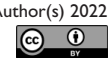
The combined effects of L1-specific and extralinguistic factors on individual performance in a tone categorization and word identification task by English-L1 and Mandarin-L1 speakers

Laméris, T.J.; Post, B.

Article

second language research

Second Language Research
2023, Vol. 39(3) 833–871



DOI: 10.1177/0267658321090068
journals.sagepub.com/home/slr


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

Tim Joris Laméris and **Brechtje Post**
University of Cambridge, UK

Abstract

Adult second language learners often show considerable individual variability in the ease with which lexical tones are learned. It is known that factors pertaining to a learner's first language (L1; such as L1 tonal status or L1 tone type) as well as extralinguistic factors (such as musical experience and working memory) modulate tone learning facility. However, how such L1-specific and extralinguistic factors affect performance together in dynamic ways is less well understood. Therefore, to unpack the potential interactions between these factors for individual learners, we assessed the combined effects of L1 tonal status, L1 tone type, and musical experience and working memory on second language (L2) tone perception and word learning in a tonal pseudolanguage by English-L1 and Mandarin-L1 adult learners, by using a pre-lexical tone categorization task and a lexical word identification task. We found that L2 tone perception and word learning were primarily facilitated by extralinguistic factors, but that the degree to which learners rely on these factors is modulated by their L1 tonal status, as for instance musical experience facilitated perception and word learning for English, but not for Mandarin participants. We also found clear effects of L1 tone type, as Mandarin participants tended to struggle with categorizing and lexically processing level tone contrasts, which do not occur in Mandarin.

Keywords

extralinguistic factors, individual variability, lexical tone, L2 speech, musical experience, perception, word learning, working memory

Corresponding author:
Tim Joris Laméris, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK
Email: tjl44@cam.ac.uk

I Introduction

In tone languages, fundamental frequency (f_0) acts as a primary acoustic cue to change a word's core lexical meaning (Yip, 2002). For adult second language (L2) learners, lexical tones are thought to be relatively difficult to master. In particular, while they may overcome difficulties in processing tones devoid of lexical meaning in tone *perception* (X. Wang, 2013), it appears that linking tones to a lexical item in *word learning* presents considerably more persistent difficulty (Pelzl et al., 2019, 2020). Yet, as with all aspects of speech, some learners appear to perceive tones and learn tone words more easily than others do, reflecting the large degree of individual variability in L2 learners' speech learning *facility*, i.e. the ease with which non-native sounds are learned in the early stages (Bowles et al., 2016, pp. 774–775; Kachlicka et al., 2019). To better understand what accounts for this individual variability, this article examines how factors pertaining to a learner's first language (L1), as well as extralinguistic factors, jointly affect L2 tone perception and word learning facility.

We will use the term 'L1-specific factors' to refer to linguistic factors pertaining to a learner's L1, and zoom in on *L1 tonal status* (i.e. does the L1 use tones for lexical purposes?) and *L1 tone type* (i.e. what types of f_0-based units, either tonal or intonational, exist in the L1?). In addition, we will use the term 'extralinguistic factors' to refer to individual factors not related to the L1, and focus in this article on musical experience and working memory. As we will review in Section II, all these factors are known to modulate L2 tone perception and word learning facility. However with a few notable exceptions[1] (Chan and Leung, 2020; D. Chang et al., 2016; S. Chen et al., 2020; Cooper and Wang, 2012), most previous studies either only assess the effects of L1-specific factors, controlling for or not measuring the effect of extralinguistic factors (Braun et al., 2014; J. Chen et al., 2020; So and Best, 2010), or they assess extralinguistic factors, but in participants of the same L1 (Bowles et al., 2016; Wong et al., 2020). Therefore, instead of looking at these factors separately, we examine the combined effects of L1-specific and extralinguistic factors to try to provide a more complete and accurate account of individual variability in L2 tone learning. More specifically, this article investigates how L1 tonal status, L1 tone type, musical experience and working memory – factors that have not been investigated simultaneously in previous studies – work together to modulate performance in a tone categorization task (representing tone *perception*) and in a pseudolanguage word identification task (representing tone *word learning*) by a group of tonal (Mandarin-L1) and non-tonal (English-L1) learners.

II Background

1 L1-specific factors in L2 tone perception

There is ample evidence that L1 tonal status modulates individual performance in tone perception. In comparison to non-tonal peers, L1 speakers of a tonal language (henceforth: 'tonal L1ers') tend to process tones predominantly in the left brain hemisphere (Klein et al., 2001; Y. Wang et al., 2004), perceive L2 tones in a categorical rather than in a psychoacoustic way (Hallé et al., 2004), and tend to be better at identifying tones spoken by multiple speakers (Y. S. Chang et al., 2017). Some studies further show that

the stronger the lexical role of f_0 in the L1, the better the sensitivity to pitch in an L2 (Schaefer and Darcy, 2014), and that not only L1 but also L2 knowledge of a tonal language can facilitate non-native pitch perception (Wiener and Goss, 2019). While this suggests that tonal L1ers perceive tones *differently* than their non-tonal peers, by no means do they always perform *better*, as evidenced by findings of tone identification and discrimination tasks in which tonal L1ers do not outperform their non-tonal peers (Cooper and Wang, 2012; Francis et al., 2008; Gandour and Harshman, 1978; So and Best, 2010; X. Wang, 2013). Note however, that there are findings that do suggest a comparative advantage in tone perception for tonal L1ers (Chan and Leung, 2020; Peng et al., 2010; Wayland and Guion, 2004).

