CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Phonology, homophony, and eyes-closed rest in Mandarin novel word learning: An eye-tracking study in adult native and non-native speakers

Wenfu Bao[1,2,3] , Anja Arnhold[1] and Juhani Järvikivi[1]

[1]Department of Linguistics, University of Alberta, Edmonton, AB, Canada, [2]Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada and [3]Rehabilitation Sciences Institute, University of Toronto, Toronto, ON, Canada
Corresponding author: Wenfu Bao; Email: wenfu.bao@mail.utoronto.ca

Abstract
This study used the visual world paradigm to investigate novel word learning in adults from different language backgrounds and the effects of phonology, homophony, and rest on the outcome. We created Mandarin novel words varied by types of phonological contrasts and homophone status. During the experiment, native ($n = 34$) and non-native speakers (English; $n = 30$) learned pairs of novel words and were tested twice with a 15-minute break in between, which was spent either resting or gaming. In the post-break test of novel word recognition, an interaction appeared between language backgrounds, phonology, and homophony: non-native speakers performed less accurately than native speakers only on non-homophones learned in pairs with tone contrasts. Eye movement data indicated that non-native speakers' processing of tones may be more effortful than their processing of segments while learning homophones, as demonstrated by the time course. Interestingly, no significant effects of rest were observed across language groups; yet after gaming, native speakers achieved higher accuracy than non-native speakers. Overall, this study suggests that Mandarin novel word learning can be affected by participants' language backgrounds and phonological and homophonous features of words. However, the role of short periods of rest in novel word learning requires further investigation.

Phonology, homophony, and eyes-closed rest in Mandarin novel word learning: An eye-tracking study in adult native and non-native speakers

When it comes to learning Mandarin, lexical tones are an inevitable topic. There are four main tones in Mandarin: tone 1 is high-level, tone 2 is high-rising, tone 3 is low-dipping, and tone 4 is high-falling (Chao, 1930; Zhang & Lai, 2010; Zhu & Wang, 2015). This tonal system is notorious for being difficult to learn for

non-native speakers. While some research shows that tonal contrasts are more difficult for learners from a non-tonal first language (L1) background than for those from a tonal L1 background (Cooper & Wang, 2012; Poltrock et al., 2018; Schaefer & Darcy, 2020), speakers of tonal L1s do not necessarily outperform speakers of non-tonal L1s in non-native tone perception and word learning (Laméris & Post, 2022; So & Best, 2010). However, another characteristic of Mandarin—the abundance of homophones—is often neglected in adult lexical learning. In the current study, we investigated the combination of these two Mandarin features, phonology and homophony, in a novel word learning experiment that tested both native and non-native speakers. In addition, we were interested in whether participants' performance would be affected by a short break with either an eyes-closed rest or a distractor task (i.e., playing a computer game). The available evidence suggests that short periods of rest help memory consolidation and facilitate learning, in a similar manner to sleep (Brokaw et al., 2016; Dewar et al., 2012; Wamsley, 2019).

Mandarin tonal contrasts and novel word learning

A number of studies have demonstrated the challenge of learning Mandarin tonal contrasts in adult non-native speakers, who come from a variety of linguistic backgrounds such as English (Hao, 2018, 2023; Shen & Froud, 2016), Dutch (Sadakata & McQueen, 2014; Zou et al., 2017), Thai (Wu et al., 2014), and Japanese (Tsukada et al., 2016). These findings (alongside others, e.g., Kann et al., 2008; Zhu et al., 2021) suggest that despite difficulties, adult speakers are capable of learning non-native tones to a certain degree. This is especially the case in tonal L1 speakers (Chan & Leung, 2020). For example, Wayland and Li (2008) reported that Mandarin-L1 speakers outperformed English-L1 speakers when discriminating the low- vs. mid-tone contrast in Thai. Similarly, using a Cantonese tone discrimination task, Qin and Mok (2011) found that Mandarin-L1 speakers performed significantly better than English-L1 and French-L1 speakers. Nonetheless, both L1 tonal status and L1 tone type affect individual performance in tonal learning (Laméris & Post, 2022), and tonal L1 speakers may perform worse than their non-tonal L1 counterparts in certain psycholinguistic tasks. For instance, Hao (2012) observed significantly lower accuracy among Cantonese-L1 speakers when perceiving Mandarin tone 4 and producing Mandarin tone 1 than among English-L1 speakers. Therefore, the specific tonal contrasts involved appear to be crucial when comparing performance between tonal and non-tonal L1 learners. However, the existing literature has primarily focused on speech perception and production (Ning et al., 2014; Shen & Froud, 2016; Tsukada et al., 2015; Wang et al., 1999), with much less research conducted on novel word learning.

Novel words can be defined as pseudowords that are phonologically legal but not associated with any meaning (Ziegler et al., 1997), i.e., they are possible, but not existing words. A majority of lexical learning studies using novel word learning designs pertain to infants and young children (Chan et al., 2011; Chen & Liu, 2014), especially comparing monolingual and bilingual infants (Burnham et al., 2018; Byers-Heinlein & Werker, 2013; Singh et al., 2016; Wewalaarachchi & Singh, 2020).

Among the relatively few studies with adults, most have focused on tonal learning. For example, Wong and Perrachione (2007) found that non-native tones can be learned by English-L1 speakers when identifying English pseudowords superimposed with pitch patterns resembling Mandarin tones. Also using a lexical identification task, Laméris and Post (2022) suggested that "tonal rather than segmental distinctions were the hardest feature to memorize in the pseudolanguage words" (p. 857) for both English-L1 and Mandarin-L1 speakers, given that most errors were tone-only errors. Similar findings were observed in Chang and Bowles (2015), who reported more tonal errors than segmental errors by English-L1 speakers while learning disyllabic Mandarin pseudowords. Note that most of these studies, like the present one, used stimuli made up of segments that occur in both Mandarin and English. Therefore, while the pseudowords were new to the learners, the segmental contrasts distinguishing them were already familiar from English, as pointed out to us by an anonymous reviewer (but see Wright & Baese-Berk, 2022, on learning Thai tone contrasts and phonotactics).

In addition to tonal learning, several studies have investigated the role of tonal and segmental information in Mandarin spoken word recognition (Li et al., 2018; Malins & Joanisse, 2010; Sereno & Lee, 2015). For example, Wiener et al. (2018, 2021) conducted eye-tracking studies in which English-L1 speakers were trained on an artificial tonal language that mimics Mandarin's syllable-tone combinations. Results indicate that both intermediate learners of Mandarin and naïve listeners used co-occurring tonal and segmental cues during spoken word recognition. Nevertheless, few studies have explicitly compared the difference between native and non-native speakers when learning Mandarin tones and segments, though findings from Cantonese learning may shed light on this question. For instance, Poltrock et al. (2018) found that native speakers learned Cantonese pseudowords better than non-native speakers; within the non-native group, Mandarin-L1 speakers outperformed their French-L1 counterparts in pseudowords that differed minimally in tones.

Homophony effects in lexical processing

Homophony is a common linguistic phenomenon, where pairs or groups of words have identical pronunciation but different meanings. However, the proportion of homophones varies across languages. It has been reported that 3.2% of English consists of homonyms (i.e., same pronunciation and spelling), whereas 11.6% of Mandarin is homophonous (Neergaard et al., 2021; Wen, 1980). According to Duanmu (2007), the average homophone density is 1.4 words per monosyllable in English (e.g., *meat* and *meet*), but five words per syllable in Mandarin. Some forms like *shì* (i.e., /ʂɻ⁴/ with the falling tone 4; the superscript number indicates the lexical tone of the syllable) correspond to nearly 30 unique homophonous morphemes (Wiener et al., 2018). An example of Mandarin disyllabic homophones is *shǒushì* /ʂoʊ³ʂɻ⁴/, which has two meanings "gesture" (occurring 37 times in a corpus of five million words) and "jewelry" (occurring 44 times in the same corpus; Chen et al., 1996; Huang et al., 1998).

Accordingly, different effects of homophony have been observed in English and Mandarin when processed by native speakers. For example, in lexical decision tasks, English homophones show longer response latencies and lower accuracy than non-homophones, especially among low-frequency words (Rubenstein et al., 1971). The size of homophony effects is found to be determined by the frequency of the homophone and its mate, as well as by the orthographic and phonological characteristics of the homophone (Pexman et al., 2001). Contrarily, Mandarin homophones exhibit facilitative effects (i.e., faster processing than non-homophones; Ziegler et al., 2000), which are also observed in non-native learners of Mandarin (Liu & Wiener, 2020, 2022). This homophony advantage is often attributed to the large amount of homophonic mates in Mandarin, which increases the phonological familiarity of a homophone and facilitates processing (Ziegler et al., 2000).

To explore whether the distinct homophony effects in English and Mandarin can be explained by script type, Hino et al. (2013) studied kanji, which are logographic Chinese characters used in Japanese writing. Native Japanese speakers showed an inhibitory effect for homophones with only one homophonic mate, but a facilitatory effect for those with multiple mates. Therefore, they concluded that the pattern is predicted by the number of homophonic mates that the target homophone has, rather than by the script type.

However, most studies in the previous literature focus on visual word recognition using the lexical decision task. Spoken language has received relatively less attention. Moreover, even though novel word learning was used in prior research, most work addresses homophony learning in infants and young children (Dautriche et al., 2016, 2018; Ramachers et al., 2017; Storkel & Maekawa, 2005). Yet until recently, some adult studies have examined non-native learning of spoken homophones in Mandarin. In Liu and Wiener (2020, 2022), for example, native English speakers from a second-semester Mandarin class learned monosyllabic tonal pairs over three days. A facilitative homophone effect was observed when the talker variability was low. That is, new words that were homophonous with previously learned words were identified more accurately and recognized faster than those that were not homophonous, if learners were familiar with a single speaker's voice.

