences in the two groups. Specifically, in the English group, novel words learned in a pair with consonant contrasts had significantly more target looks than those acquired in the other two contrast types in the time window from approximate 450ms to 950ms, see panels (c) and (d). However, the Mandarin group presented only one significant difference for the early effect of consonant contrasts, in that novel words learned in a pair with consonant contrasts had significantly more target looks than those acquired in a pair with consonant & tone contrasts in the time window from 606ms to 879ms, see panel (e). Also, late in the time window, there were significantly fewer target looks for novel words learned in a pair with consonant contrasts compared to the other two phonological contrasts in native English speakers, see panels (c) and (d), and compared to words learned in a pair with tonal contrast in native Mandarin speakers, see panel (f). Additionally, the difference curves comparing tone contrasts and consonant & tone contrasts in both language groups did not show any significant differences, thus they were not included here.

To sum up, there was an early effect of consonant contrasts on target looks across both language groups, compared to the other two phonological contrast types. From 0ms to around 1100ms after target word onset, participants tended to look more to the target if the novel word was learned in a pair with consonant contrasts, which was indicated by the significant difference time windows.

4.3.2 Interaction between homophone and language background

In order to investigate how *Homophone* interacted with *Language Background* to predict target looks over time in Test Phase II, we concatenated *Homophone* and *Language Background* as a new factor (i.e. having four levels, homophones and non-homophones learned by native Mandarin and English speakers, respectively). The final model is given in Table 4.5 below.

As can be seen from the smooth terms summary, there was a non-linear curve which changes over time for every condition within the interaction variable. To further explore the non-linearity, we plotted the model's results in Figure 4.9 below. Panel (a) shows the non-linear smooths for the four levels of the newly created variable. Panels (b) and (c) present the differences between homophones and non-homophones among native Mandarin speakers and native English speakers, respectively. These differences were not significant, though participants in both language groups had more target looks to non-homophones than to homophones. Panels (d) and (e) show the differences between native English and Mandarin speakers for homophones and non-homophones, respectively. Panel (d) reveals a clear pattern that native English speakers generally had more target looks to homophones than native Mandarin speakers, with a significant difference in the time window between 1758ms to 2152ms.

Table 4.5 Generalized additive mixed model with *Homophone* and *Language Background* fitted to the target looks over time in Test Phase II, reporting parametric coefficients and smooth terms.

Model formula: IA_Target_ELogit ~ HomoLang + s (Time, by = HomoLang) + s (Event, bs = "re") + s (Time, Subject, bs = "fs", m = 1) + s (Time, Word, bs = "fs", m = 1)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.38392	0.18925	-2.029	0.0425 *
HomoLangYes.NativeMandarin	-0.41707	0.32965	-1.265	0.2058
HomoLangNo.NativeEnglish	0.06169	0.13322	0.463	0.6433
HomoLangYes.NativeEnglish	-0.22296	0.34964	-0.638	0.5237

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s (Time):HomoLangNo.NativeMandarin	7.48782	8.037	7.875	1.91e-10 ***
s (Time):HomoLangYes.NativeMandarin	6.56233	7.159	2.341	0.0273 *
s (Time):HomoLangNo.NativeEnglish	7.08197	7.706	7.332	2.49e-09 ***
s (Time):HomoLangYes.NativeEnglish	6.88628	7.485	4.506	6.76e-05 ***
s (Event)	0.08759	3284.000	0.000	1.0000
s (Time, Subject)	316.47132	602.000	1.557	< 2e-16 ***
s (Time, Word)	202.28706	252.000	5.619	< 2e-16 ***

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

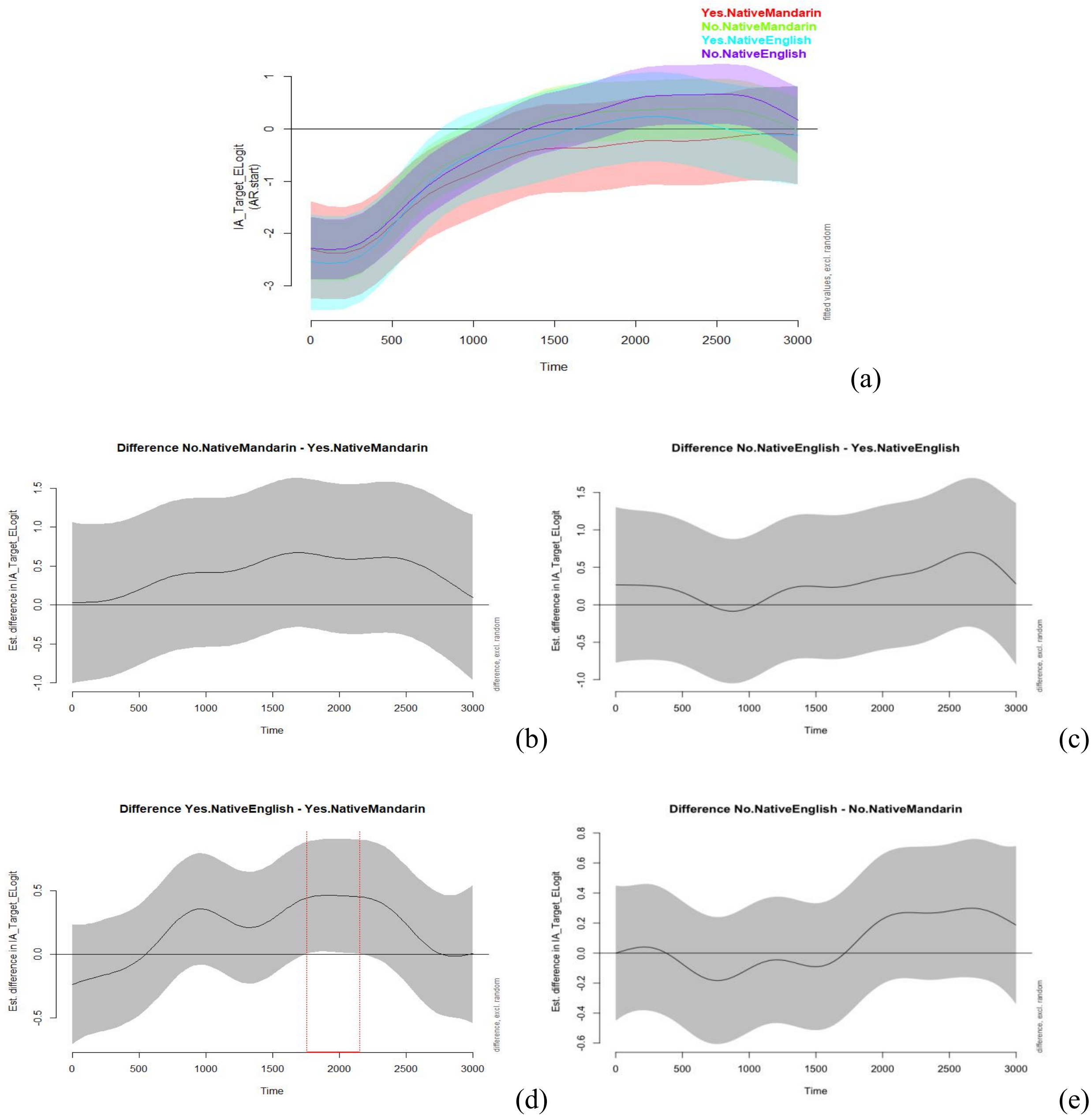


Figure 4.9 Panel (a) shows non-linear smooths (fitted values) for the four levels of the interaction between *Homophone* and *Language Background*. Shaded bands represent the pointwise 95%-confidence interval. Panels (b) and (c) show the difference between the two non-linear smooths comparing homophones to non-homophones in native Mandarin speakers and native English speakers, respectively. Panels (d) and (e) show the difference between the two non-linear smooths comparing the two language groups with regard to homophones and non-homophones, respectively. When the shaded pointwise 95%-confidence interval does not overlap with the *x*-axis (i.e. the value is significantly different from zero), this is indicated by a red line on the *x*-axis and vertical dotted lines.

4.3.3 Three-way interaction between language background, homophone, and phonological contrast

Furthermore, we explored how the three-way interaction between *Language Background*, *Homophone* and *Phonological Contrast* affected the target looks over time in Test Phase II. We concatenated *Language Background*, *Homophone* and *Phonological Contrast* as a new variable (i.e. having 12 levels, homophones and non-homophones learned in novel word pairs with three types of phonological contrasts by native Mandarin and English speakers, respectively). The resulting model is shown in Table 4.6 below.