One reason why L1 tonal status alone may not explain individual differences in L2 tone perception is because the factor of L1 tone type needs to be considered. Simply put, rather than L2 tones overall, it is often specific L2 tones that may be easy or difficult to perceive, depending on the tone types in a learner's L1. Note that in this article, we will use the term L1 tone type as an overarching expression to describe specific f0-based units (which can be either lexical or intonational tones) occurring in the L1 in terms of 1) phonological-categorical and 2) phonetic-acoustic properties, following the distinction proposed by K. Yu et al. (2017).

Previous studies have suggested that L1 tone type (in phonological-categorical terms) affects L2 tone perception because listeners may assimilate non-native tones to f_0-based categories in the L1 (S. Chen et al., 2020; Hao, 2012; So and Best, 2010). This notion of categorical assimilation is rooted in models of L2 speech learning such as the Perceptual Assimilation Model (Best, 1995; Best and Tyler, 2007) that propose that the ease with which non-native sounds are learned depends on the relative similarity between L1 and L2 sounds. For example, L1 speakers of Mandarin, which only has one high-level tone, appear to struggle with discriminating Cantonese mid-level and low-level tones (Qin and Jongman, 2016; Zhu et al., 2021). It has been suggested that this is because Mandarin listeners tend to assimilate Cantonese level tones to the single Mandarin level tone, making them therefore relatively difficult to perceive accurately (Qin and Jongman, 2016, p. 334; Zhu et al., 2021, p. 4224).

Crucially, non-tonal listeners may be less affected by categorical assimilation because they simply do not have competing lexical tone categories in their L1. Although they may assimilate L2 tones to intonational categories, effects of such assimilation on L2 tone perception may be relatively weak (Best, 2019, p. 5; Reid et al., 2015; So and Best, 2010, 2014), arguably because intonational categories have a 'weaker (less categorical) mental representation' than lexical tone categories (Francis et al., 2008, p. 269). As a result, even though they may fail to form abstract L2 tone categories (Chan and Leung, 2020, p. 10), non-tonal listeners may in some instances perceive L2 tones more accurately than tonal listeners by processing them in a psychoacoustic manner (A. Chen et al., 2018; Peng et al., 2010; X. Wang, 2013; K. Yu et al., 2019).

An alternative account describing the effect of L1 tone type on L2 tone perception focuses on phonetic-acoustic rather than phonological-categorical properties. For instance, speakers of Mandarin appear to pay relatively more attention to differences in f_0 contour and direction, whereas English speakers may pay relatively more attention to f_0 height when processing pitch, which could potentially explain the difficulty for

Mandarin speakers to perceive level tone contrasts in an L2 (Francis et al., 2008; Gandour and Harshman, 1978; Qin and Jongman, 2016).

Finally, we note that attentional differences between listeners of different L1s to secondary cues of lexical tones may also modulate L2 tone perception (S. Chen et al., 2017). For instance, laryngeal phonation (creaky voice) facilitates perception of low-register tones in Cantonese-L1 listeners (K. M. Yu and Lam, 2014) and of low-dipping tones in Mandarin-L1 listeners (Yang, 2015). In this study, we will zoom in on f_0 as the primary acoustic cue to lexical tone and only manipulate f_0 between the stimuli, but we will consider the possible effect of the absence of other acoustic cues on participants' tone perception in the discussion.

2 L1-specific factors in tone word learning

Whereas accounting for individual differences in L2 tone *perception* based on L1 tonal status alone remains relatively complex, particularly because of the effect of L1 tone type on the perception of specific L2 tones, it appears that individual differences in L2 tone *word learning* can be more easily accounted for by L1 tonal status.

For instance, Pelzl et al. (2019) report that English-L1 advanced L2 learners of Mandarin can accurately perceive pitch in a pre-lexical tone categorization task, but may not all be able to 'repurpose it as a lexical cue' (p. 80) in lexical tasks. In an eye-tracking study, Ling and Grüter (2020) similarly found that English-L1 intermediate learners of Mandarin had 'considerably more difficulty in using tone alone to distinguish between words' (p. 19).

It is crucial to note that these studies involved Mandarin participants listening to their own L1, thereby perhaps naturally yielding an advantage of L1 tonal status in comparison to non-tonal participants. However, evidence for a facilitative effect of L1 tonal status in L2 tone word learning is also found in studies in which tonal L1ers were exposed to a different tone language. For instance, Poltrock et al. (2018) showed that Mandarin participants outperformed French listeners in recalling Cantonese pseudowords that contrasted in tone. Chan and Leung (2020) investigated the effects of L1 tonal status on the incidental 'phonological learning', which was defined as an intermediate step between tone perception and tone word learning (p. 4). They show that Cantonese participants outperformed English participants in the phonological learning of Thai tones, and suggest that Cantonese L1 tonal status facilitated the formation of syllable-level tone categories required for utilizing tones at the word level.