Memory consolidation following learning

Studies have established sleep's effects on memory enhancement after learning, when compared to an equivalent period of wakefulness (Diekelmann & Born, 2010; Korman et al., 2007; Stickgold, 2005). This benefit has also been validated in lexical learning research for English (Fenn et al., 2003; Weighall et al., 2017) and Cantonese (Qin & Zhang, 2019). For example, Kurdziel and Spencer (2016) found that in a task of recalling novel words after a 12-hour delay that included sleep, English speakers outperformed their peers who spent the time awake. In terms of non-native tonal learning, Qin and Zhang (2019) trained Mandarin listeners on identifying Cantonese tonal contrasts and found that those who had an overnight sleep between training and later post-tests showed improved performance in identification accuracy. Recently, Qin et al. (2022) demonstrated that this overnight tone

consolidation effect can be facilitated by a learner's ability to detect pitch height differences.

Furthermore, even short periods of unoccupied rest can induce memory consolidation. Previous studies often categorize participants into two conditions: the waking rest condition where participants have an unoccupied rest and the distractor task condition where participants play a computer game. For instance, Brokaw et al. (2016) suggested that a 15-minute interval of waking rest can facilitate verbal declarative memory in a story recalling test, relative to playing a computer game as distraction (also see Humiston & Wamsley, 2018). Likewise, brief periods of wakeful resting enhance memory retention in a range of sensorimotor and cognitive tasks (Dewar et al., 2012; Wamsley, 2019). However, little research has addressed whether a 15-minute period of unoccupied rest can affect novel word learning, especially among learners from both native and non-native backgrounds.

The present study

The first aim of this study was to investigate the roles of phonology and homophony in Mandarin novel word learning among native and non-native (English) speakers. By combining different Mandarin tones and segments, we created a set of novel words that consisted of homophones and non-homophones and varied in types of phonological contrasts: consonant contrasts, tone contrasts, or both. To assess the time course of online processing after the learning, we tested novel spoken word recognition using the visual world paradigm, which allows for examining participants' interpretation of linguistic stimuli without interruption (Dahan & Tanenhaus, 2005). The second aim was to explore whether brief periods of unoccupied rest would aid memory consolidation following the learning session. To this end, we designed the experiment with a break between the learning and test phases and compared the performance of participants who spent this interval either resting or gaming.

The novel word learning task consisted of three sections. The first one was an integrated Learning Phase/Test Phase I. In each trial, participants learned novel words via word-object associations and were then tested on the target word by selecting its matching object. The second section was a 15-minute break, during which participants either had an eyes-shut rest or played a computer game. The last section, Test Phase II, was a visual world eye-tracking experiment. Specifically, participants attended to three objects on a computer screen while listening to a novel word and selected the target object corresponding to what they heard. The target object was presented in two settings: in the competitor setting, it was presented with a competitor picture and a distractor picture; in the no-competitor setting, it was presented with two distractor pictures. Participants' response accuracy in Test Phases I and II was collected, and their eye movements in Test Phase II were also analyzed.

We asked the following four research questions: (i) Do participants' language backgrounds modulate their novel word learning outcome? Given that native speakers are familiar with the Mandarin phonological system and the task simulates their native word learning, we predicted that they would outperform non-native

speakers overall, reflected by higher accuracy and more/earlier looks to the target object. (ii) Do phonological contrasts between novel words predict the outcome, and does this depend on participants' language backgrounds? Considering previous evidence on the difficulty of tonal learning in non-native speakers (Chang & Bowles, 2015; Laméris & Post, 2022), an interaction between phonological contrasts and language backgrounds was anticipated; in particular, English speakers would likely respond to novel words learned with tone contrasts least accurately. (iii) Is the learning outcome affected by the homophone status of novel words, and does this depend on participants' language backgrounds? In light of different homophone effects in English (Pexman et al., 2001; Rubenstein et al., 1971) and Mandarin (Ziegler et al., 2000) as well as non-native speakers' lack of Mandarin experience, it was predicted that homophony would interact with language backgrounds to affect the outcome, and a facilitative homophone effect may occur in native speakers only. (iv) Do different break types predict the outcome, and is this affected by participants' language backgrounds? Since tonal learning can be consolidated overnight (Qin & Zhang, 2019; Qin et al., 2022) and brief periods of rest enhance memory as well (Brokaw et al., 2016; Dewar et al., 2012; Wamsley, 2019), it was hypothesized that participants who had a rest would outperform those who played the game, regardless of their language backgrounds.

Method

Participants

We recruited 68 adults via an online participant pool or flyers at the University of Alberta in Edmonton, Canada: 67 were students who participated in exchange for course credits, and one was an employee who received a small payment. Four native English speakers were excluded due to previous experience with Mandarin and/or Japanese, considering that the pitch features of Japanese may assist Mandarin tone learning (Caldwell-Harris et al., 2015; So & Best, 2010). Therefore, a total of 64 participants were included in the final analysis: 34 native Mandarin speakers who were from mainland China (25 female; $M_{age} = 20.5$ years, range = 18–27, $SD = 1.86$), and 30 native English speakers who were from Canada and without tonal language background or experience (21 female; $M_{age} = 22.9$ years, range = 18–63, $SD = 9.86$). While an a priori power analysis was not performed, the sample size of each language group in the current study was largely comparable to that reported in previous research (Brokaw et al., 2016; Poltrock et al., 2018; Qin & Zhang, 2019; Ziegler et al., 2000).

Given the facilitative effects of musicality on linguistic tone learning (Alexander et al., 2005; Wong & Perrachione, 2007), we collected information about participants' musical background using the Goldsmiths Musical Sophistication Index, which consists of five dimensions: active engagements, perceptual abilities, musical training, singing abilities, and emotions (i.e., emotional responses to music; Müllensiefen et al., 2014). This index was included as a variable of interest in statistical modeling. All participants reported normal or corrected-to-normal vision and normal hearing and gave written informed consent prior to participation. The

Table 1. Mandarin novel words

Phonological contrast		Novel word pair	
Tone contrast	1	**/pʰa¹sa¹/**	/pʰa²sa¹/
	2	**/tʰa¹sa¹/**	/tʰa³sa¹/
	3	**/kʰu¹sa¹/**	/kʰu⁴sa¹/
	4	**/mu²sa¹/**	/mu³sa¹/
	5	**/ni²sa¹/**	/ni⁴sa¹/
	6	**/li³sa¹/**	**/li⁴sa¹/**
Consonant contrast	1	**/pʰa¹sa¹/**	/kʰa¹sa¹/
	2	/tʰa²sa¹/	/ma²sa¹/
	3	/kʰu³sa¹/	/nu³sa¹/
	4	/mu⁴sa¹/	/lu⁴sa¹/
	5	/ni¹sa¹/	/tʰi¹sa¹/
	6	**/li⁴sa¹/**	/pʰi⁴sa¹/
Consonant and tone contrast (Both contrasts)	1	**/pʰa¹sa¹/**	/na²sa¹/
	2	**/tʰa¹sa¹/**	/la³sa¹/
	3	**/kʰu¹sa¹/**	/pʰu⁴sa¹/
	4	**/mu²sa¹/**	/tʰu³sa¹/
	5	**/ni²sa¹/**	/mi⁴sa¹/
	6	**/li³sa¹/**	/kʰi⁴sa¹/

Note. Novel words in bold are homophones.

experiment protocol and consent procedures were reviewed and approved by the Research Ethics Board of the University of Alberta.

Materials

Novel words

We created 18 pairs of Mandarin disyllabic novel words (see Table 1). The first syllables were generated by integrating the six consonants /pʰ, tʰ, kʰ, m, n, l/, three vowels /a, u, i/, and four lexical tones in Mandarin, whereas the second syllables were constantly /sa¹/. We avoided phonological combinations that are real words in Mandarin (i.e., existing Mandarin disyllabic words were not used). Thus, 28 novel words were created, of which seven (25%) are homophones: /tʰa¹sa¹/, /kʰu¹sa¹/, /mu²sa¹/, /ni²sa¹/, /li³sa¹/, /li⁴sa¹/, and /pʰa¹sa¹/. We were not interested in controlling potential frequency effects; participants encountered phonological forms of homophones more often than those of non-homophones since they learned homophonous forms with two (for /pʰa¹sa¹/, three) different meanings.

There are three phonological contrast groups. First, pairs in the tone contrast group differ in the first syllables' tones only, representing all six possible tonal

comparisons in Mandarin (i.e., tone 1/tone 2, tone 1/tone 3, tone 1/tone 4, tone 2/tone 3, tone 2/tone 4, tone 3/tone 4). Second, novel words in the consonant contrast group differ in the first syllables' consonants only. To even out the number of pairs in each group, tones 1 and 4 were used more than once in this group because they are less difficult than the other two tones for native English speakers (Chang & Bowles, 2015). Third, the consonant and tone contrast (hereafter referred to as both contrasts) group incorporates segmental and tonal differences within the first syllables.

Further, another two novel words /musa/ and /tʰusa/, carrying the Mandarin neutral tone in both syllables, were created for the practice trial. The first author, a native Mandarin speaker from northern China, recorded all novel words in a sound-attenuated booth and manually split them into separate sound files using *Praat* (Boersma & Weenink, 2020).