Table 4.6 Generalized additive mixed model with *Language Background*, *Homophone* and *Phonological Contrast* fitted to the target looks over time in Test Phase II, reporting parametric coefficients and smooth terms.

Model formula: IA_Target_ELogit ~ HomoLangPhono + s (Time, by = HomoLangPhono) + s (Event, bs = "re") + s (Time, Subject, bs = "fs", m = 1) + s (Time, Word, bs = "fs", m = 1)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.35937	0.33469	-1.074	0.2829
HomoLangPhonoYes.NativeMandarin.Cons&Ton_Contrast	-0.85009	0.39763	-2.138	0.0325 *
HomoLangPhonoNo.NativeEnglish.Cons&Ton_Contrast	0.04613	0.16875	0.273	0.7846
HomoLangPhonoYes.NativeEnglish.Cons&Ton_Contrast	-0.68406	0.42198	-1.621	0.1050
HomoLangPhonoNo.NativeMandarin.Cons_Contrast	-0.16680	0.41495	-0.402	0.6877
HomoLangPhonoYes.NativeMandarin.Cons_Contrast	-0.78186	0.41535	-1.882	0.0598 .
HomoLangPhonoNo.NativeEnglish.Cons_Contrast	0.01058	0.43055	0.025	0.9804
HomoLangPhonoYes.NativeEnglish.Cons_Contrast	-0.61072	0.44320	-1.378	0.1682
HomoLangPhonoNo.NativeMandarin.Ton_Contrast	0.19547	0.46750	0.418	0.6759
HomoLangPhonoYes.NativeMandarin.Ton_Contrast	-0.79389	0.39447	-2.013	0.0442 *
HomoLangPhonoNo.NativeEnglish.Ton_Contrast	-0.00587	0.48087	-0.012	0.9903
HomoLangPhonoYes.NativeEnglish.Ton_Contrast	-0.57532	0.42055	-1.368	0.1713

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s (Time):HomoLangPhonoNo.NativeMandarin.Cons&Ton_Contrast	13.27979	16.129	3.452	3.18e-06 ***
s (Time):HomoLangPhonoYes.NativeMandarin.Cons&Ton_Contrast	10.25716	13.406	1.365	0.1867
s (Time):HomoLangPhonoNo.NativeEnglish.Cons&Ton_Contrast	7.09491	9.180	2.103	0.0240 *
s (Time):HomoLangPhonoYes.NativeEnglish.Cons&Ton_Contrast	4.20927	5.543	2.156	0.0487 *
s (Time):HomoLangPhonoNo.NativeMandarin.Cons_Contrast	14.00360	16.593	3.690	1.27e-06 ***
s (Time):HomoLangPhonoYes.NativeMandarin.Cons_Contrast	1.00809	1.016	1.246	0.2630
s (Time):HomoLangPhonoNo.NativeEnglish.Cons_Contrast	15.34199	17.553	5.521	1.03e-12 ***
s (Time):HomoLangPhonoYes.NativeEnglish.Cons_Contrast	11.02740	14.204	3.243	3.17e-05 ***
s (Time):HomoLangPhonoNo.NativeMandarin.Ton_Contrast	9.29550	12.181	1.532	0.1072
s (Time):HomoLangPhonoYes.NativeMandarin.Ton_Contrast	1.01173	1.020	5.085	0.0232 *
s (Time):HomoLangPhonoNo.NativeEnglish.Ton_Contrast	2.97654	3.652	0.962	0.4351
s (Time):HomoLangPhonoYes.NativeEnglish.Ton_Contrast	13.07468	16.125	3.083	2.99e-05 ***
s (Event)	0.01053	3273.000	0.000	1.0000
s (Time, Subject)	317.52614	603.000	1.562	< 2e-16 ***
s (Time, Word)	197.92607	252.000	5.784	< 2e-16 ***

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

According to the smooth terms summary above, a non-linear curve which changes over time was found in homophones learned with tone contrasts and non-homophones involved with consonant contrasts among native Mandarin speakers, and in homophones learned with any type of phonological contrasts and non-homophones involved with consonant contrasts among native English speakers. To examine the non-linear pattern of these curves, we plotted all difference graphs that compare homophones and non-homophones learned with the three types of phonological contrasts by native Mandarin speakers (the first three panels) and native English speakers (the last three panels) in Figure 4.10 below. The two language groups showed a consistent pattern: non-homophones generally had more target looks than homophones across all phonological contrast types for both language groups. More specifically, in the Mandarin group, there was a significant difference in the time windows from 1152ms to 2515ms for novel word pairs learned with consonant & tone contrasts, from 1303ms to 2394ms for novel word pairs acquired with consonant contrasts, and from 576ms to 1667ms for novel words pairs studied with tone contrasts, as shown in panels (a), (b) and (c), respectively. In the English group, a significant difference was found in the time windows between 1848ms to 2333ms for novel word pairs with consonant & tone contrasts, and between 1152ms to 1909ms and 2727ms to 2818ms for novel words pairs with consonant contrasts, as panels (d) and (e) illustrate, respectively. However, there was no significant difference for novel words studied in a pair with tone contrasts among native English speakers, see panel (f) below.