It thus appears that L1 tonal status on its own may facilitate L2 tone word learning, given tonal L1ers' familiarity with the use of pitch to indicate lexical meaning (Cooper and Wang, 2012, p. 4765). However, to the best of our knowledge, there are no studies that examine whether in addition to L1 tonal status, L1 tone type also modulates L2 tone word learning in a similar way that it is known to modulate L2 tone perception. To address this gap in the literature, we first ask:

- Research question 1: How do Mandarin participants' L1 tonal status and L1 tone types affect individual performance in a tone categorization task and a word identification task in a pseudolanguage with a rising, a falling, a mid-level, and a low-level tone, and how does this compare to performance by English participants?

3 Extralinguistic factors: Musical experience and working memory

There has been an increasing interest in recent years to explain individual variability in L2 tone learning by not only looking at learners' L1-specific, but also extralinguistic factors. Here, we focus on two of these factors, musical experience and working memory, and review previous studies that have investigated their role in L2 tone perception and word learning.

Musical experience is one of the most investigated extralinguistic factors in the L2 tone perception and word learning literature, possibly due to the shared cognitive processing of pitch in music and language (Perrachione et al., 2013; Sadakata et al., 2020). For tone perception, studies with Mandarin speakers have revealed improved pitch sensitivity and tone discrimination abilities in trained musicians compared to non-musicians (Tang et al., 2016; H. Wu et al., 2015). In a large-scale study involving over 400 Cantonese native speakers, years of musical training was found to be the strongest predictor of performance in a tone discrimination task (Wong et al., 2020). However, some studies show no clear effect of musical experience on tone perception (Chan and Leung, 2020), and it has been suggested that a facilitative effect of musical experience on L2 tone perception may be task-dependent (D. Chang et al., 2016).

Studies on L2 tone word learning generally find a facilitative effect of musical experience. In one of the earliest studies on the subject, Wong and Perrachione (2007) report that English learners with musical experience performed better than non-musicians, both in pre-lexical perception of tones on meaningless syllables and in the learning of tonal words in a pseudolanguage. Bowles et al. (2016) found similar facilitative effects of musical experience in a large study of Mandarin word learning by 160 English-L1 participants.

As this article focuses on the combined effects of L1-specific and extralinguistic factors, a key question is whether L1-specific factors such as L1 tonal status interact with extralinguistic factors like musical experience. Studies that have investigated this suggest that this is indeed the case. For instance, S. Chen et al. (2020) showed that English-L1 musicians had a stronger categorical perception of tones than non-musicians, whereas no such difference was found between Mandarin-L1 musicians and non-musicians. This suggests that the facilitative effect of musical experience on L2 tone perception may be weaker for tonal L1ers. Such an interaction between L1 tonal status and musical experience was also found in L2 tone word learning by Cooper and Wang (2012), who showed that musical experience only benefited English, but not Thai participants in Cantonese tone word learning. The authors suggest that English participants may have drawn on their pitch acuity gained through musical practice 'to enhance their ability to utilize linguistic pitch in a higher-level linguistic context' (p. 4765). By contrast, the Thai participants may not have needed to additionally draw on skills gained through musical experience because they already benefited from their L1 tonal status in tone word learning, making musical experience less relevant. This suggests that there is a dynamic interplay between L1-specific and extralinguistic factors in tone word learning, and highlights the importance of accounting for both of these types of factors in investigating L2 tone learning facility.

As a second extralinguistic factor, we assessed the effect of individual learners' working memory (WM) on performance in our tone categorization and tone word

identification tasks. We deemed it necessary to include a measure of WM because our word identification task replicates vocabulary learning, for which WM has been found to be facilitative (Baddeley, 2003; Kormos and Sáfár, 2008). In addition, we want to further investigate the role of WM in facilitating pre-lexical and lexical processing of pitch following conflicting findings in the literature. Findings from previous studies suggest that WM may not facilitate pre-lexical pitch processing, either in language or in music, although this may depend on how cognitively demanding the task is (Bidelman et al., 2013; Hutka et al., 2015). As for lexical pitch processing, studies with English-L1 participants suggest that WM facilitates word-level processing of Japanese pitch (Goss, 2020), and moderately facilitates Mandarin tone word learning (Bowles et al., 2016). However, findings from Chinese-L1 and Korean-L1 advanced learners of Japanese lexical pitch (Goss and Tamaoka, 2019) and English-L1 beginners learning tonal pseudolanguage words (Perrachione et al., 2011) revealed no such facilitative effect. Given this relatively unclear link between WM and pre-lexical and lexical pitch processing, we therefore re-assess whether WM facilitates performance in tone perception and tone word learning in English and Mandarin participants.