Novel objects
From the *Novel Object and Unusual Name Database* (Horst & Hout, 2016), we selected 47 novel objects that control for complexity in color, shape, and material. These objects were randomly assigned to the created novel words for participants to learn (see Appendix A in Supplementary Materials for a list of the word-object mappings). Except for the two objects in the practice trial, all were used in experimental trials: 36 as target or competitor objects and nine as distractor objects (see Appendix B in Supplementary Materials for a list of the distractor objects).

Procedure
Each experimental session was composed of three sections, as illustrated in Figure 1: the integrated Learning Phase/Test Phase I, Break, and Test Phase II. It took approximately 35 minutes to complete the whole session.

Learning Phase/Test Phase I
In each trial, participants learned one of the 18 pairs of novel words and then tested on one of them (cf. Figure 1). During the learning, a novel object randomly appeared at four different locations on a display screen (24 inches, 1,920 × 1,080 pixel resolution), with its name (i.e., the novel word) presented auditorily twice at each place. Then, the second object went through the same process. Thereby, within each trial, participants were presented with the novel word/object 8 times (4 visual displays × 2 audio plays). During the test, the two novel objects simultaneously appeared at the center of the screen, but only one novel word was played. By pulling one of the two triggers on a gamepad, participants matched the novel word they heard to its corresponding object. Within the paired novel words, the tested one was labeled as the target (referring to the target object) and the other one as the competitor (referring to the competitor object). No feedback was given to participants after they responded.

This section was programmed with *Experiment Builder* (SR Research, 2015). Participants' eye movements were tracked with an SR Research *EyeLink 1000* desktop mount eye tracker, which was run on a Dell PC with a sampling rate of
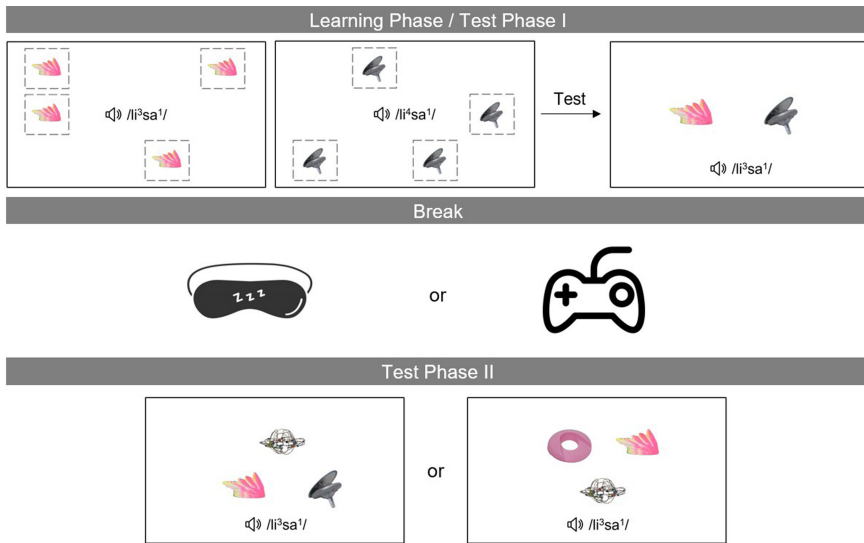
Figure 1. Novel word learning task.

Note. The task consisted of three sections: Learning Phase/Test Phase I, Break, and Test Phase II. During the integrated Learning Phase/Test Phase I, participants learned pairs of novel words (e.g., /li^{3}sa^1/ and /li^{4}sa^1/; the glosses illustrate novel words played auditorily and did not appear on the screen during the experiment) and then tested on one of them (e.g., /li^{3}sa^1/). Afterward, there was a 15-min break when participants either had an eyes-shut rest or played a game. In Test Phase II, participants were tested on one of the novel words (e.g., /li^{3}sa^1/) again, which was presented either with its competitor (/li^{4}sa^1/) and a distractor (i.e., the competitor setting on the left) or with two distractors (i.e., the no-competitor setting on the right).

either 500 Hz or 1,000 Hz. There was one practice trial for participants to familiarize themselves with the task, followed by 18 experimental trials (see Appendix C in Supplementary Materials). It took around 10 minutes to complete this section. We collected participants' responses and eye movements, but only analyzed the former.

Break

During the break, participants were randomly assigned to either the game or the rest group. In the game group, participants played the computer game "Snood" as a distractor task for 15 minutes. In this game, players clear blocks of colors by joining three or more icons of the same color. It was chosen because the game is engaging and involves only minimal hand and eye movements. In the rest group, participants wore an eye mask while sitting back or lying down on a reclining chair and had an eyes-closed rest for the same duration.

Test Phase II

At each trial, three novel objects simultaneously appeared on the screen, some of which had occurred before; the name of one of them (i.e., the target word) was played auditorily, while participants' eye gaze to the presented pictures was tracked. The target object was presented in two settings, either with or without its

competitor. Specifically, in the competitor setting, the target object occurred with its competitor and one unrelated distractor object; in the no-competitor setting, the target object appeared with two unrelated distractor objects that participants had not encountered during the learning. Participants were instructed to use the computer mouse to click on the target object that matched the spoken novel word they heard. No feedback was provided after their responses.

This section was programmed with *Experiment Builder* (SR Research, 2015) as well, and participants' eye movements were recorded using the same eye tracker. Since each novel word (cf. Table 1) served once as the target and once as the competitor, there were altogether 72 trials (18 novel word pairs × 2 targets × 2 competition settings; see Appendix D in Supplementary Materials for a list of the 72 trials). It took around 10 minutes to complete this section. We collected and analyzed participants' response and eye movement data.

Data analysis

Response data

We analyzed response accuracy in Test Phases I and II. Using the *lme4* package (v1.1-28; Bates et al., 2015), we built generalized linear mixed-effects models in *R* (v4.1.1; R Core Team, 2021). Estimated *p*-values were obtained through the *lmerTest* package (v3.1-3; Kuznetsova et al., 2017). The dependent variable was *Accuracy*, which was coded as a binary variable (Correct vs. Incorrect). Accordingly, we specified the family argument as binomial in model construction. For fixed effects, the independent variables consisted of *Language Background* (Native Mandarin vs. Native English), *Homophony* (Homophone vs. Non-homophone), *Phonological Contrast* (Consonant Contrast vs. Tone Contrast vs. Both Contrasts), and *Musical Sophistication* (Active Engagements vs. Perceptual Abilities vs. Musical Training vs. Singing Abilities vs. Emotions) in Test Phase I, as well as another two: *Break Type* (Rest vs. Game) and *Competition* (Competitor vs. No-competitor) in Test Phase II. For random effects, we included random intercepts for *Subject* and *Word* and tested random slopes for all of the manipulated factors mentioned above.

To attain the optimal model, fixed-effects and random-effects structures were fitted separately using a stepwise backward approach. Specifically, we used the *anova* function to compare a complex model to its simpler version which had one component removed, by inspecting the estimated *p*-value and Akaike Information Criterion (AIC) value. 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 (Matuschek et al., 2017). After the optimal model was determined, we used the *emmeans* package (v1.7.2; Lenth, 2022) to conduct multiple comparisons for significant factors and interactions and applied the Tukey adjustment method to guard against Type I error. We report significant comparisons ($p < 0.05$) below.

To visualize data, *Accuracy* was converted to a numeric variable as *Accuracy Proportion*. Effects of significant factors and interactions were visualized using the

ggplot2 (v3.3.5; Wickham, 2016) and *lattice* (v0.20-45; Sarkar, 2008) packages. To indicate significant group differences on the graphics, the *ggsignif* package (v0.6.3; Constantin & Patil, 2021) was used.

Eye movement data

Eye movements in Test Phase II only were analyzed. For preprocessing, data were output as a sample report using the SR Research EyeLink *Data Viewer* and then preprocessed using the *VWPre* package (v1.2.4; Porretta et al., 2016). To index each unique recording sequence of eye movements, *Event* was created as a combination of *Subject* and *Trial*. Prior to binning the data, we identified events with excessive trackloss. Specifically, 171 events (3.71%) were detected with more than 25% trackloss data and thus were removed. We were only interested in eye fixations that occurred in the target interest area (i.e., target looks). In order to obtain proportion looks, we aggregated the sample data into 20 ms bins, and each bin was coded as *Target* or *Nontarget*. Since proportions are not suitable for analysis, empirical logits (Barr, 2008) were calculated for the binned target proportion data, within the time window from 200 to 1,700 ms after the target stimulus onset. We used 200 ms as the starting point because it usually takes around that duration of time to program a saccade and execute an eye movement (Fischer, 1992).

Using the *mgcv* package (v1.8-39; Wood, 2017), we built generalized additive mixed-effects models (GAMMs), as they are well designed for analyzing time series data (Baayen et al., 2017). Eye movements in the correct trials ($n = 3,304$) only were modeled. The dependent variable was the logit-transformed proportions of target looks. The independent variables were the same as those in the accuracy analysis. The fixed-effects structure consisted of a significant predictor and its nonlinear interaction with time (i.e., smooth). The random-effects structure contained a random intercept for *Event*, a random smooth for *Subject* by time, and a random smooth for *Word* by time.

The fixed-effects and random-effects structures were fitted separately in a stepwise forward manner. Specifically, we employed the maximum likelihood (ML) estimation approach for model comparisons, through the *compareML* function in the *itsadug* package (v2.4; van Rij et al., 2020). The contribution of a new component was evaluated based on the estimated *p*-value obtained from comparing the models' ML scores, and the component was included only when the *p*-value was smaller than the conventional threshold of 0.05. After the model structure was determined, considering possible autocorrelation, an AR(1) (i.e., first-order autoregression) model for the residuals was constructed by specifying the *rho* parameter and the starting point for each time series to account for the fact that each value in a time series is affected by the preceding value.