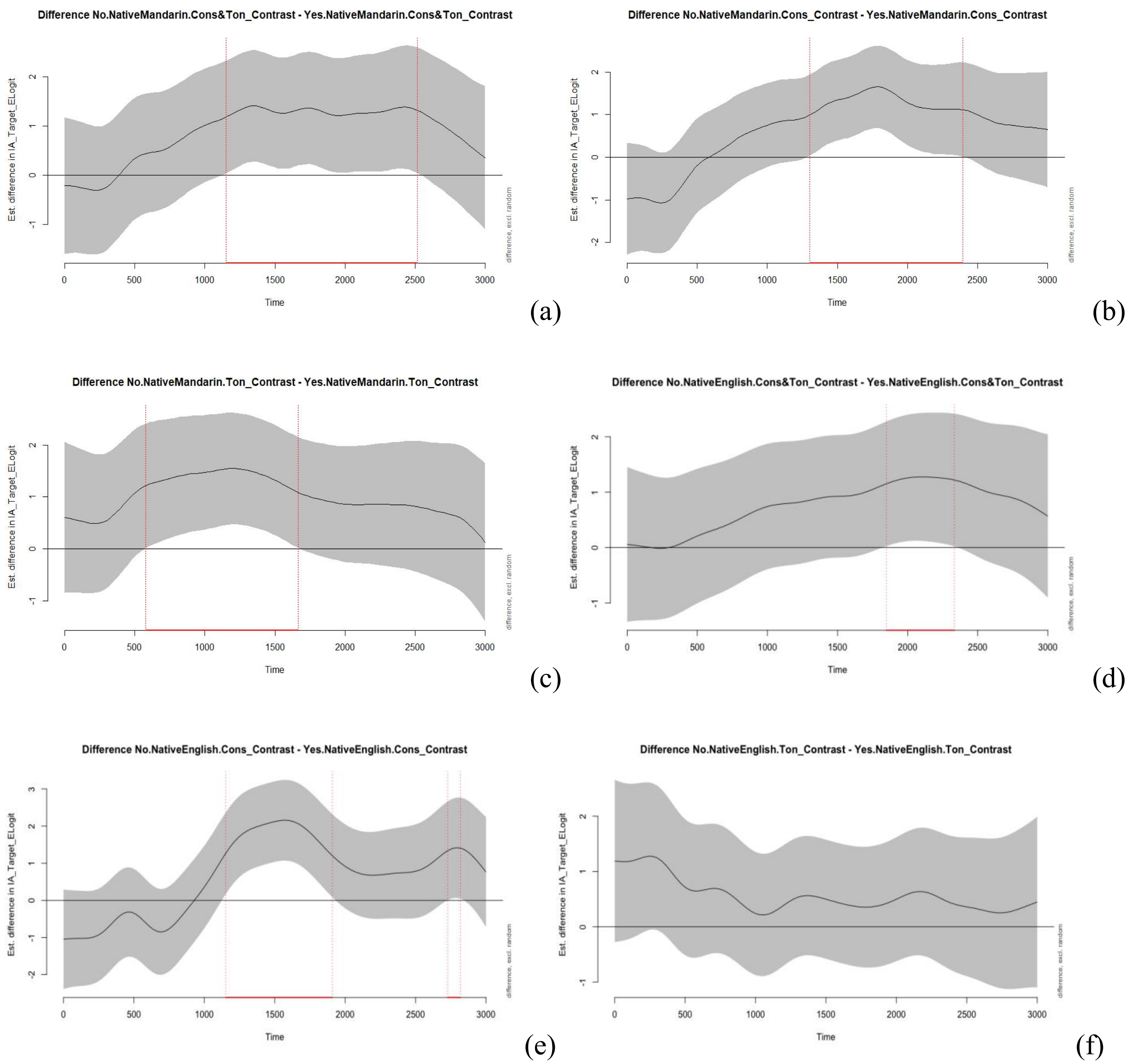


Figure 4.10 Panels (a) to (c) illustrate the difference between the two non-linear smooths comparing homophones to non-homophones learned by native Mandarin speakers in a pair with consonant & tone contrasts, consonant contrasts, and tone contrasts, respectively. Panels (d) to (f) present the difference between the two non-linear smooths comparing homophones to non-homophones learned by native English speakers in a pair with consonant & tone contrasts, consonant contrasts, and tone contrasts, respectively. When the shaded pointwise 95%-confidence interval does not overlap with the *x*-axis (i.e. the value is significantly different from zero), this is indicated by a red line on the *x*-axis and vertical dotted lines.

4.3.4 Competition effects

In order to capture potentially different trends over time for the two visual conditions, we converted *Competition* into an ordered factor *OFCompetition* before building the model. An ordered factor model allows us to assess whether the non-linear difference between the two levels of the factor is significant or not (Wieling, 2018). The resulting model is presented in Table 4.7 below.

Table 4.7 Generalized additive mixed model with *Competition* fitted to the target looks over time in Test Phase II, reporting parametric coefficients and smooth terms.

Model formula: IA_Target_ELogit ~ OFCompetition + s (Time) + s (Time, by = OFCompetition) + s (Event, bs = "re") + s (Time, Subject, bs = "fs", m = 1) + s (Time, Word, bs = "fs", m = 1)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.46678	0.15432	-3.025	0.00249 **
OFCompetitionNo	0.13448	0.05015	2.682	0.00732 **

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s (Time)	7.87866	8.186	11.535	4.84e-15 ***
s (Time):OFCompetitionNo	6.68729	7.932	6.733	1.36e-08 ***
s (Event)	0.01042	3284.000	0.000	1
s (Time, Subject)	318.66065	603.000	1.562	< 2e-16 ***
s (Time, Word)	205.40254	252.000	5.776	< 2e-16 ***

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

According to the smooth terms summary above, there was a significantly different trend over time between the competitor and no-competitor conditions in Test Phase II ($F = 6.733$, $p < 0.001$). Figure 4.11 below illustrates the partial effects of *OFCompetition* and the smooths difference between the competitor and no-competitor conditions, which reveals that the no-competitor condition generally had more target looks than the competitor condition and that there was a significant difference in the time window from 455ms to 1545ms, despite a brief early effect of significantly more target looks in the competitor context before 182ms.

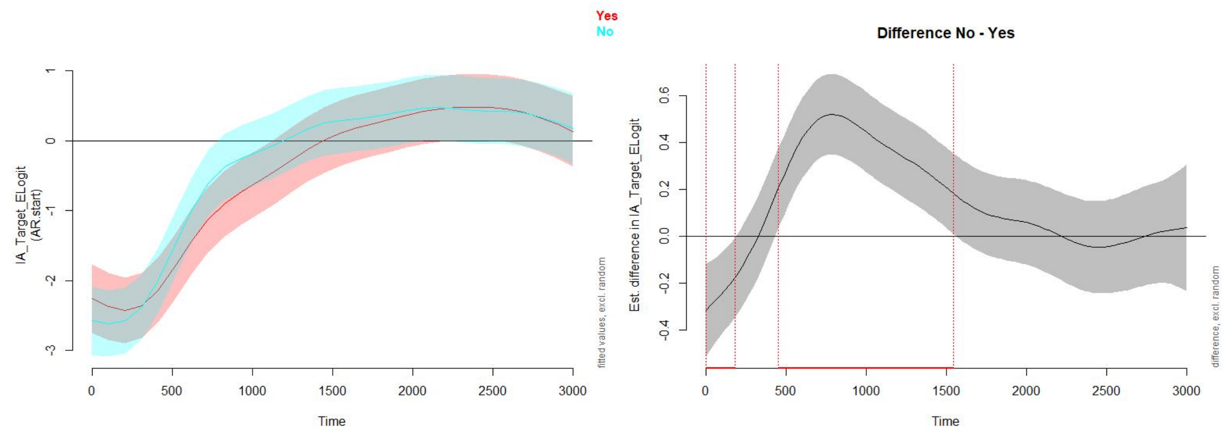


Figure 4.11 Left: Non-linear smooths (fitted values) for the competitor and no-competitor conditions. Shaded bands represent the pointwise 95%-confidence interval. Right: Difference between the two non-linear smooths comparing the competitor condition to the no-competitor condition. When the shaded pointwise 95%-confidence interval does not overlap with the *x*-axis (i.e. the value is significantly different from zero), this is indicated by a red line on the *x*-axis and vertical dotted lines.

In addition, we explored how *Competition* interacted with *Language Background* to predict the target looks over time in Test Phase II. To test this, we concatenated *Language Background* and *Competition* as a new variable (i.e. having four levels, the competitor and no-competitor conditions for native Mandarin and English speakers, respectively). The final model is presented in Table 4.8 below.

Table 4.8 Generalized additive mixed model with *Competition* and *Language Background* fitted to the target looks over time in Test Phase II, reporting parametric coefficients and smooth terms.

Model formula: IA_Target_ELogit ~ LangComp + s (Time, by = LangComp) + s (Event, bs = "re") + s (Time, Subject, bs = "fs", m = 1) + s (Time, Word, bs = "fs", m = 1)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.66760	0.16678	-4.003	6.26e-05 ***
LangCompNativeEnglish.Yes	0.23262	0.13519	1.721	0.08532 .
LangCompNativeMandarin.No	0.24605	0.06988	3.521	0.00043 ***
LangCompNativeEnglish.No	0.25542	0.13336	1.915	0.05547 .

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s (Time):LangCompNativeMandarin.Yes	6.2565	7.128	6.144	4.25e-07 ***
s (Time):LangCompNativeEnglish.Yes	6.6364	7.512	9.019	1.42e-11 ***
s (Time):LangCompNativeMandarin.No	8.0042	8.538	11.260	9.52e-15 ***
s (Time):LangCompNativeEnglish.No	7.7017	8.334	10.860	5.33e-15 ***
s (Event)	0.0147	3280.000	0.000	1
s (Time, Subject)	316.4923	601.000	1.564	< 2e-16 ***
s (Time, Word)	207.0703	251.000	5.851	< 2e-16 ***

Note: Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

As shown in the smooth terms summary, there was a non-linear curve which changes over time for every level of this interaction variable. To further examine the non-linear pattern of these curves, we plotted the model in Figure 4.12 below. Panel (a) shows the non-linear smooths for the two visual settings in native Mandarin and English speakers. Panels (b) and (c) show the difference between the two non-linear smooths comparing the competitor and no-competitor conditions in the English group and in the Mandarin group, respectively. It was found that in both language groups, there were more target looks in the no-competitor condition than in the competitor condition. Specifically, there was a significant difference in the time windows from 485ms to 1091ms in native English speakers and from 515ms to 1515ms in native Mandarin speakers. Moreover, panel (d) presents the difference between the two non-linear smooths comparing the English group and the Mandarin group in the competitor setting. We found that native English speakers tended to have more target looks than native Mandarin speakers when there was a competitor in the visual array, and there was a significant difference between the two language groups in the time window from 1970ms to 2667ms. However, no significant difference was found while comparing the non-linear smooths for the two language groups in the no-competitor condition, so we did not include the plot below.

In conclusion, participants in both language groups tended to look significantly more at the target when there was no competitor in the visual setting, compared to the case when a competitor occurred. Additionally, when there was a competitor in the visual context, native English speakers had significantly more target looks than native Mandarin speakers.