Finally, since our study measured both tone perception (in a tone categorization task) and tone word learning performance (in a word identification task), we will also investigate whether performance in one task predicts performance in the other. Indeed, studies that investigated the link between pre-lexical and lexical pitch processing suggest that L2 tone perception performance (i.e. individual pitch perception ability) may in fact be one of the strongest facilitators of L2 tone word learning in English speakers (Bowles et al., 2016; Ling and Grüter, 2020; Perrachione et al., 2011; Wong and Perrachione, 2007, p. 565). However, evidence from the cross-linguistic study by Cooper and Wang (2012) suggests that L1 tonal status may attenuate the facilitative effect of individual pitch perception ability on tone word learning, as English-L1 participants did but Thai-L1 participants did not benefit from pitch perception ability in Cantonese tone word learning. This leaves it relatively unclear what extralinguistic factors do facilitate tone word learning in tonal L1 participants given that, based on Cooper and Wang (2012), neither musical experience nor pitch perception ability appear to strongly do so.

In sum, the literature to date has mainly investigated how individual variability in L2 tone perception and word learning is modulated by learners' L1-specific or extralinguistic factors, but only a handful of studies have examined the combined effect of such factors. Yet, findings that suggest that musical experience facilitates L2 tone word learning in English but not in Thai listeners (Cooper and Wang, 2012) highlight that simultaneously accounting for an array of L1-specific and extralinguistic factors may provide a more refined view of how individual factors modulate L2 tone learning facility. Therefore, our study combines L1-specific and extralinguistic factors, which were only partially addressed in previous studies, to better understand the relative weighting of and interactions between these factors on performance in L2 tone perception and in word learning. We therefore ask as our second research question:

- Research question 2: How do Mandarin participants' L1 tonal status and L1 tone types interact with musical experience and working memory to determine performance in our tone categorization and word identification tasks, and how does this compare to English participants?

III Methods

We assessed the combined effects of L1 tonal status, L1 tone type, musical experience and working memory (WM) in tone perception and word learning by means of two behavioral tasks: A tone categorization task and a tone word identification task.

Table 1. Participant demographics.

	English (n = 21)		Mandarin (n = 20)	
Age (years)	20.98 (1.56)		22.63 (3.32)	
WM Score (%)	57.90 (23.19)		72.09 (23.54)	
Musical experience (years)				
MU = musicians, NM = non-musicians	MU (n = 11)	NM (n = 10)	MU (n = 10)	NM (n = 10)
	13.32 (2.84)	0.90 (1.66)	14.84 (5.09)	0.98 (0.97)

Notes. Values are means with standard deviations in brackets. Equivalence tests (Lakens et al., 2018); Cohen's d set at 0.5 revealed no significant differences in the measures between the two groups. Test results: Age: $t(39) = -0.453$, $p = .673$. WM: $t(39) = -0.344$, $p = .634$. Musical experience (musicians): $t(19) = 0.288$, $p = .388$. Musical experience (non-musicians): $t(18) = 0.986$, $p = .168$.

1 Participants

The study was approved by the ethics board of the University of Cambridge. 21 native speakers of English (11 female; mean age: 20.98) and 20 native speakers of Mandarin Chinese (10 female; mean age: 22.63) participated in this study. Participants were all recruited at the University of Cambridge, participated voluntarily and were paid for their participation. Within each group, half of the participants were musicians, which we defined as participants who were actively practicing music and who had more than 6 years of formal musical training (Cooper and Wang, 2012; Wong and Perrachione, 2007). An overview of the participants is given in Table 1 and a detailed description is provided in Appendices 1–2.

None of the participants claimed to be simultaneous bilinguals (i.e. being fully proficient in two languages acquired since birth), but many had knowledge of a second language and some had some exposure to a heritage language. Some speakers in the Mandarin group reported to have some knowledge of another Chinese language or dialect (including Wu and Cantonese).[2] None of the English participants had knowledge of a pitch accent or tone language.

Participants' working memory was estimated by a backwards digit span task, as outlined later in this section. The measure of musical experience was computed as the number of years of playing a musical instrument including formal instruction.

Table 2. Pseudolanguage words.

	Tone 1 (rising 15)	Tone 2 (falling 51)	Tone 3 (mid-level 22)	Tone 4 (low-level 11)
Segment 1	/nɔn15/	/nɔn51/	/nɔn22/	/nɔn11/
Meaning	*television*	*book*	*cat*	*fork*
Segment 2	/lɔn15/	/lɔn51/	/lɔn22/	/lɔn11/
Meaning	*chair*	*leg*	*apple*	*church*
Segment 3	/jɑɹ15/	/jɑɹ51/	/jɑɹ22/	/jɑɹ11/
Meaning	*mountain*	*kite*	*leaf*	*shirt*
Segment 4	/juɹ15/	/juɹ51/	/juɹ22/	/juɹ11/
Meaning	*door*	*guitar*	*car*	*hammer*

2 Stimuli

Two sets of audio stimuli were used: a set of vowels (/i/ /a/ and /ɛ/) for the tone categorization task and a set of pseudolanguage words (/nɔn/, /lɔn/, /jɑɹ/ and /juɹ/; see Table 2) for the word identification task. These stimuli carried either a rising, a falling, a mid-level, or a low-level tone, resulting in $3 \times 4 = 12$ tone stimuli and $4 \times 4 = 16$ word stimuli (sound files are in supplemental material 3). The four tones were chosen explicitly to assess the effect of L1 tone type on Mandarin participants, with the rising and falling being exemplars of the rising and falling tones in Mandarin, but mid-level and low-level tones both being similar to the single Mandarin high-level tone in terms of pitch contour.