In terms of data visualization, averaged proportion looks to different interest areas were plotted using the *VWPre* package (v1.2.4; Porretta et al., 2016). Additionally, as the summary of GAMMs does not indicate the exact shape of regression lines and whether they differ from each other significantly, we used the *itsadug* package (v2.4; van Rij et al., 2020) to plot the nonlinear smooths and significant difference curves.

Table 2. Model summary: Accuracy in Test Phase I

Generalized linear mixed-effects model fit by maximum likelihood, family: binomial				
Fixed effects				
	Estimate	Standard Error	z-value	p-value
Intercept	6.6767	0.9183	7.271	3.58e-13***
Language Background: Native English	−1.8209	0.4752	−3.832	0.000127***
Phonological Contrast: Consonant Contrast	−1.2418	1.0606	−1.171	0.241684
Phonological Contrast: Tone Contrast	−1.3878	0.5414	−2.563	0.010362*
Random effects				
Groups	Name	Variance	Standard Deviation	
Subject	(Intercept)	0.6471	0.8045	
Word Number of observations: 1152, groups: Subject, 64; Word, 15	(Intercept)	2.4020	1.5498	

*Note.* ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

Results

There were neither significant effects of musical sophistication on participants' performance, nor did the Mandarin and English speaker groups differ significantly with respect to musical sophistication. Thus, we do not further discuss this factor in the rest of the paper.

Response accuracy

Test Phase I

There were 1,152 trials (64 participants × 18 trials). Model comparisons suggested that *Language Background* ($p < 0.001$) and *Phonological Contrast* ($p = 0.021$) affected response accuracy, whereas *Homophony* did not ($p = 0.908$); no significant interaction was found among the three factors. The random-effects structure consisted of random intercepts for *Subject* and *Word*, as adding random slopes did not improve the model fit. A summary of the optimal model is presented in Table 2.

For the significant effects, multiple comparisons indicated that Mandarin speakers had significantly higher accuracy than English speakers ($p = 0.0001$; see Figure 2 left panel). Mandarin speakers achieved near-ceiling performance (98.5% mean accuracy), probably because this test phase occurred right after the learning. Further, novel words learned in pairs with both contrasts received significantly more correct responses than those learned in pairs with tone contrasts only ($p = 0.028$; see Figure 2 right panel). In the consonant contrast condition, accuracy was slightly lower than the both contrasts condition and higher than the tone contrast condition, but neither difference reached the significance level.
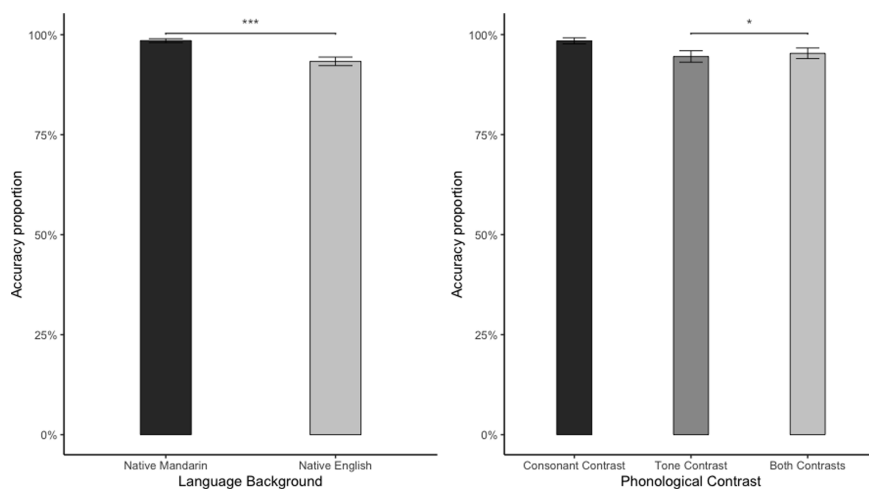
Figure 2. Significant effects on accuracy in Test Phase I.
Note. Left and right panels show the effects of *Language Background* and *Phonological Contrast* on accuracy proportions (y-axis), respectively. Error bars represent the standard error of mean; asterisks indicate a significant difference (*$p < 0.05$, ****$p < 0.001$).

Test Phase II
There were 4,608 trials (64 participants × 72 trials). In the final model, the fixed-effects structure consisted of three significant interactions: a three-way interaction among *Language Background*, *Phonological Contrast*, and *Homophony* ($p < 0.001$); an interaction between *Language Background* and *Break Type* ($p = 0.011$); and an interaction between *Language Background* and *Competition* ($p = 0.040$). The random effects included only random intercepts for *Subject* and *Word*, since none of the random slopes improved model performance. A summary of the best model is shown in Table 3.

Figure 3 displays the effects of these significant interactions. Panel (a) shows the three-way interaction among *Language Background*, *Phonological Contrast*, and *Homophony*. Multiple comparisons suggested that when learning non-homophones in pairs with tone contrasts, English speakers had significantly lower accuracy than Mandarin speakers ($p = 0.002$). Within Mandarin speakers, homophones presented in pairs with consonant contrasts received significantly higher accuracy than those presented in pairs with tone contracts ($p = 0.007$), and for words learned in pairs with consonant contrasts, homophones were significantly more accurate than non-homophones ($p = 0.042$). Surprisingly, Mandarin speakers had lower accuracy than English speakers when learning homophones with tonal contrasts, though this difference did not reach the significance level. Panel (b) illustrates the interaction between *Language Background* and *Break Type*. Interestingly, English speakers had higher accuracy after resting, whereas Mandarin speakers achieved better results after gaming; yet neither of these differences between resting and gaming was significant. However, after playing the game, Mandarin speakers performed significantly better than English speakers ($p = 0.002$). In addition, the interaction between *Language Background* and

Table 3. Model summary: Accuracy in Test Phase II

Generalized linear mixed-effects model fit by maximum likelihood, family: binomial				
Fixed effects				
	Estimate	Standard Error	z-value	p-value
Intercept	1.29960	0.24726	5.256	1.47e-07***
Language Background: Native English	−1.00540	0.28476	−3.531	0.000415***
Phonological Contrast: Consonant Contrast	0.80459	0.28697	2.804	0.005051**
Phonological Contrast: Tone Contrast	−0.22691	0.16929	−1.340	0.180132
Homophony: Non-homophone	−0.08743	0.27257	−0.321	0.748383
Break Type: Rest	−0.43940	0.22060	−1.992	0.046387*
Competition: No-competitor	0.47765	0.09807	4.871	1.11e-06***
Language Background: Native English & Phonological Contrast: Consonant Contrast	−0.01398	0.35905	−0.039	0.968945
Language Background: Native English & Phonological Contrast: Tone Contrast	0.48915	0.23698	2.064	0.039005*
Language Background: Native English & Homophony: Non-homophone	−0.06788	0.24182	−0.281	0.778944
Phonological Contrast: Consonant Contrast & Homophony: Non-homophone	−0.98310	0.37661	−2.610	0.009045**
Phonological Contrast: Tone Contrast & Homophony: Non-homophone	0.19661	0.33375	0.589	0.555792
Language Background: Native English & Break Type: Rest	0.83448	0.31938	2.613	0.008979**
Language Background: Native English & Competition: No-competitor	0.28823	0.13943	2.067	0.038720*
Language Background: Native English & Phonological Contrast: Consonant Contrast & Homophony: Non-homophone	0.50021	0.41674	1.200	0.230023
Language Background: Native English & Phonological Contrast: Tone Contrast & Homophony: Non-homophone	−0.94233	0.34007	−2.771	0.005589**
Random effects				
Groups	Name	Variance	Standard Deviation	
Subject	(Intercept)	0.3260	0.571	
Word	(Intercept)	0.1406	0.375	
Number of observations: 4608, groups: Subject, 64; Word, 28				

Note. ***p < 0.001, **p < 0.01, *p < 0.05.
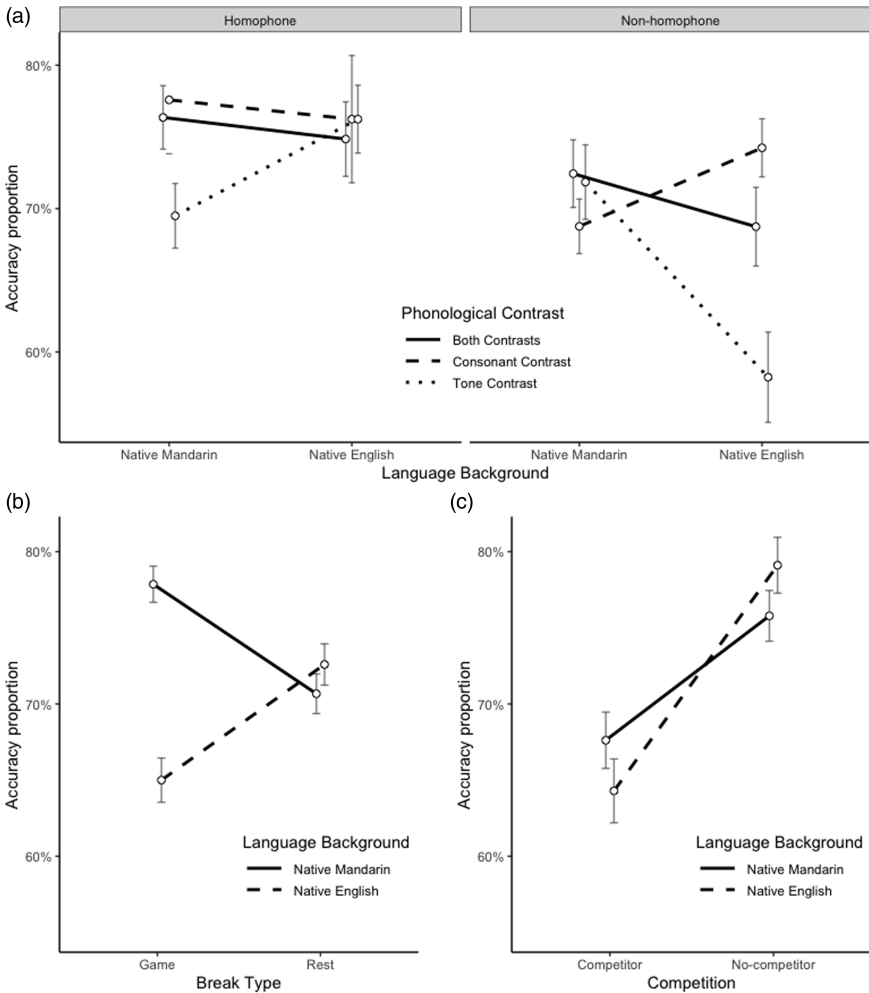
Figure 3. Significant effects on accuracy in Test Phase II.