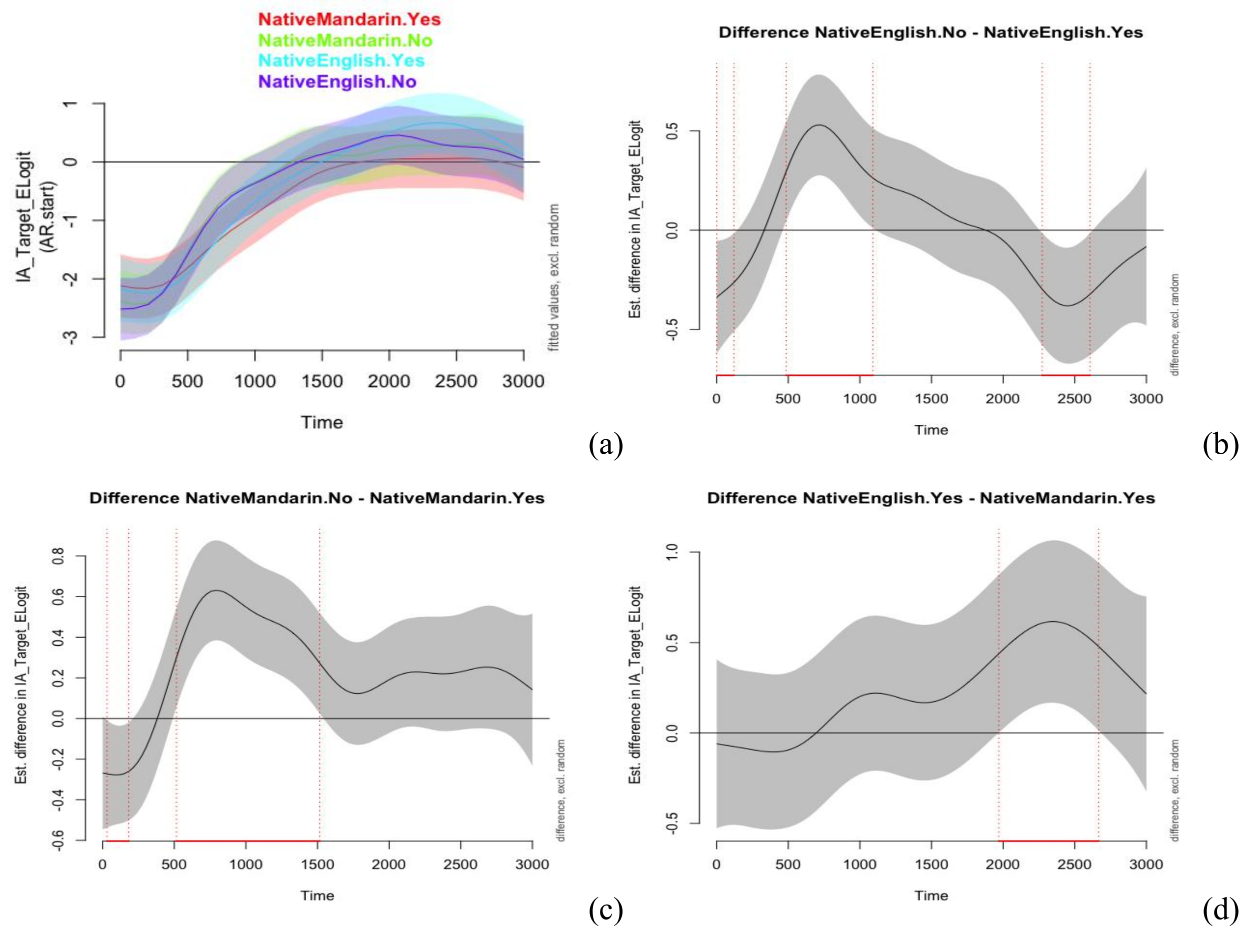


Figure 4.12 Panel (a) shows non-linear smooths (fitted values) for the four levels of the interaction between *Competition* and *Language Background*. Shaded bands represent the pointwise 95%-confidence interval. Panel (b) shows the difference between the two non-linear smooths comparing the competitor condition to the no-competitor condition among native English speakers. Panel (c) shows the difference between the two non-linear smooths comparing the competitor condition to the no-competitor condition among native Mandarin speakers. Panel (d) shows the difference between the two non-linear smooths comparing native English speakers to native Mandarin speakers when there was a competitor in the visual setting. When the shaded pointwise 95%-confidence interval does not overlap with the *x*-axis (i.e. the value is significantly different from zero), this is indicated by a red line on the *x*-axis and vertical dotted lines.

Chapter 5: Discussion

In this study, we used the visual world eye-tracking paradigm to investigate how adults from native and non-native backgrounds learn novel words which contain homophones and non-homophones, and vary in three types of Mandarin phonological contrasts: consonant contrasts, tone contrasts, consonant & tone contrasts. We also investigated whether short periods of unoccupied rest following learning affect the outcome, in comparison to completing a distractor task for an equivalent duration.

First of all, we asked whether participants' language background would modulate their learning outcome, and whether that would predict their response accuracy and RTs in Test Phases I and II, as well as their eye movements in Test Phase II. To that end, we tested native speakers of Mandarin and English, and hypothesized that participants' language background was a significant predictor for all three measurements above. We also predicted that Mandarin speakers would have significantly faster and more accurate responses, and more target looks than English speakers.

With regard to the novel words, we asked how participants' learning would be affected by whether the word was a homophone or a non-homophone, and whether that would depend on their language background. Our hypothesis was that homophony would interact with participants' language background to predict their RTs, response accuracy and eye movements, with the two language groups showing different tendencies.

In addition, we asked how different types of phonological contrasts entailed in the novel word pairs would affect participants' outcome, and whether that would depend on their language background. We hypothesized that phonological contrasts would interact with participants' language backgrounds to predict the three aforementioned measurements, and that for native

English speakers, words learned in a pair with tone contrasts would probably receive the slowest and least accurate responses.

Lastly, we explored how the two types of break would influence participants' learning outcome in Test Phase II, and whether that would depend on their language background. Our prediction was that regardless of their language background, participants who had an eyes-closed rest would have faster and more accurate responses, and more target looks than those who played the computer game.

The effects of each manipulated variable and its role in interactions are discussed below.

5.1 Language background effects

We first analyzed the effect of participants' language background on their response accuracy. In Test Phase I, native Mandarin speakers had significantly higher accuracy than native English speakers, as we hypothesized. This can be attributed to the native phonological knowledge that Mandarin speakers could use when learning these novel words that were constructed according to Mandarin phonology, relative to their native English-speaking counterparts. In Test Phase II, *Language Background* interacted with *Phonological Contrast* to predict response accuracy significantly. As expected, the two language groups showed different patterns. Native English speakers were significantly more accurate on novel words learned with consonant contrasts than those acquired with tone or consonant & tone contrasts, whereas there were no significant differences in native Mandarin speakers or between the two language groups.

These findings have two implications. First of all, the two language groups probably used different strategies in novel word learning. Specifically, the English group performed well on novel words involved in consonant contrasts, but their learning outcome was impeded when there were tones in the phonological conditions. However, the Mandarin group showed no significant differences between the three types of phonological contrasts. Second, it can be speculated that native Mandarin speakers mainly benefitted from their familiarity with the Mandarin phonology. Since the Mandarin segments used in the novel word stimuli (i.e. the six consonants /p^h, t^h, k^h, m, n, l/ and three vowels /a, u, i/) are similar to their English counterparts in terms of pronunciation, it is plausible that Mandarin tones are comparable to phonemes like consonants and vowels for Mandarin listeners but not for English listeners.

Additional significant interactions that involve *Language Background* affecting participants' responses and eye movements are reported below.

5.2 Homophone effects

We investigated how participants' learning was affected by whether the novel word was a homophone or a non-homophone, and whether that would depend on their language background. On one hand, we found a significant interaction between *Homophone* and *Language Background* predicting response accuracy in Test Phase II: Homophones had higher accuracy than non-homophones across both language groups, though the difference did not reach significance for the Mandarin speakers. This suggests that regardless of their language background, participants learned the lexical mappings between form and meaning, despite them sometimes being one-to-one and other times one-to-many. This is totally possible in natural language learning settings, especially for English and Mandarin speakers who have around 20% and 80% of homophones in their native languages, respectively (Ke et al., 2002). Interestingly, a significant difference between homophones and non-homophones was found in the English group, who demonstrated a stronger homophony advantage than the Mandarin group. Further, with non-homophones, native Mandarin speakers were significantly more accurate than native English speakers. Therefore, it appears that the homophony learning advantage was particularly enhanced in native English speakers, whereas native Mandarin speakers maintained their edge in acquiring non-homophones due to the benefit of native phonological knowledge.