To avoid bias that may arise from listening to stimuli produced by a speaker of one's native language (Braun and Johnson, 2011), stimuli were recorded by two native speakers of Italian, who were trained singers. To ensure that participants would not be influenced by voice familiarity across tasks and to help abstract away from the f_0 traces to tone categories, the female voice was used in the tone categorization task and the male voice in the word identification task.

Stimuli were recorded in a sound-attenuated booth at a sampling frequency of 48 KHz. The speakers were instructed to produce stimuli with a flat tone at a comfortable pitch level. The f_0 contour of this naturally produced flat tone was taken as a baseline tone (the mid-level tone). The speakers were also instructed to naturally produce stimuli with a rising, falling, and low-level tone. Based on the f_0 onset and end values of these natural productions, the mid-level tone stimuli were then resynthesized using Pitch-Synchronous Overlap and Add (PSOLA) in Praat (Boersma and Weenink, 2019) to create stimuli for the other tones. This ensured that tone minimal quadruplets only differed in f_0 and not in other acoustic cues. Both the male and the female tones had the same relative tone values in terms of Chao numerals (Chao, 1968) and the vowel stimuli in the tone categorization task and the pseudolanguage word stimuli in the word identification task were therefore deemed to belong to the same four tone categories: namely 15 (rise); 51 (fall); 22 (mid-level); and 11 (low-level). For visualization, the f_0 and Chao-normalized contours of the tones are shown in Figure 1.

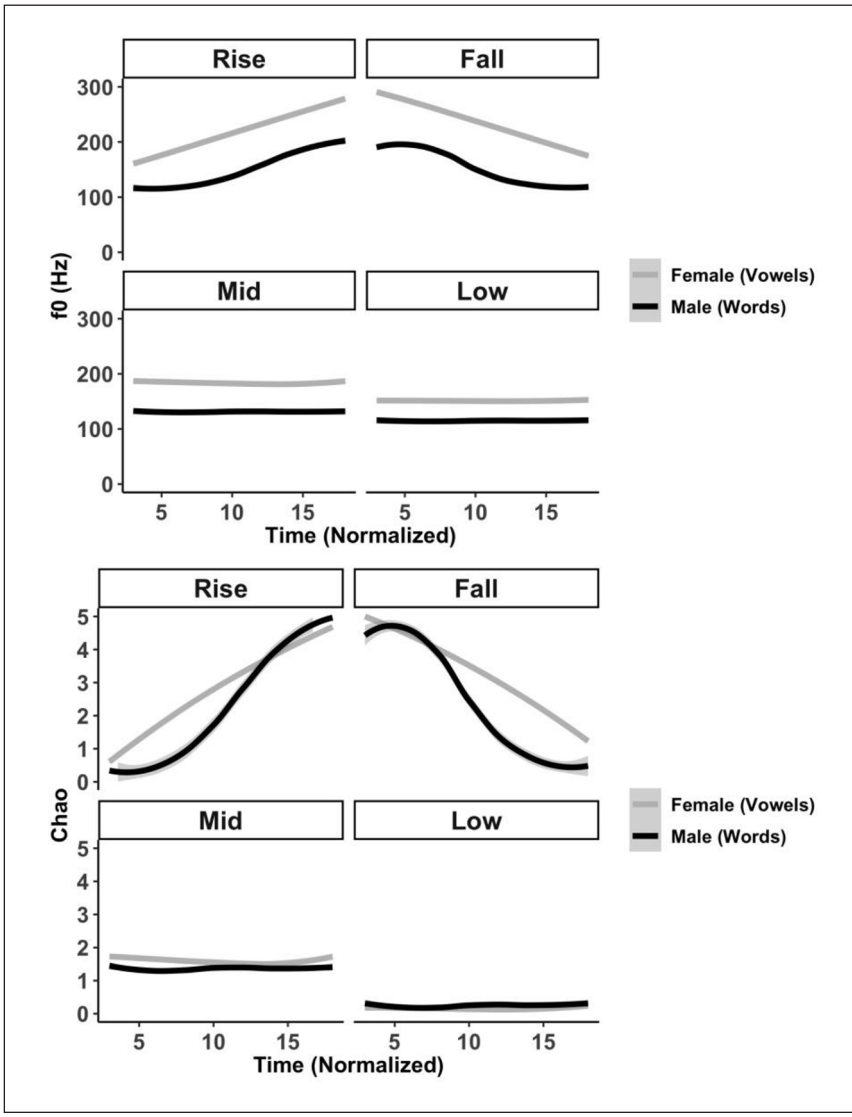


Figure 1. f_0 and Chao numeral curves for the four tones. Ribbons, where applicable, indicate a 95% confidence interval.

After resynthesis, the average intensity of stimuli was set to 70 dB (using the 'scale intensity' command in Praat). Five trained phoneticians deemed the synthesized stimuli to sound as natural as the original mid-level stimuli.