Note. Panel (a) presents the significant three-way interaction among *Language Background* (x-axis), *Phonological Contrast*, and *Homophony*; (b) presents the interaction between *Language Background* and *Break Type* (x-axis); (c) presents the interaction between *Language Background* and *Competition* (x-axis). Across panels, y-axis indicates accuracy proportions and error bars represent the standard error of mean.

Competition is presented in panel (c). The no-competitor setting had significantly higher accuracy than the competitor setting in both language groups (Mandarin, $p < 0.001$; English, $p < 0.001$), although Mandarin speakers outperformed English speakers in the competitor setting ($p = 0.012$).

Eye movements

Eye movement data in Test Phase II consisted of 4,608 events (64 participants × 72 trials). After removing the 171 events identified with more than 25% trackloss data,

Table 4. Model summary: Logit-transformed proportions of target looks in the competitor setting

| Generalized additive mixed-effects model (GAMM), family: Gaussian				
Parametric coefficients	Estimate	Standard Error	*t*-value	*p*-value
Condition: Native Mandarin. Non-homophone. Both Contrasts	−0.74485	0.16788	−4.437	9.14e−06***
Condition: Native English. Non-homophone. Both Contrasts	−0.03602	0.19591	−0.184	0.8541
Condition: Native Mandarin. Homophone.Both Contrasts	−0.53050	0.22124	−2.398	0.0165*
Condition: Native English. Homophone.Both Contrasts	−0.40599	0.24314	−1.670	0.0950
Condition: Native Mandarin. Non-homophone. Consonant Contrast	−0.42993	0.20236	−2.125	0.0336*
Condition: Native English. Non-homophone. Consonant Contrast	−0.53525	0.21897	−2.444	0.0145*
Condition: Native Mandarin. Homophone.Consonant Contrast	−0.22654	0.28585	−0.793	0.4281
Condition: Native English. Homophone.Consonant Contrast	−0.29623	0.31586	−0.938	0.3483
Condition: Native Mandarin. Non-homophone. Tone Contrast	0.34627	0.23561	1.470	0.1417
Condition: Native English. Non-homophone. Tone Contrast	−0.14093	0.28597	−0.493	0.6221
Condition: Native Mandarin. Homophone.Tone Contrast	−0.38057	0.22152	−1.718	0.0858
Condition: Native English. Homophone.Tone Contrast	−0.13460	0.23690	−0.568	0.5699
Smooth terms	*edf*	Ref.df	*F*-value	*p*-value
s(Time): Condition: Native Mandarin. Non-homophone. Both Contrasts	1.001e+00	1.001	5.099	0.02390*
s(Time): Condition: Native English. Non-homophone. Both Contrasts	2.774e+00	3.520	1.501	0.16559
s(Time): Condition: Native Mandarin. Homophone.Both Contrasts	6.066e+00	7.363	2.352	0.02769*
s(Time): Condition: Native English. Homophone.Both Contrasts	1.001e+00	1.002	3.549	0.05958
s(Time): Condition: Native Mandarin. Non-homophone. Consonant Contrast	2.015e+00	2.483	3.798	0.01613*
s(Time): Condition: Native English. Non-homophone. Consonant Contrast	1.002e+00	1.002	9.666	0.00187**
s(Time): Condition: Native Mandarin. Homophone.Consonant Contrast	3.022e+00	3.898	1.216	0.28014

(*Continued*)

Table 4. (*Continued*)

Smooth terms	edf	Ref.df	F-value	p-value
s(Time): Condition: Native English. Homophone.Consonant Contrast	4.780e+00	6.087	3.395	0.00235**
s(Time): Condition: Native Mandarin. Non-homophone. Tone Contrast	2.815e+00	3.510	1.624	0.17626
s(Time): Condition: Native English. Non-homophone. Tone Contrast	1.002e+00	1.002	3.060	0.08010
s(Time): Condition: Native Mandarin. Homophone.Tone Contrast	3.570e+00	4.545	1.183	0.35034
s(Time): Condition: Native English. Homophone.Tone Contrast	7.445e+00	8.447	3.078	0.00343**
s(Time, Subject)	2.798e+02	565.000	1.096	< 2e-16***
s(Time, Word)	2.001e+02	248.000	5.718	< 2e-16***
s(Event)	5.714e-03	1406.000	0.000	1.00000

Note. edf : effective degrees of freedom; Ref.df: reference degrees of freedom; ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

the remaining dataset contained 326,599 observations. Then, we focused on data with correct responses, which resulted in a total of 232,564 observations.

For GAMMs analysis, we generated empirical logits of the binned target proportion data from 200 ms after the target stimulus onset until 1,700 ms, which served as the dependent variable. The random-effects structure contained a random intercept for *Event*, a random smooth for *Subject* by time, and a random smooth for *Word* by time. For the fixed-effects structure, model comparison suggested a significant three-way interaction among *Language Background*, *Phonological Contrast*, and *Homophony* ($p < 0.001$). In order to inspect the effect of time on this interaction, the three factors were concatenated as a new variable *Condition* with 12 (2 × 3 × 2) levels. A smooth term indicating its nonlinear interaction with time was also significant ($p < 0.001$). Moreover, model selection revealed a significant interaction between *Condition* and *Competition* ($p < 0.001$). This was further evidenced by visualizing the averaged proportions of target looks in different conditions for the competitor and no-competitor settings (see Appendix E in Supplementary Materials). Therefore, we analyzed how *Condition* affected target looks over time in the two settings separately.

The competitor setting contained 106,774 observations. A summary of the optimal model is presented in Table 4. According to the smooth terms section, a significant nonlinear curve over time was found for six levels of *Condition*, as shown by the *edf* (i.e., effective degrees of freedom) value greater than 1 (indicating wiggliness) and the *p*-value lower than 0.05. These nonlinear smooths were plotted in Figure 4. Different patterns were found between the two language groups: overall, target looks increased over time in Mandarin speakers but fluctuated in English speakers.
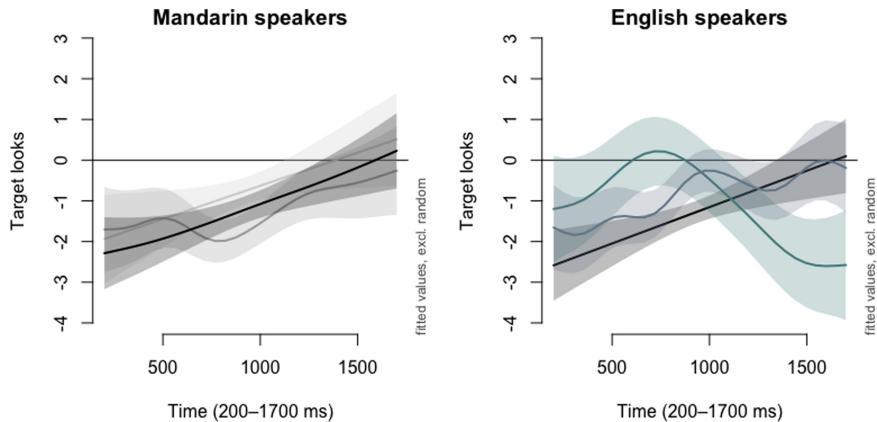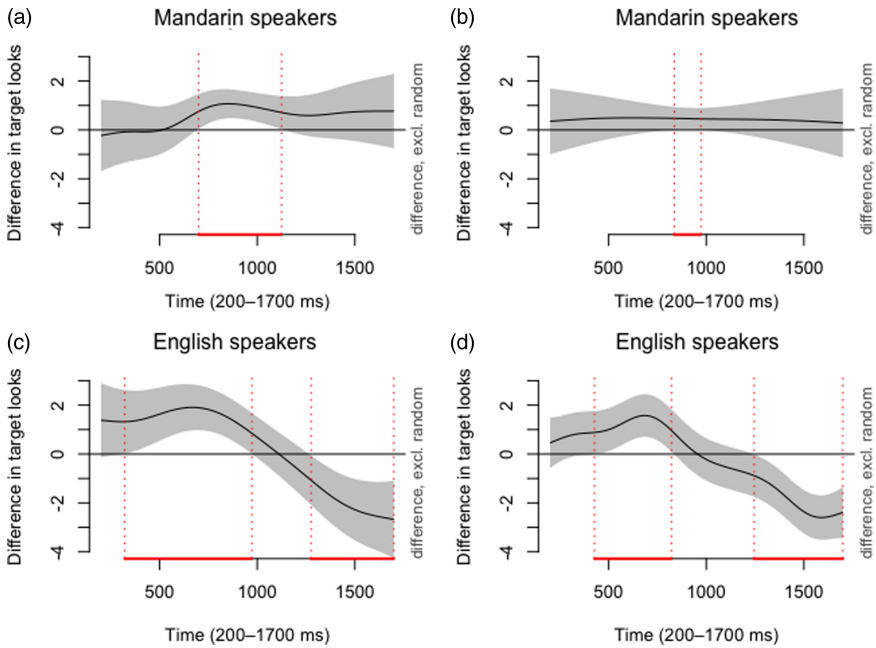
Figure 4. Significant nonlinear smooths over time in the competitor setting.
Note. The x-axis indicates the time after the target stimulus onset, from 200 to 1,700 ms; the y-axis indicates logit-transformed proportions of target looks. Each curve represents a nonlinear smooth, and the shaded area indicates 95% confidence interval. Left panel shows nonlinear smooths for Mandarin speakers. Black curve: non-homophones learned in pairs with consonant contrasts; light gray curve: non-homophones learned in pairs with both contrasts; dark gray curve: homophones learned in pairs with both contrasts. Right panel shows nonlinear smooths for English speakers. Black curve: non-homophones learned in pairs with consonant contrasts; light slate gray curve: homophones learned in pairs with tone contrasts; dark slate gray curve: homophones learned in pairs with consonant contrasts.