On the other hand, we found a significant three-way interaction between *Language Background*, *Homophone* and *Phonological Contrast* affecting response latencies in Test Phase II, which showed different processing patterns for homophones in the two language groups. Specifically, native Mandarin speakers responded more slowly to homophones than to non-homophones across all three types of phonological contrasts, though the difference was only significant for novel words acquired in a pair with tone contrasts or consonant & tone contrasts. In the English group by contrast, there was little difference in RTs between homophones and non-homophones learned in a pair with consonant contrasts or consonant & tone contrasts, but homophones received faster responses when they were learned in a pair with tone contrasts, although the difference was not significant. These findings are in contrast with the facilitatory homophone effects found in previous Chinese studies (e.g., Ziegler et al., 2000; Chen et al., 2009), and the processing disadvantage reported in prior work on English homophones (e.g., Rubenstein et al., 1971; Pexman et al., 2001). A possible reason is that previous researchers tended to use real Chinese or English homophones in their experiments, however, the homophones used in this study were novel words, which might lead to a different pattern. In addition, given that English speakers had the lowest accuracy on tone contrasts, it is possible that the English group acquired tone contrasts so unsuccessfully that they tended to neglect the tonal differences, especially in homophones. This is further evidenced by English speakers' eye movement data, which suggested that tone contrasts were the only type of phonological contrasts with which no significant difference in target looks was found between homophones and non-homophones.

The visual world experiment provides a new insight into how adults acquire homophones in a novel word learning task. Our findings for native Mandarin speakers complement existing studies of homophone effects in lexical processing, showing that homophones with only one homophonic mate are processed more slowly than non-homophones (e.g., Hino et al., 2013). It is important to be able to show that homophony interacts with the phonological context in which homophones are learned, as well as with learners' language background, as described above. However, it is not clear whether this homophony acquisition pattern could be extended into a learning environment where a higher percentage of homophones are used in the novel words, or where most of the homophones have more than one homophonic mate. Directions for future work are discussed in the final chapter.

5.3 Phonological contrast effects

Regarding different types of phonological contrasts, we found a significant interaction between *Language Background* and *Phonological Contrast* affecting response accuracy in Test Phase II. As discussed above in section 5.1, the two language groups showed different patterns in processing novel words varied in types of phonological contrasts. The English-speaking adults learned novel words with consonant contrasts significantly better than those with the other two types of contrasts, but they experienced challenges when tones were involved in the phonological learning environment. In contrast, the Mandarin-speaking adults showed no significant differences in accuracy between the three types of phonological contrasts. These findings corroborate our hypothesis and provide further evidence for what has been found in the prior literature. That is, in line with Li et al. (2019), tonal information might play a less important role than segmental information in non-native Mandarin lexical processing.

Moreover, as stated above in section 5.2, the significant three-way interaction between *Language Background*, *Homophone* and *Phonological Contrast* affecting response latencies in Test Phase II indicated different tendencies in the two language groups. Native Mandarin speakers processed homophones more slowly than non-homophones regardless of the phonological contrast type, although the difference only reached significance for novel words involving tonal differences. Since each tone is associated with far more lexical entries than each segment is in Mandarin, and minimal word pairs that differ only in tone are abundant, it is very likely that tonal information provides fewer contrasting cues than segmental information in lexical access. Further, eye movements suggested that there was an early effect of consonant contrasts on target looks across both language groups, which was indicated by significantly more early target looks than in the other two phonological conditions in Test Phase II.

Taken together, in the frame of the novel word learning task, our findings build on existing evidence that tonal information might play a weaker role in spoken word processing, among native English speakers in this study. In comparison to segmental information like consonants, tonal information tends to raise challenges for native English speakers during novel word learning. However, it is beyond the scope of this study to investigate how well specific segments and tones are acquired by participants. Future research could build on the current work by addressing these research questions.

5.4 Break type effects

In terms of different types of break between the two phases of the experiment, we found a significant interaction between *Break Type* and *Language Background* predicting response accuracy in Test Phase II, which indicated opposing tendencies in the two language groups. Contrary to expectations, only native English-speaking adults benefited from a brief period of unoccupied rest, relative to completing a distractor task for an equivalent duration. Interestingly, Mandarin adults had higher accuracy after playing the game than following an eyes-closed rest.

This finding provides some further evidence for the rest-induced memory enhancement, which is found in novel word learning among adult non-native learners in the present study (here, native English speakers). Looking at prior research, Kurdziel and Spencer (2016) demonstrated the benefit of sleep in novel word learning among English-speaking adults, while they were acquiring very low-frequency native words. Meanwhile, some studies revealed memory consolidation during wakeful resting, such as declarative memory (Brokaw et al., 2016) and procedural memory (Humiston & Wamsley, 2018). Nevertheless, none of them studied the function of moments of unoccupied rest in novel word learning in both native and non-native adult learners, which was the gap that this study attempted to fill. In contrast, the Mandarin adults did not benefit from the rest following learning, but from completing the distractor task. A possible explanation is that the Mandarin speakers were asked to learn native-like nonwords in this study, so rest did not come into play. Therefore, our results only lent support to waking rest's benefit among non-native learners in novel word learning. Further data collection would be needed to determine whether another group of adult non-native learners, who are from non-English speaking backgrounds, would benefit from brief wakeful resting.

5.5 Competition effects

Using the visual world paradigm, we designed two display settings: the competitor scene consisting of the target, its competitor and a distractor, and the no-competitor scene composed of the target and two distractors. We asked participants to complete a visual search task to locate the target object on the display as they listened to the auditory input. Participants' response and eye movement data suggested that the no-competitor condition had significantly faster and more accurate responses, as well as more target looks than the competitor condition. Since competition was used as an indicator for participants' learning success, this competition effect has provided support for adults' ability to learn novel words, no matter whether they are from a native or a non-native background.

In addition, a possible reason for the competition effect is that the presence of the competitor interfered with the perception of the target object. Essentially, the phonological similarity between the names of the target and the competitor gives rise to the competition effect. Upon hearing the target word during the test, participants were influenced by the visual representation of the competitor once it occurred in the visual array. Eye movements generated during this process reflect an ongoing matching between the phonological and visual representations associated with the displayed objects in the visual workspace. Although we used different tasks from previous studies (e.g., Allopenna et al., 1998), the phonological competition between the target and its related competitors was clearly evidenced. Moreover, the interaction between *Competition* and *Language Background* affecting target looks indicated that this interference was stronger in Mandarin adults than in English adults. This can be justified by the fact that native Mandarin speakers were more familiar with the phonological learning environment in this experiment than their English counterparts. Thus, it is more likely for Mandarin adults to be interfered during lexical processing as they required additional effort to make sure that they had made the correct responses.

Chapter 6: Conclusion and future directions

The present study used visual world eye-tracking to investigate how adults from native and non-native backgrounds learn novel words that involve homophones and non-homophones, and vary in different types of phonological contrasts present in Mandarin. The role of brief periods of unoccupied rest in consolidation of novel word learning was also investigated. In terms of the effect of participants' language background, results suggested that in the test phase integrated with the training, native speakers had significantly higher accuracy than non-native speakers. However, in the test phase following the learning, different tendencies were shown in native and non-native speakers for different types of phonological contrasts. Specifically, in this study, English speakers were significantly more accurate on novel words involved with consonant contrasts than those involving tones, whereas there was no significant difference in accuracy between different phonological conditions in Mandarin speakers. The pattern found among English speakers can be probably generalized to non-native learners of Mandarin who have no prior experience with or knowledge of tone languages.

Regarding homophony acquisition, homophones received more accurate responses than non-homophones in all participants, but interestingly the difference was only significant among non-native speakers. Moreover, native speakers responded to homophones more slowly than to non-homophones across all types of phonological contrasts, though the difference only reached significance for phonological conditions involving tones. Given the abundance of homophones in Mandarin, it is plausible that homophony causes processing ambiguity in native speakers, especially when tonal information is involved. With respect to the benefits of wakeful resting, only observations of non-native learners lent support to the consolidation of novel word learning after brief periods of unoccupied rest. It is possible that wakeful resting facilitates acquiring of novel words in a non-native language, but does not come into play in native-like learning. Furthermore, our findings highlighted the visual world competition effects. When there was no competitor in the visual array, participants had significantly faster and more accurate responses, as well as more target looks than the context with the presence of a competitor. Taken together, these findings have contributed to our understanding about novel word learning in adult native and non-native learners.