In the tone categorization task, each tone was represented by an arrow (Figure 2). In the word identification task, each pseudolanguage word was linked to an image to establish a sound-meaning connection (Figure 3). The images were gathered from a database

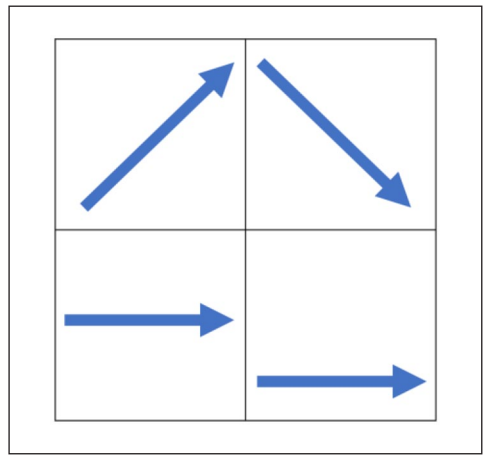


Figure 2. Visual stimuli in tone categorization task.

by Rossion and Pourtois (2004) and represent 16 high-frequency nouns (Battig and Montague, 1969; Van Overschelde et al., 2004). Care was taken to select words that were semantically unrelated to each other to facilitate word learning (Nation, 2000).

3 Procedure

A battery of eight tasks (including training sessions) was conducted over two consecutive days (Table 3). Note that in addition to the tone categorization and word identification tasks, participants also completed a word production task, which is not reported in this article.

Participants were told that they were taking part in a study that investigated the effects of audiovisual presentation on L2 vocabulary learning. After signing a consent form, participants completed the tasks individually. The first author only intervened at the start of new tasks to provide instructions. Written instructions for each task were in English or Mandarin. The experiment was carried out over two days to limit the total time spent in one session and to facilitate word recall after a night of sleep (Dumay and Gaskell, 2007).

All tasks were administered in a sound-attenuated booth and run on a touchscreen tablet laptop (*DELL Inspiron 13 5000 Series*) through the *OpenSesame* software (Mathôt et al., 2012). Participants listened to audio stimuli over *Beyerdynamic DT 990* headphones at a comfortable listening level.

a Tone categorization task. In the tone categorization task, participants listened to a vowel carrying one of the four tones and were asked to identify the tone by touching the corresponding arrow on the touchscreen. They were encouraged to make their choice as quickly as possible and to guess if unsure. Time-out was 5,000 ms after presentation of the audio stimulus.

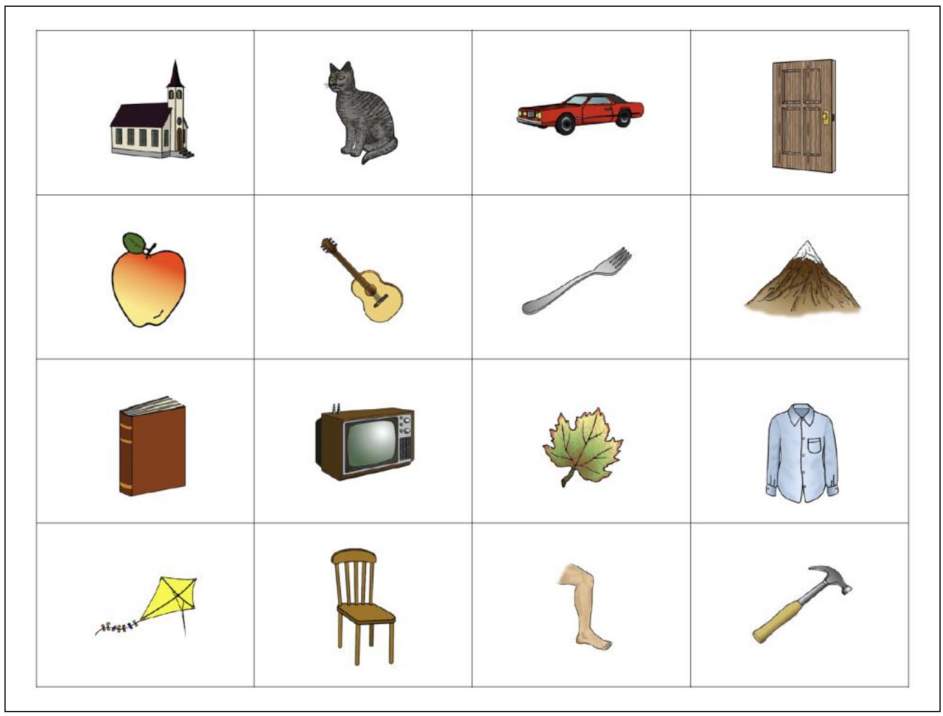


Figure 3. Visual stimuli for pseudolanguage words.

One practice session with 16 trials (4 presentations per tone) including feedback was held at the beginning. In the practice session, the vowel /o/ was used, which was not used in the main session. The practice session was followed by a main session in which there were 72 trials (6 presentations per stimulus) without feedback in a randomized order.

b Word training. The word training consisted of mimicry (listen-and-repeat), which was expected to be a relatively effective way to quickly memorize novel L2 words (Baills et al., 2019; M. Li and Dekeyser, 2017).[3] Participants were presented with the individual pseudolanguage words (the audio stimuli) and their meaning (the images). They were asked to repeat the words out loud and pronounce them as accurately as possible, whilst simultaneously trying to memorize the words. No feedback was given regarding their pronunciation.