Further, significant difference curves are illustrated in Figure 5. Among Mandarin speakers, non-homophones generally received more target looks than homophones when learned in pairs with both contrasts, with a significant difference from 700 to 1,124 ms (see panel a); non-homophones learned in pairs with both contrasts had more target looks continuously than those with consonant contrasts, with a brief period (836–973 ms) showing a significant difference (panel b). Among English speakers, in the consonant contrast environment, homophones received significantly more target looks than non-homophones within 321–973 ms, though the latter had more target looks from 1,276 ms to the end (panel c); homophones learned in pairs with consonant contrasts received significantly more target looks in 427–821 ms, but less from 1,245 ms until the end, compared to those with tone contrasts (panel d). In summary, where differences occurred, Mandarin speakers showed more looks to non-homophones than to homophones learned in the same conditions, whereas it was the other way around for English speakers. Additionally, the data provide some evidence that consonant contrasts facilitate early target identification compared to tone contrasts for English speakers.

The no-competitor setting consisted of 125,790 observations. Table 5 presents a summary of the final model. The smooth terms section suggests two nonlinear curves with significant changes over time for *Condition*, both in English speakers, which were plotted in the left panel of Figure 6; the right panel shows the significant differences between these curves. Specifically, we observed a similar pattern as in the competitor setting: among English speakers, homophones learned in pairs with consonant contrasts received significantly more target looks than those with tone

Figure 5. Significant difference curves in the competitor setting.

Note. The x-axis indicates the time after the target stimulus onset, from 200 to 1,700 ms; the y-axis indicates the difference in logit-transformed proportions of target looks. For each curve, the red interval (with vertical dotted lines) indicates the time window with a significant difference. The shaded area indicates 95% confidence interval. Panel (a) compares non-homophones and homophones learned in pairs with both contrasts by Mandarin speakers, with a significant difference within 700–1,124 ms. Panel (b) compares consonant contrasts and both contrasts in non-homophones learned by Mandarin speakers, with a significant difference in 836–973 ms. Panel (c) compares homophones and non-homophones learned in pairs with consonant contrasts by English speakers, with significant differences within 321–973 ms and 1,276–1,700 ms. Panel (d) compares consonant contrasts and tone contrasts in homophones learned by English speakers, with significant differences within 427–821 ms and 1,245–1,700 ms.

contrasts from 609 to 1,033 ms, but less from 1,579 ms until the end. However, there were no significant nonlinear curves found in Mandarin speakers.

Discussion

This study used the visual world eye-tracking paradigm to explore how adults from native and non-native backgrounds learn Mandarin novel words, which had homophones and non-homophones and varied in types of phonological contrasts: consonant contrasts, tone contrasts, or both. We also examined if brief periods of unoccupied rest would affect participants' learning outcome. In Test Phase I, which occurred immediately after the learning, accuracy results show near-ceiling performance, with Mandarin speakers outperforming English speakers. In Test Phase II, which occurred after a 15-minute break, several significant interactions among the manipulated factors were observed, and Mandarin speakers only outperformed English speakers under certain circumstances. Further, eye movement findings reveal that segmental and tonal information may contribute differently to novel spoken word recognition.

Table 5. Model summary: Logit-transformed proportions of target looks in the no-competitor setting

Generalized additive mixed-effects model (GAMM), family: Gaussian				
Parametric coefficients	Estimate	Standard Error	*t*-value	*p*-value
Condition: Native Mandarin. Non-homophone. Both Contrasts	−0.58187	0.16677	−3.489	0.000485***
Condition: Native English. Non-homophone. Both Contrasts	−0.16856	0.20214	−0.834	0.404355
Condition: Native Mandarin. Homophone.Both Contrasts	−0.11338	0.21018	−0.539	0.589565
Condition: Native English. Homophone.Both Contrasts	−0.08932	0.23592	−0.379	0.704995
Condition: Native Mandarin. Non-homophone. Consonant Contrast	−0.06726	0.19308	−0.348	0.727573
Condition: Native English. Non-homophone. Consonant Contrast	−0.09514	0.21945	−0.434	0.664629
Condition: Native Mandarin. Homophone.Consonant Contrast	−0.10613	0.27419	−0.387	0.698708
Condition: Native English. Homophone.Consonant Contrast	0.01861	0.30581	0.061	0.951487
Condition: Native Mandarin. Non-homophone. Tone Contrast	0.21722	0.22577	0.962	0.335991
Condition: Native English. Non-homophone. Tone Contrast	−0.26618	0.25138	−1.059	0.289650
Condition: Native Mandarin. Homophone.Tone Contrast	−0.28557	0.20360	−1.403	0.160747
Condition: Native English. Homophone.Tone Contrast	−0.24486	0.23006	−1.064	0.287177
Smooth terms	*edf*	Ref.df	*F*-value	*p*-value
s(Time): Condition: Native Mandarin. Non-homophone. Both Contrasts	1.004e+00	1.007	2.768	0.09532
s(Time): Condition: Native English. Non-homophone. Both Contrasts	1.003e+00	1.006	2.696	0.10081
s(Time): Condition: Native Mandarin. Homophone.Both Contrasts	6.476e+00	7.784	1.478	0.14515
s(Time): Condition: Native English. Homophone.Both Contrasts	1.019e+00	1.037	2.554	0.11153
s(Time): Condition: Native Mandarin. Non-homophone. Consonant Contrast	6.365e+00	7.544	2.127	0.05664
s(Time): Condition: Native English. Non-homophone. Consonant Contrast	1.009e+00	1.013	3.770	0.05108
s(Time): Condition: Native Mandarin. Homophone.Consonant Contrast	1.703e+00	2.161	0.854	0.45917

(*Continued*)

Table 5. (*Continued*)

Smooth terms	*edf*	Ref.df	*F*-value	*p*-value
s(Time): Condition: Native English. Homophone.Consonant Contrast	5.142e+00	6.527	3.061	0.00415**
s(Time): Condition: Native Mandarin. Non-homophone. Tone Contrast	3.204e+00	4.087	1.783	0.12908
s(Time): Condition: Native English. Non-homophone. Tone Contrast	1.004e+00	1.005	0.070	0.79557
s(Time): Condition: Native Mandarin. Homophone.Tone Contrast	1.002e+00	1.003	3.536	0.05999
s(Time): Condition: Native English. Homophone.Tone Contrast	1.002e+00	1.003	5.939	0.01471*
s(Time, Subject)	3.100e+02	565.000	1.405	< 2e-16***
s(Time, Word)	2.089e+02	248.000	7.680	< 2e-16***
s(Event)	4.462e-03	1659.000	0.000	1.00000

Note. edf : effective degrees of freedom; Ref.df: reference degrees of freedom; ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.
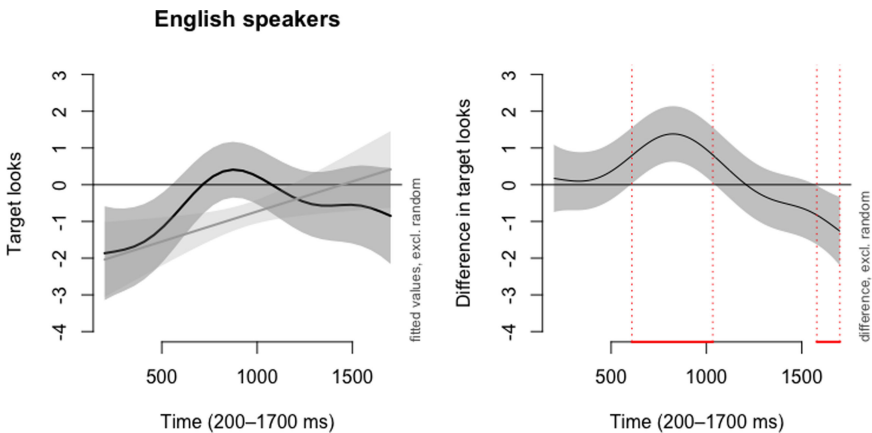

Figure 6. Significant nonlinear smooths and difference curve in the no-competitor setting.
Note. Left panel shows two significant nonlinear smooths, both found in English speakers. The x-axis indicates the time after the target stimulus onset, from 200 to 1,700 ms; the y-axis indicates logit-transformed proportions of target looks. Each curve represents a nonlinear smooth, and the shaded area indicates 95% confidence interval. Black curve: homophones learned in pairs with consonant contrasts; dark gray curve: homophones learned in pairs with tone contrasts. Right panel shows the significant difference curve. It compares consonant contrasts and tone contrasts in homophones learned by English speakers, with significant differences within 609–1,033 ms and 1,579–1,700 ms.