This research has also raised many questions in need of further investigation. To start with, additional work is needed to explore the homophony learning pattern in a scenario where a greater proportion of homophones among novel words is used, and more novel words with multiple homophonic mates are involved. In the current study, homophones constituted 25% of the linguistic stimuli, which is close to the percentage of homophones in English. It will be of interest to examine what happens if a much larger number of homophones are used, such as the approximate proportion of 80% in Mandarin. Moreover, it is recommended that further research take vowel contrasts into consideration while creating the novel words and compare adult learners' performance on specific segments and tones. Another important issue to resolve is to assess the learning patterns in non-native speakers with other language history, such as Cantonese-speaking adults who have a more complex tonal system than Mandarin tones. Future work should also be undertaken to examine the rest-induced memory consolidation in another group of non-native learners, other than the English background, to establish a fuller picture.

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*, 419−439.

Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457−474. doi: 10.1016/j.jml.2007.09.002

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.11, retrieved 6 April 2020 from http://www.praat.org/

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch Ability As an Aptitude for Tone Learning. *Language Learning, 66*(4), 774–808. doi: 10.1111/lang.12159

Brokaw, K., Tishler, W., Manceor, S., Hamilton, K., Gaulden, A., Parr, E., & Wamsley, E. J. (2016). Resting state EEG correlates of memory consolidation. *Neurobiology of Learning and Memory, 130*, 17–25. http://dx.doi.org/10.1016/j.nlm.2016.01.008

Burnham, D., Singh, L., Mattock, K., Woo, P. J., & Kalashnikova, M. (2018). Constraints on Tone Sensitivity in Novel Word Learning by Monolingual and Bilingual Infants: Tone Properties Are More Influential than Tone Familiarity. *Frontiers in Psychology, 8*: 2190. doi: 10.3389/fpsyg.2017.02190

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition, 128*, 407–416.

Caldwell-Harris, C. L., Lancaster, A., Ladd, D. R., Dediu, D., & Christiansen, M. H. (2015). Factors influencing sensitivity to lexical tone in an artificial language: implications for second language learning. *Studies in Second Language Acquisition, 37*, 335–357. doi: 10.1017/S0272263224000849

Chan, C. C. Y., Tardif, T., Chen, J., Pulverman, R. B., Zhu, L., & Meng, X. (2011). English- and Chinese-Learning Infants Map Novel Labels to Objects and Actions Differently. *Developmental Psychology, 47*(5), 1459–1471.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language, 108*, 1–9. doi: 10.1016/j.bandl.2008.02.001

Chang, C. B., & Bowels, A. R. (2015). Context effects on second-language learning of tonal contrasts. *Journal of the Acoustical Society of America, 138*(6), 3703–3716. doi: 10.1121/1.4937612

Chao, Y.-R. (1930). A system of tone-letters. *Le Maitre Phonetique, 45*, 24–27.

Chen, C.-Y. (1984). Neutral Tone in Mandarin: Phonotactic description and the issue of the norm. *Journal of Chinese Linguistics, 12*(2), 299–333.

Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes, 24*(7–8), 967–982. doi: 10.1080/01690960902804515

Chen, Y., & Liu, H.-M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities, 35*, 10–20.

Chinese Academy of Social Sciences Institute of Linguistics. (1985). *Xianda Hanyu Cidian (A Dictionary of Current Chinese)*. Beijing: The Commercial Press.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America, 131*(6), 4756–4769. doi: 10.1121/1.4714355

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *Journal of the Acoustical Society of America, 134*(2), EL 133–139. doi: 10.1121/1.4812435

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review, 12*, 453−459.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word Learning: Homophony and the Distribution of Learning Exemplars. *Language Learning and Development, 12*(3), 231–251.

Dautriche, I., Fibla, L., Fievet, A.-C., Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive Psychology, 104*, 83–105. https://doi.org/10.1016/j.cogpsych.2018.04.001

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief Wakeful Resting Boosts New Memories Over the Long Term. *Psychological Science, 23*(9), 955–960.

Dobson, D. M. (1996). Snood [Computer game]. Calabasas, CA: Word of Mouse Games.

Duanmu, S. (2007). *The Phonology of Standard Chinese, Second Edition.* New York, NY: Oxford University Press.

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

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning, 60*(S2), 5-24. doi: 10.1111/j.1467-9922.2010.00598.x

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

Hao, Y.-C. (2018). Second Language Perception of Mandarin Vowels and Tones. *Language and Speech, 61*(1), 135–152. doi: 10.1177/0023830917717759

Hino, Y., Kusunose, Y., Lupker, S. J., & Jared, D. (2013). The Processing Advantage and Disadvantage for Homophones in Lexical Decision Tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 529–551.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393–1409.

Humiston, G. B., & Wamsley, E. J. (2018). A brief period of eye-closed rest enhances motor skill consolidation. *Neurobiology of Learning and Memory, 155*, 1–6. https://doi.org/10.1016/j.nlm.2018.06.002

Kann, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neuroscience, 9*:53. doi: 10.1186/1471-2202-9-53

Ke, J., Wang., F., & Coupe, C. (2002). The rise and fall of homophones: a window to language evolution. In *Proceedings of 4th International Conference on the Evolution of Language*. Harvard, MA.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica, 70*(1–2), 1–23.

Kong, Y.-Y., & Zeng, F.-G. (2006). Temporal and spectral cues in Mandarin tone recognition. *Journal of the Acoustical Society of America, 120*(5), 2830–2840. doi: 10.1121/1.2346009

Kucera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.

Kurziel, L. B. F., & Spencer, R. M. C. (2016). Consolidation of novel word learning in native English-speaking adults. *Memory, 24*(4), 471–481. doi: 10.1080/09658211.2015.1019889

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. doi: 10.18637/jss.v082.i13

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.3.01. https://CRAN.R-project.org/package=emmeans.

Li, C., Lin, C. Y., Wang, M., & Jiang, N. (2013). The activation of segmental and tonal information in visual word recognition. *Psychonomic Bulletin and Review, 20*(4), 773–779. doi: 10.3758/s13423-013-0395-2

Li, C., Wang, M., Davis, J. A., & Guan, C. Q. (2019). The role of segmental and tonal information in visual word recognition with learners of Chinese. *Journal of Research in Reading, 42*(2), 213–238. doi: 10.1111/1467-9817.12137

Lin, W. C. J. (1985). Teaching Mandarin Tones to Adult English Speakers: Analysis of Difficulties with Suggested Remedies. *RELC Journal, 16*(2), 31–47. https://doi.org/10.1177/003368828501600207

Lin, Y.-H. (2007). *The Sounds of Chinese*. New York, NY: Cambridge University Press.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech, 47*(2), 109–138.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62*, 407–420. doi: 10.1016/j.jml.2010.02.004

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

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 576–585.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. http://dx.doi.org/10.1016/j.jml.2017.01.001

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*(4), 710–716.

Ning, L.-H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication, 63–64*, 55–69. http://dx.doi.org/10.1016/j.specom.2014.05.001

Norman, J. (1988). *Chinese*. New York, NY: Cambridge University Press.

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophones Effects in Lexical Decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 139–156.

Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult Learning of Novel Words in a Non-native Language: Consonants, Vowels, and Tones. *Frontiers in Psychology, 9*:1211. doi: 10.3389/fpsyg.2018.01211

Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2016). VWPre: Tools for Preprocessing Visual World Data. Version 1.2.2, updated 2019-12-18, https://CRAN.R-project.org/package=VWPre.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ramachers, S., Brouwer, S. & Fikkert, P. (2017). How Native Prosody Affects Pitch Processing during Word Learning in Limburgian and Dutch Toddlers and Adults. *Frontiers in Psychology, 8*:1652. doi: 10.3389/fpsyg.2017.01652

Rubenstein, H., Lewis, S. S., Rubenstein, M. A. (1971). Evidence for Phonemic Recoding in Visual Word Recognition. *Journal of Verbal Learning and Verbal Behavior, 10*, 645–657.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology, 5*:1318. https://doi.org/10.3389/fpsyg.2014.01318

Salverda, A. P., & Tanenhaus, M. K. (2017). The Visual World Paradigm. In A. M. B. de Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the neurobiology of language: A practical guide* (pp. 89–110). Hoboken, NJ: Wiley Blackwell.