After a familiarization with the images and their meanings in participants' native language to ensure that participants considered the images to be analogous to a word in their L1, each of the 16 pseudolanguage words was audiovisually presented 4 times, resulting in 64 trials in total. Participants had 5,000 ms to repeat the word before the next audiovisual stimulus was presented. The first two presentations were in a pseudorandomized order for all participants: each audiovisual stimulus was presented twice in a row (e.g.

Table 3. Overview of tasks.

DAY 1	
Description	Duration (minutes)
Tone categorization	5
Word training	10
*Word production**	*5*
Word identification	15
DAY 2	
Description	Duration (minutes)
Working memory	5–10
Word training	10
*Word production**	*5*
Word identification	15

Note. *Not reported in this article.

the word for 'cat', followed by the word for 'cat'), and the order was such that no segmental or tonal minimal pair followed one another. The last two presentations were fully randomized for each participant individually.

The same word training was conducted on day 2. The only difference was that the image familiarization was not conducted, and that the pseudorandomized presentation order was the reverse of that of day 1.

c Word identification task. The word identification task involved image-matching to replicate L2-to-L1 word recall and tone word learning, following Barcroft and Sommers (2014); Cooper and Wang (2012). Participants would hear a pseudolanguage word and were then prompted to identify the meaning of that word by making a 16-way choice on the touchscreen. The options were displayed on a 4 × 4 answer board, similar to Figure 3. Participants were encouraged to make their choice as quickly as possible and to guess if unsure. Time-out per trial was set to 10s.

Participants started with a practice block in which they received feedback to familiarize themselves with the task format, but also to further help them memorize the words through perceptual training (M. Li and Dekeyser, 2017). The feedback showed whether the participant's answer was correct or incorrect, and presented once more the correct sound-image combination. Each stimulus was presented twice, totaling 32 trials, in a randomized order. This practice block lasted about 5 minutes.

The practice block was followed by a main block without feedback. To avoid that participants would associate the audio stimulus with the physical position of the image on the answer board rather than with the actual image, the images' positions were shuffled in the main block. In the main block, each stimulus was presented 6 times, totaling 96 trials, in a randomized order. There was a small break after the participants had completed two-thirds of the task. The exact same task was repeated on Day 2, with the only

difference being that the images' positions on the answer boards were again shuffled in the practice and main blocks.

d Working memory task. WM was operationalized through a backwards digit span task, as one of the proxies of WM associated with retention of phonological and lexical information required for L2 perception and word learning (Baddeley, 2003; Goss, 2020, p. 28; Kormos and Sáfár, 2008).

Participants were instructed to repeat out loud in their native language and in backward order a sequence of digits presented to them on the screen. After a practice session, they were presented with a block of five 2-digit sequences (e.g. 1–7; 6–3; 2–5; 8–4; 9–5). Participants would move onto a next block of five n+1-digit sequences (e.g. 5–8–2; 6–9–4; etc.) and continue to do so if they correctly repeated at least three sequences per block. If participants did not reach this threshold, the task was aborted at the end of a block. The maximum attainable block consisted of five 8-digit sequences.

A percentage working memory score was calculated by dividing the total number of digits from fully correctly recalled sequences by the maximum attainable score (175). Mean working memory scores per group are reported in Table 1.

4 Data analysis

All analyses were performed in *R 4.0.1* (R Core Team, 2020). Figures were generated with the *ggplot2* package (Wickham, 2016). We present descriptive statistics and results from mixed-effects models to assess the effects of L1-specific and extralinguistic factors on performance in the tone categorization and word identification tasks. Null responses and responses with unnaturally fast reaction times ($<$ 250 ms) were removed, excluding 0.84% and 1.42% of data points from each task, respectively. Because accuracy scores in the tone categorization task revealed a ceiling effect, we analysed reaction times (RTs) as a main proxy of performance. For RT data, only data for correctly categorized items were analysed. RT data were log-transformed and outliers (2.5 SDs from the mean) were removed, following Chan and Leung (2020). For the word identification task, in which there was considerably more variability in accuracy (% correctly recalled words), accuracy scores rather than RT were analysed as a proxy of performance.

Models were computed in the *lme4* package (Bates et al., 2015) and fitted with the *bobyqa* optimizer where applicable. Model diagnosis (observation of residual QQ plots) was carried out with the *DHARMa* package (Hartig, 2020). We adhered to a maximum Variance Inflation Factor (VIF) threshold of 5 (O'Brien, 2007) in all final models. None of the models showed multicollinearity. Post-hoc power simulations were carried out using the *simr* package (Green and MacLeod, 2016).[4]

We built models based on our research questions, including fixed effects and interactions of interest. The model for tone categorization (dependent variable: log RT) contained fixed effects for *L1* (English, Mandarin; contrast-coded), *tone* (rise, fall, mid-level, low-level; contrast-coded), *musical experience* (a continuous variable expressing years of playing a musical instrument; scaled and centered), and *working*

memory (a continuous variable expressing WM score; scaled and centered), and the three-way interactions *L1*tone*musical experience* and *L1*tone*working memory*.