Language background effects

We first asked whether participants' language backgrounds would modulate their novel word learning outcome and predicted that native speakers would outperform non-native speakers in general. In Test Phase I, this hypothesis was borne out, as Mandarin speakers had significantly higher accuracy than English speakers. This

result suggests that Mandarin speakers possessed an advantage due to their native phonological knowledge, which is in line with Poltrock et al. (2018) who found that native speakers performed better than non-native (Mandarin-L1 and French-L1) speakers when learning Cantonese novel words.

In Test Phase II, however, the picture becomes more complicated. Participants' language backgrounds interacted with phonology and homophony to affect the accuracy measure. Although English speakers were generally less accurate than Mandarin speakers, the group difference was only significant when learning non-homophones in pairs with tone contrasts (see further discussion in the following sections). Thus, the language background effects here should be interpreted within the context, where participants were tested on novel words learned a short period of time before. That said, our results provide further evidence for the relative difficulty of tonal learning in adult non-native speakers (in line with Hao, 2018; Laméris & Post, 2022; Pelzl, 2019).

In addition, eye-tracking data showed some patterns unique to English speakers. For example, when comparing homophones learned in different phonological environments, English speakers had more target looks to those learned in the consonant contrast condition during an early stage, but more target looks to those learned in the tone contrast condition later. This result indicates that English speakers were able to use not only the segmental information but also tonal cues in novel word recognition (in line with Wiener et al., 2018, 2021), although the latter seems to be less straightforward and readily accessible. It is possible that the later effect of tonal contrasts is simply due to the fact that tonal cues appear mainly on vowels, i.e., later than consonantal ones, which were localized to syllable onsets (Cutler and Chen, 1997; Ling and Grüter, 2022, but also see Zhao et al., 2011; Zou et al., 2022). Yet more evidence is needed to verify this inference, for example, by comparing the learning of novel words starting with or without an initial consonant. Overall, our findings suggest that beyond accuracy, participants' language backgrounds affected their eye movement behavior, which demonstrates differences in online language processing.

Phonology effects

We also examined whether different types of phonological contrasts entailed in the novel word pairs would predict the outcome and whether this would depend on participants' language backgrounds. In Test Phase I, different from our hypothesis, there was a significant effect of phonological contrasts on accuracy, but no interaction with language backgrounds. Specifically, across language groups, novel words learned in pairs with both contrasts received significantly higher accuracy than those with tone contrasts only. This finding can be explained in part by the fact that the coexistence of segmental and tonal distinctions facilitates novel word recognition and learning simply due to the availability of two cues rather than just one. Further, tonal contrasts not only are difficult to learn for native English speakers (Hao, 2018; Pelzl et al., 2019) but also may pose a challenge for native Mandarin speakers to identify or process (Taft & Chen, 1992; Ye & Connine, 1999). For example, Taft and Chen (1992) found that native Mandarin speakers had significantly longer reaction latency and higher error rates when discriminating

word pairs that differed in tones than those differed in vowels. Similarly, Ye and Connine (1999) reported faster reaction time when listening to Mandarin syllables mismatched on vowels than those mismatched on tones among native speakers. It has been suggested that tones may be a weaker cue than segments in Mandarin lexical access, because tones are primarily realized on vowels and thus are available later than segments in CV syllables (Cutler & Chen, 1997; Ling & Grüter, 2022). Yet, there are studies showing parallel processing of tonal and segmental information in Mandarin word recognition (Malins & Joanisse, 2010, 2012; Zhao et al., 2011; Zou et al., 2022). From another perspective, it can also be hypothesized that tones provide fewer cues than segments since each tone is associated with far more lexical entries than each segment is. Ultimately, it is possible that these inherent differences may account for some of the differences we observed in the learning of novel words that contrast in Mandarin tones and segments (and their combinations).

In Test Phase II, also different from what we expected, phonological contrasts interacted with language backgrounds and homophony to influence accuracy. As stated above, Mandarin speakers outperformed English speakers when learning non-homophones in pairs with tone contrasts; however, surprisingly, they did not outperform English speakers in all other conditions. Further, within Mandarin speakers, homophones learned in pairs with consonant contrasts received more accurate responses than those with tone contrasts. In other words, tonal information complicates Mandarin homophone learning, especially in native speakers. This might be due to a less informative role of tones in Mandarin word retrieval than segments, and a larger interference from their mental lexicon in native speakers than non-native speakers. A further study could assess other tonal language speakers with a different phonology (e.g., Cantonese speakers who have a more complex tonal system) and consider vowel contrasts to compare learners' performance on specific segments and tones.

Homophony effects

We investigated whether the homophone status of a novel word would affect the learning outcome and whether this would be dependent on participants' language backgrounds. We predicted that homophony would interact with language backgrounds. However, this was not the case in Test Phase I, as we did not find any homophony effects on accuracy. By contrast, in Test Phase II, homophony and language backgrounds participated in a three-way interaction. For Mandarin speakers, homophones received more accurate responses than non-homophones when novel words were learned in pairs with consonant contrasts. Though a significant difference only appeared in this particular phonological condition, it is in line with previous research showing a facilitative effect of homophony on processing in Mandarin (Ziegler et al., 2000). Nonetheless, this facilitative effect was not observed in English speakers at all, which is different from the findings of Liu and Wiener (2020, 2022). A possible explanation could be that the English speakers in our study did not have any tonal language experience, whereas those in Liu and Wiener (2020, 2022) had learned Mandarin for at least several months.

Since previous research has mainly used lexical decision tasks to explore different homophone effects in Mandarin and English (e.g., Ziegler et al., 2000), our study is

the first empirical investigation of homophony in Mandarin novel word learning among native and non-native adults. Importantly, our results demonstrate that the phonological information embedded in Mandarin homophones can interact with listeners' language backgrounds to modulate their learning outcome. In particular, English speakers did not profit from homophony when learning Mandarin novel words in terms of the accuracy they achieved, though they did show more early target looks in homophones compared to non-homophones when presented with a competitor differing in a consonant. Whether the absence of a clear facilitative effect is due to the lower proportion of homophones and the inhibitory effect associated with homophones in English (Pexman et al., 2001; Rubenstein et al., 1971) or to these participants being non-native speakers without experience with tonal languages is a matter for future research.

Rest effects

Lastly, we asked if different types of break would predict the learning outcome and whether this would be contingent on participants' language backgrounds. Our hypothesis was that the rest group would perform better than the game group, irrespective of language backgrounds, based on previous research showing the benefits of sleep or rest in various tasks (Brokaw et al., 2016; Dewar et al., 2012; Kurdziel & Spencer, 2016; Qin & Zhang, 2019; Wamsley, 2019). However, this was not corroborated. We found that participants' break type interacted with their language backgrounds, with different patterns in the two language groups. Interestingly, Mandarin speakers benefited from spending the break gaming, whereas English speakers performed better after resting, though the difference between resting and gaming was not significant within either language group. However, Mandarin speakers outperformed English speakers in the gaming condition, but not in the rest condition. Thus, the present study only provides limited statistical evidence for an effect of break type. This might be attributed to the difficulty of the task or the duration of the rest, which raises questions for future research about the role of rest in novel word learning. Nevertheless, to our knowledge, this is the first study addressing how brief periods of rest affect novel word learning in both native and non-native adult speakers.

Limitations

One limitation of our design lies in the number of investigated factors, which have led to the complexity of statistical models and undermined the statistical power, especially considering the relatively limited sample size. We addressed multiple questions related to adult novel word learning in the current study, and most of these questions are complicated by themselves. Future work could examine them more closely by focusing on one or two factors or having a larger sample size. In addition, we acknowledge the exploratory nature of the question about rest. Despite recent research on sleep-dependent consolidation of learning non-native tones (Qin & Zhang, 2019; Qin et al., 2022), the role of a short period of rest was not explored yet, which was the motivation and novelty of our work. Nevertheless, different from our expectation, the rest interacted with linguistic factors even though they may

seem loosely related to each other. Thus, a future study could take another approach to analyze their association further, for example, by first examining the rest effects in Mandarin novel word learning and then extending to learning tones and segments, and lastly comparing native and non-native speakers.

Conclusions

Using visual world eye-tracking, this study is the first comprehensive investigation of Mandarin novel word learning in adult native and non-native speakers. We integrated two fundamental features of Mandarin, phonology and homophony, and assessed the effects of rest following lexical learning. Notably, these linguistic and non-linguistic variables interacted with the learner's language backgrounds to affect their learning outcome. For example, native speakers outperformed non-native speakers only in certain conditions, and the role of short periods of rest remains unclear. The findings of this research provide insights into how various factors play a part in novel word learning.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0142716424000031

Replication package. Replication data and materials for this article can be found at https://osf.io/dt2sk/.

Competing interests. We have no known conflict of interest to disclose.