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

Shen, G., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *Journal of the Acoustical Society of America, 140*(6), 4396–4403. doi: 10.1121/1.4971765

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*(2), 145–156.

Shittu, S. (2019). Perception and Production of Yorùbá Tones by Young and Adult native Yorùbá speakers and native speakers of non-tone languages. PhD Thesis, University of Alberta. https://doi.org/10.7939/r3-k3j2-pz56

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research, 29*(2), 185–200. doi: 10.1177/0267658313480154

Singh, L., Poh, F. L. S., & Fu, C. S. L. (2016). Limits on Monolingualism? A Comparison of Monolingual and Bilingual Infants' Abilities to Integrate Tone in Novel Word Learning. *Frontiers in Psychology, 7*: 667. doi: 10.3389/fpsyg.2016.00667

SR Research Experiment Builder 1.10.1630 [Computer software]. (2015). Mississauga, Ontario, Canada: SR Research Ltd.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature, 437*, 1272–1278.

Storkel, H. L., & Maekawa, J. (2005). A comparison of homophony and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language, 32*(4), 827–853.

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

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*(5), 689–708. doi: 10.1080/01690960701728261

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *Journal of the Acoustical Society of America, 120*(4), 2285–2294. doi: 10.1121/1.2339290

Tsukada, K., Kondo, M., & Sunaoka, K. (2016). The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience. *Journal of Second Language Pronunciation, 2*(2), 225–252. doi: 10.1075/jslp.2.2.05tsu

Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research, 4*(2), 141–161. https://doi.org/10.1515/caslar-2015-0009

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2017). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.3.

Wamsley, E. J. (2019). Memory Consolidation during Waking Rest. *Trends in Cognitive Sciences, 23*(3), 171–173.

Wang, X., Wang, J., & Malins, J. G. (2017). Do you hear 'feather' when listening to 'rain'? Lexical tone activation during unconscious translation: Evidence from Mandarin-English bilinguals. *Cognition, 169*, 15–24.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*(6), 3649–3658. doi: 10.1121/1.428217

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36*, 250–267. doi: 10.1016/j.wocn.2007.06.004

Wee, L.-H., & Li, M. (2015). Modern Chinese Phonology. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics* (pp. 474–489). New York, Oxford University Press.

Wewalaarachchi, T. D., & Singh, L. (2020). Vowel, consonant, and tone variation exert asymmetrical effects on spoken word recognition: Evidence from 6-year-old monolingual and bilingual learners of Mandarin. *Journal of Experimental Child Psychology, 189*: 104698. https://doi.org/10.1016/j.jecp.2019.104698

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86–116.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R, 2 edition.* Boca Raton, Florida: Chapman and Hall/CRC Press.

Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics, 46*, 86–100. http://dx.doi.org/10.1016/j.wocn.2014.06.005

Yip, M. C. W., & Zhai, M. (2018). Context Effects and Spoken Word Recognition of Chinese: An Eye-Tracking Study. *Cognitive Science, 42*, 1134–1153.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia, 49*, 1761–1770. doi: 10.1016/j.neuropsychologia.2011.02.054

Zhu, X., & Wang, C. (2015). Tone. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics* (pp. 503–515). New York, NY: Oxford University Press.

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology Matters: The Phonological Frequency Effect in Written Chinese. *Psychological Science, 11*(3), 234–238.

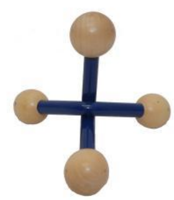
Zou, T., Chen, Y., & Caspers, J. (2017). The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Bilingualism: Language and Cognition, 20*(5), 1017–1029. doi: 10.1017/S1366728916000791

Appendix

A. List of the novel word-object mappings

No.	Novel word	Novel object
Practice 1	/musa/	
Practice 2	$/t^{h}usa/$	
1	$/p^{h}a^{1}sa^{1}/$	 (learned in tone contrast with $/p^{h}a^{2}sa^{1}/$)
		 (learned in consonant contrast with $/k^{h}a^{1}sa^{1}/$)
		 (learned in tone & consonant contrast with $/na^{2}sa^{1}/$)

2	$/p^ha^2sa^1/$	
3	$/t^ha^1sa^1/$	 (learned in tone contrast with $/t^ha^3sa^1/$)
		 (learned in tone & consonant contrast with $/la^3sa^1/$)
4	$/t^ha^3sa^1/$	
5	$/k^hu^1sa^1/$	 (learned in tone contrast with $/k^hu^4sa^1/$)
		 (learned in tone & consonant contrast with $/p^hu^4sa^1/$)

6	$/k^hu^4sa^1/$	
7	$/mu^2sa^1/$	 (learned in tone contrast with $/mu^3sa^1/$)
		 (learned in tone & consonant contrast with $/t^hu^3sa^1/$)
8	$/mu^3sa^1/$	
9	$/ni^4sa^1/$	

<table>
<tr><td rowspan="2">10</td><td rowspan="2">/ni^{2}sa^{1}/</td><td>

(learned in tone contrast with /ni^{4}sa^{1}/)</td></tr>
<tr><td>

(learned in tone & consonant contrast with /mi^{4}sa^{1}/)</td></tr>
<tr><td rowspan="2">11</td><td rowspan="2">/li^{3}sa^{1}/</td><td>

(learned in tone contrast with /li^{4}sa^{1}/)</td></tr>
<tr><td>

(learned in tone & consonant contrast with /k^{h}i^{4}sa^{1}/)</td></tr>
<tr><td>12</td><td>/li^{4}sa^{1}/</td><td>

(learned in tone contrast with /li^{3}sa^{1}/)</td></tr>
</table>

		 (learned in consonant contrast with $/p^hi^4sa^1/$)
13	$/k^ha^1sa^1/$	
14	$/t^ha^2sa^1/$	
15	$/ma^2sa^1/$	
16	$/k^hu^3sa^1/$	
17	$/nu^3sa^1/$	
18	$/mu^4sa^1/$	

19	/lu^{4}sa^{1}/	
20	/ni^{1}sa^{1}/	
21	/t^{h}i^{1}sa^{1}/	
22	/p^{h}i^{4}sa^{1}/	
23	/na^{2}sa^{1}/	
24	/la^{3}sa^{1}/	
25	/p^{h}u^{4}sa^{1}/	

26	/tʰu³sa¹/	
27	/mi⁴sa¹/	
28	/kʰi⁴sa¹/	

B. List of the novel objects used as distractors in Test Phase II

No.	Distractors
1	
2	
3	
4	
5	
6	

7	
8	
9	

C. List of the novel word pairs and target novel words in Test Phase I

Trial No.	Novel word pairs learned	Phonological contrast	Target novel word	Word code
1	$/p^{h}a^{1}sa^{1}/$ - $/k^{h}a^{1}sa^{1}/$	Cons_Contrast	$/p^{h}a^{1}sa^{1}/$	1
2	$/li^{3}sa^{1}/$ - $/li^{4}sa^{1}/$	Ton_Contrast	$/li^{4}sa^{1}/$	2
3	$/t^{h}a^{1}sa^{1}/$ - $/la^{3}sa^{1}/$	Cons&Ton_Contrast	$/t^{h}a^{1}sa^{1}/$	3
4	$/k^{h}u^{1}sa^{1}/$ - $/k^{h}u^{4}sa^{1}/$	Ton_Contrast	$/k^{h}u^{1}sa^{1}/$	4
5	$/p^{h}a^{1}sa^{1}/$ - $/na^{2}sa^{1}/$	Cons&Ton_Contrast	$/na^{2}sa^{1}/$	5
6	$/mu^{4}sa^{1}/$ - $/lu^{4}sa^{1}/$	Cons_Contrast	$/mu^{4}sa^{1}/$	6
7	$/k^{h}u^{3}sa^{1}/$ - $/nu^{3}sa^{1}/$	Cons_Contrast	$/nu^{3}sa^{1}/$	7
8	$/p^{h}a^{1}sa^{1}/$ - $/p^{h}a^{2}sa^{1}/$	Ton_Contrast	$/p^{h}a^{2}sa^{1}/$	8
9	$/li^{3}sa^{1}/$ - $/k^{h}i^{4}sa^{1}/$	Cons&Ton_Contrast	$/k^{h}i^{4}sa^{1}/$	9
10	$/t^{h}a^{1}sa^{1}/$ - $/t^{h}a^{3}sa^{1}/$	Ton_Contrast	$/t^{h}a^{1}sa^{1}/$	3
11	$/ni^{1}sa^{1}/$ - $/t^{h}i^{1}sa^{1}/$	Cons_Contrast	$/t^{h}i^{1}sa^{1}/$	10
12	$/k^{h}u^{1}sa^{1}/$ - $/p^{h}u^{4}sa^{1}/$	Cons&Ton_Contrast	$/k^{h}u^{1}sa^{1}/$	4
13	$/t^{h}a^{2}sa^{1}/$ - $/ma^{2}sa^{1}/$	Cons_Contrast	$/t^{h}a^{2}sa^{1}/$	11
14	$/mu^{2}sa^{1}/$ - $/t^{h}u^{3}sa^{1}/$	Cons&Ton_Contrast	$/t^{h}u^{3}sa^{1}/$	12
15	$/ni^{2}sa^{1}/$ - $/ni^{4}sa^{1}/$	Ton_Contrast	$/ni^{2}sa^{1}/$	13
16	$/li^{4}sa^{1}/$ - $/p^{h}i^{4}sa^{1}/$	Cons_Contrast	$/p^{h}i^{4}sa^{1}/$	14
17	$/mu^{2}sa^{1}/$ - $/mu^{3}sa^{1}/$	Ton_Contrast	$/mu^{3}sa^{1}/$	15
18	$/ni^{2}sa^{1}/$ - $/mi^{4}sa^{1}/$	Cons&Ton_Contrast	$/ni^{2}sa^{1}/$	13

Note: In the “Phonological contrast” variable, “Cons_Contrast” indicates novel words learned in a pair with with consonant contrasts; “Ton_Contrast” with tone contrasts; “Cons&Ton_Contrast” with both consonant and tone contrasts.

D. List of the target novel words in Test Phase II

Trial No.	Target novel word	Phonological contrast	Competition	Word code
1	/p^{h}a^{1}sa^{1}/	Cons_Contrast	Yes	1
2	/li^{3}sa^{1}/	Ton_Contrast	No	2
3	/t^{h}a^{1}sa^{1}/	Cons&Ton_Contrast	Yes	3
4	/k^{h}u^{1}sa^{1}/	Ton_Contrast	No	4
5	/p^{h}a^{1}sa^{1}/	Cons&Ton_Contrast	Yes	1
6	/mu^{4}sa^{1}/	Cons_Contrast	No	5
7	/k^{h}u^{3}sa^{1}/	Cons_Contrast	Yes	6
8	/p^{h}a^{1}sa^{1}/	Ton_Contrast	No	1
9	/li^{3}sa^{1}/	Cons&Ton_Contrast	Yes	2
10	/t^{h}a^{1}sa^{1}/	Ton_Contrast	No	3
11	/ni^{1}sa^{1}/	Cons_Contrast	Yes	7
12	/k^{h}u^{1}sa^{1}/	Cons&Ton_Contrast	No	4
13	/t^{h}a^{2}sa^{1}/	Cons_Contrast	Yes	8
14	/mu^{2}sa^{1}/	Cons&Ton_Contrast	No	9
15	/ni^{2}sa^{1}/	Ton_Contrast	Yes	10
16	/li^{4}sa^{1}/	Cons_Contrast	No	11
17	/mu^{2}sa^{1}/	Ton_Contrast	Yes	9
18	/ni^{2}sa^{1}/	Cons&Ton_Contrast	No	10
19	/mi^{4}sa^{1}/	Cons&Ton_Contrast	Yes	12
20	/k^{h}a^{1}sa^{1}/	Cons_Contrast	No	13
21	/li^{4}sa^{1}/	Ton_Contrast	Yes	11
22	/la^{3}sa^{1}/	Cons&Ton_Contrast	No	14
23	/k^{h}u^{4}sa^{1}/	Ton_Contrast	Yes	15

24	/na^2sa^1/	Cons&Ton_Contrast	No	16
25	/lu^4sa^1/	Cons_Contrast	Yes	17
26	/nu^3sa^1/	Cons_Contrast	No	18
27	/$p^ha^2sa^1$/	Ton_Contrast	Yes	19
28	/$k^hi^4sa^1$/	Cons&Ton_Contrast	No	20
29	/$t^ha^3sa^1$/	Ton_Contrast	Yes	21
30	/$t^hi^1sa^1$/	Cons_Contrast	No	22
31	/$p^hu^4sa^1$/	Cons&Ton_Contrast	Yes	23
32	/ma^2sa^1/	Cons_Contrast	No	24
33	/$t^hu^3sa^1$/	Cons&Ton_Contrast	Yes	25
34	/ni^4sa^1/	Ton_Contrast	No	26
35	/$p^hi^4sa^1$/	Cons_Contrast	Yes	27
36	/mu^3sa^1/	Ton_Contrast	No	28
37	/ni^2sa^1/	Cons&Ton_Contrast	Yes	10
38	/$p^ha^1sa^1$/	Cons_Contrast	No	1
39	/li^3sa^1/	Ton_Contrast	Yes	2
40	/$t^ha^1sa^1$/	Cons&Ton_Contrast	No	3
41	/$k^hu^1sa^1$/	Ton_Contrast	Yes	4
42	/$p^ha^1sa^1$/	Cons&Ton_Contrast	No	1
43	/mu^4sa^1/	Cons_Contrast	Yes	5
44	/$k^hu^3sa^1$/	Cons_Contrast	No	6
45	/$p^ha^1sa^1$/	Ton_Contrast	Yes	1
46	/li^3sa^1/	Cons&Ton_Contrast	No	2
47	/$t^ha^1sa^1$/	Ton_Contrast	Yes	3
48	/ni^1sa^1/	Cons_Contrast	No	7

49	/$k^{h}u^{1}sa^{1}$/	Cons&Ton_Contrast	Yes	4
50	/$t^{h}a^{2}sa^{1}$/	Cons_Contrast	No	8
51	/$mu^{2}sa^{1}$/	Cons&Ton_Contrast	Yes	9
52	/$ni^{2}sa^{1}$/	Ton_Contrast	No	10
53	/$li^{4}sa^{1}$/	Cons_Contrast	Yes	11
54	/$mu^{2}sa^{1}$/	Ton_Contrast	No	9
55	/$k^{h}a^{1}sa^{1}$/	Cons_Contrast	Yes	13
56	/$li^{4}sa^{1}$/	Ton_Contrast	No	11
57	/$la^{3}sa^{1}$/	Cons&Ton_Contrast	Yes	14
58	/$k^{h}u^{4}sa^{1}$/	Ton_Contrast	No	15
59	/$na^{2}sa^{1}$/	Cons&Ton_Contrast	Yes	16
60	/$lu^{4}sa^{1}$/	Cons_Contrast	No	17
61	/$nu^{3}sa^{1}$/	Cons_Contrast	Yes	18
62	/$p^{h}a^{2}sa^{1}$/	Ton_Contrast	No	19
63	/$k^{h}i^{4}sa^{1}$/	Cons&Ton_Contrast	Yes	20
64	/$t^{h}a^{3}sa^{1}$/	Ton_Contrast	No	21
65	/$t^{h}i^{1}sa^{1}$/	Cons_Contrast	Yes	22
66	/$p^{h}u^{4}sa^{1}$/	Cons&Ton_Contrast	No	23
67	/$ma^{2}sa^{1}$/	Cons_Contrast	Yes	24
68	/$t^{h}u^{3}sa^{1}$/	Cons&Ton_Contrast	No	25
69	/$ni^{4}sa^{1}$/	Ton_Contrast	Yes	26
70	/$p^{h}i^{4}sa^{1}$/	Cons_Contrast	No	27
71	/$mu^{3}sa^{1}$/	Ton_Contrast	Yes	28
72	/$mi^{4}sa^{1}$/	Cons&Ton_Contrast	No	12

Note: In the "Competition" variable, "Yes" means that there was a competitor in the visual stimuli, i.e. the three objects were the target, its competitor and a distractor; "No" means that there was no competitor, i.e. the three objects were the target and two distractors.