References

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Proceedings of Interspeech 2005*, 397–400.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2017). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed effects regression models in linguistics* (pp. 49–69). Berlin: Springer.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (Version 6.1.11). http://www.praat.org.

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. https://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. https://doi.org/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*(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

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**(2), 335–357. https://doi.org/10.1017/s0272263114000849

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. https://doi.org/10.1037/a0024049

Chan, R. K., & Leung, J. H. (2020). Why are lexical tones difficult to learn?: Insights from the incidental learning of tone-segment connections. *Studies in Second Language Acquisition*, **42**(1), 33–59. https://doi.org/10.1017/S0272263119000482

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

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

Chen, K.-J., Huang, C.-R., Chang, L.-P., & Hsu, H.-L. (1996). SINICA CORPUS: Design methodology for balanced corpora. In *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation*, 167–176.

Chen, Y., & Liu, H.-M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities*, **35**(1), 10–20. https://doi.org/10.1016/j.ridd.2013.10.010

Constantin, A.-E., & Patil, I. (2021). ggsignif: R package for displaying significance brackets for 'ggplot2'. *PsyArxiv.* https://doi.org/10.31234/osf.io/7awm6

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. https://doi.org/10.1121/1.4714355

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Psychophysics*, **59**(2), 165–179. https://doi.org/10.3758/BF03211886

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**(3), 453–459. https://doi.org/10.3758/bf03193787

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*, **12**(3), 231–251. https://doi.org/10.1080/15475441.2015.1127163

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. https://doi.org/10.1177/0956797612441220

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, **11**(2), 114–125. https://doi.org/10.1038/nrn2762

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York: Oxford University Press.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, **425**(6958), 614–616. https://doi.org/10.1038/nature01951

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). New York: Springer. https://doi.org/10.1007/978-1-4612-2852-3_3

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, **40**(2), 269–279. https://doi.org/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. https://doi.org/10.1177/0023830917717759

Hao, Y.-C. (2023). Phonolexical processing of Mandarin segments and tones by English speakers at different Mandarin proficiency levels [Advance online publication]. *Second Language Research.* https://doi.org/10.1177/02676583231167790

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. https://doi.org/10.1037/a0029122

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. https://doi.org/10.3758/s13428-015-0647-3

Huang, C.-R., Chen, K.-J., Gao, Z.-M., Chen, F.-Y., & Shen, Z.-C. (1998). Mandarin Chinese Word Frequency Dictionary. *Technical Report 98-01*. Nankang: Academia Sinica.

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**(1), 53. https://doi.org/10.1186/1471-2202-9-53

Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., & Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature Neuroscience*, **10**(9), 1206–1213. https://doi.org/10.1038/nn1959

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laméris, T. J., & Post, B. (2022). The combined effects of L1-specific and extralinguistic factors on individual performance in a tone categorization and word identification task by English-L1 and Mandarin-L1 speakers. *Second Language Research*, 026765832210900. https://doi.org/10.1177/02676583221090068

Lenth, R. V. (2022). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.7.2]. https://CRAN.R-project.org/package=emmeans

Li, C., Wang, M., Davis, J. A., & Guan, C. Q. (2018). The role of segmental and tonal information in visual word recognition with learners of Chinese. *Journal of Research in Reading*, **42**(2), 213–238. https://doi.org/10.1111/1467-9817.12137

Ling, W., & Grüter, T. (2022). From sounds to words: The relation between phonological and lexical processing of tone in L2 Mandarin. *Second Language Research*, **38**(2), 289–313. https://doi.org/10.1177/0267658320941546

Liu, J., & Wiener, S. (2020). Homophones facilitate lexical development in a second language. *System*, **91**, 102249. https://doi.org/10.1016/j.system.2020.102249

Liu, J., & Wiener, S. (2022). Effects of phonological and talker familiarity on second language lexical development. *The Mental Lexicon*, **17**(1), 132–153. https://doi.org/10.1075/ml.20024.liu

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, **62**(4), 407–420. https://doi.org/10.1016/j.jml.2010.02.004

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, **50**(8), 2032–2043. https://doi.org/10.1016/j.neuropsychologia.2012.05.002

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. https://doi.org/10.1016/j.jml.2017.01.001

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, **9**(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Neergaard, K. D., Xu, H., German, J. S., & Huang, C.-R. (2021). Database of word-level statistics for Mandarin Chinese (DoWLS-MAN). *Behavior Research Methods*, **54**(2), 987–1009. https://doi.org/10.3758/s13428-021-01620-7

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. https://doi.org/10.1016/j.specom.2014.05.001

Pelzl, E. (2019). What makes second language perception of Mandarin tones hard? Chinese as a Second Language. *The Journal of the Chinese Language Teachers Association, USA*, **54**(1), 51–78. https://doi.org/10.1075/csl.18009.pel

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, **41**, 59–86. https://doi.org/10.1017/S0272263117000444

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. https://doi.org/10.1037/0278-7393.27.1.139

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. https://doi.org/10.3389/fpsyg.2018.01211

Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2016). *VWPre: Tools for preprocessing visual world data* [Version 1.2.4]. https://CRAN.R-project.org/package=VWPre

Qin, Z., Jin, R., & Zhang, C. (2022). The effects of training variability and pitch aptitude on the overnight consolidation of lexical tones. *Journal of Speech, Language, and Hearing Research*, **65**(9), 3377–3391. https://doi.org/10.1044/2022_JSLHR-22-00058

Qin, Z., & Mok, P. P.-K. (2011). Perception of Cantonese tones by Mandarin, English and French speakers. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, 1654–1657.

Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLOS ONE*, **14**(12), 1–23. https://doi.org/10.1371/journal.pone.0221498

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://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. https://doi.org/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**(6), 645–657. https://doi.org/10.1016/s0022-5371(71)80071-3

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, **5**(2), 1318. https://doi.org/10.3389/fpsyg.2014.01318

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R* [ISBN 978-0-387-75968-5]. Springer. http://lmdvr.r-forge.r-project.org

Schaefer, V., & Darcy, I. (2020). Applying a newly learned second language dimension to the unknown: The influence of second language Mandarin tones on the naïve perception of Thai tones. *Psychology of Language and Communication*, **24**(1), 90–123. https://doi.org/10.2478/plc-2020-0007

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, **58**(2), 131–151. https://doi.org/10.1177/0023830914522956

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. https://doi.org/10.1121/1.4971765

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. https://doi.org/10.3389/fpsyg.2016.00667

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, **53**(2), 273–293. https://doi.org/10.1177/0023830909357156

SR Research. (2015). *SR research experiment builder 1.10.1630*. Mississauga, Ontario, Canada: SR Research Ltd.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, **437**(7063), 1272–1278. https://doi.org/10.1038/nature04286

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, **32**(4), 827–853. https://doi.org/10.1017/S0305000905007099

Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen & O. J. Tzeng (Eds.), *Language processing in Chinese* (pp. 151–172). Oxford: North-Holland. https://doi.org/10.1016/S0166-4115(08)61891-9

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. https://doi.org/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. H., & van Rijn, H. (2020). itsadug: Interpreting time series and autocorrelated data using gamms [R package version 2.4].

Wamsley, E. J. (2019). Memory consolidation during waking rest. *Trends in Cognitive Sciences*, **23**(3), 171–173. https://doi.org/10.1016/j.tics.2018.12.007

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. https://doi.org/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. https://doi.org/10.1016/j.wocn.2007.06.004

Weighall, A., Henderson, L., Barr, D., Cairney, S., & Gaskell, M. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, **167**, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Wen, W. (1980). Cong yingwen de tongxingci lai kan hanyu pinyin wenzi de tongyinci [A study of Chinese homophones from the view of English homographs]. *Yuwen Xiandaihua [Modernizing Our Language]*, **2**, 120–124.

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*. New York: Springer-Verlag. https://ggplot2.tidyverse.org

Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech*, **61**(4), 632–656. https://doi.org/10.1177/0023830918761762

Wiener, S., Ito, K., & Speer, S. R. (2021). Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinations: An eye-tracking study. *Studies in Second Language Acquisition*, **43**(1), 155–180. https://doi.org/10.1017/S0272263120000418

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, **28**(4), 565–585. https://doi.org/10.1017/S0142716407070312

Wood, S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC Press.

Wright, J., & Baese-Berk, M. (2022). The impact of phonotactic features on novel tone discrimination. *Journal of Phonetics*, **93**, 101151. https://doi.org/10.1016/j.wocn.2022.101151

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. https://doi.org/10.1016/j.wocn.2014.06.005

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, **14**(5/6), 609–630. https://doi.org/10.1080/016909699386202

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, **27** (1), 153–201. https://doi.org/10.1017/S0952675710000060

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, **49**(7), 1761–1770. https://doi.org/10.1016/j.neuropsychologia.2011.02.054

Zhu, M., Chen, X., & Yang, Y. (2021). The effects of native prosodic system and segmental context on Cantonese tone perception by Mandarin and Japanese listeners. *The Journal of the Acoustical Society of America*, **149**(6), 4214–4227. https://doi.org/10.1121/10.0005274

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. https://doi.org/10.1111/1467-9280.00247

Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience*, **9**(6), 758–775. https://doi.org/10.1162/jocn.1997.9.6.758

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. https://doi.org/10.1017/S1366728916000791

Zou, T., Caspers, J., & Chen, Y. (2022). Perception of different tone contrasts at sub-lexical and lexical levels by Dutch learners of Mandarin Chinese. *Frontiers in Psychology*, **13**. https://doi.org/10.3389/fpsyg.2022.891756