The final model for word identification (dependent variable: correct/incorrect) contained the same fixed effects and interactions as the tone categorization model, but in addition contained a fixed effect of *tone categorization* (a continuous variable expressing log RTs in the tone categorization task; centered and scaled), and an *L1*tone*tone categorization* interaction to see to what extent tone perception predicts performance in tone word learning. All final models contained *Subject* (individual participant) and *Item* (stimulus) as random intercepts. Attempts were made to include random slopes but this led to convergence issues. To assess the interactions in more detail, Bonferroni-corrected multiple comparisons were generated using the *emmeans* package (Lenth, 2020).

IV Predictions

Based on the literature reviewed in Section II, we make the following predictions for our tasks in response to our research questions:

- Research question 1: Mandarin participants are expected to have slower reaction times for mid-level and low-level tones. English participants may be better at quickly categorizing level tones as opposed to contour tones. We therefore expect an interaction between *L1 tone type* and *L1 tonal status* in the tone categorization task. Although we are not aware of any previous literature that has investigated the effect of *L1 tone type* in tone word learning, we expect the general familiarity with associating f_0 to lexical meaning (i.e. *L1 tonal status)*, rather than the familiarity with specific pitch contours (i.e. *L1 tone type*), to be a stronger predictor of performance in the word identification task. Mandarin participants are thus expected to overall outperform English participants in accurately recalling tonal pseudolanguage words.
- Research question 2: It is expected that musical experience will not necessarily facilitate tone categorization in Mandarin speakers, but it may do so for mid-level and low-level tones, which are expected to be relatively challenging and may be identified faster by musicians than by non-musicians. Musical experience is not expected to strongly predict word identification performance in Mandarin speakers. For English speakers however, musical experience is expected to be a strong predictor of performance in both tone categorization and word identification. We therefore expect an interaction between *L1 Tonal Status* and *musical experience*. In both groups, working memory is only expected to facilitate word identification performance.

V Results

We first present an overview of performance in the tone categorization and word identification tasks in Section V.1, after which we present model results in Section V.2 to investigate how our predictors of interest (*L1*, *tone*, *musical experience and working memory*) affected variability in performance.

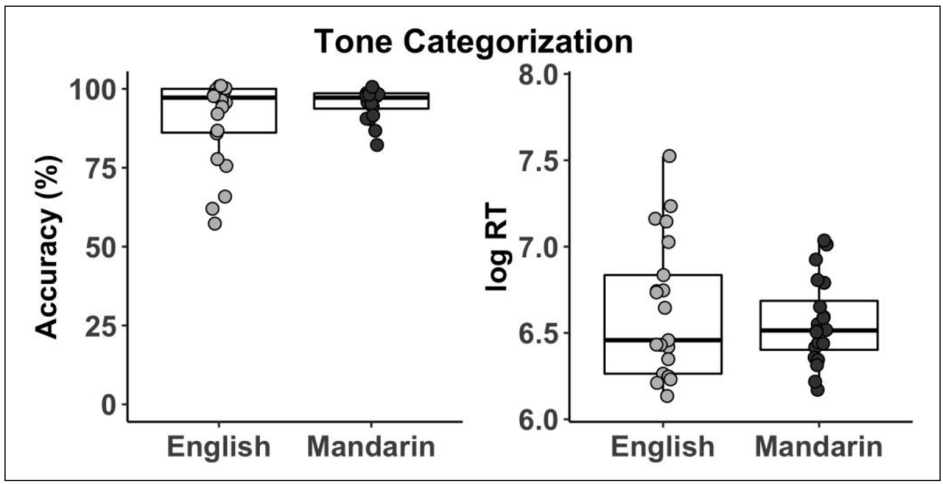


Figure 4. Accuracy and log reaction times (RT) for tone categorization per group.

1 Overview of performance and individual variability

a Tone categorization. Figure 4 shows accuracy scores and log-transformed reaction times (RTs) for the tone categorization task. A visual inspection reveals no stark difference between the English and Mandarin group, either in terms of accuracy or reaction time. As mentioned earlier, because of a ceiling effect observed for the accuracy scores, we will focus on log RTs in subsequent analyses as a measure of tone categorization performance (for an alternative analysis of tone categorization performance based on accuracy scores, see supplemental material 5).

b Word identification. Figure 5 shows accuracy and log RT for the word identification on days 1 and 2. A visual inspection suggests that participants improved their accuracy scores over the two sessions, but that large individual differences exist both in the English and the Mandarin group. RTs were not the focus of our analysis for the word identification task, but a visual inspection suggests that log RTs did not differ greatly between groups or across days.

2 Model results

To account for the observed individual variability in the tone categorization and the word identification tasks, this section highlights significant effects and interactions found in our models. Note that we only present data from the main block on day 2 of the word identification task. This is for brevity but also because we consider data from day 1 to be intermediate, as the word training had not been fully completed then.

A summary of all significant ($p < .05$) effects and interactions is provided in Table 4 (full details are in Appendices 3–4). Following our research questions, we will first address the effects and interactions of *L1* and *tone* in Sections V.II.a–V.II.c, after which

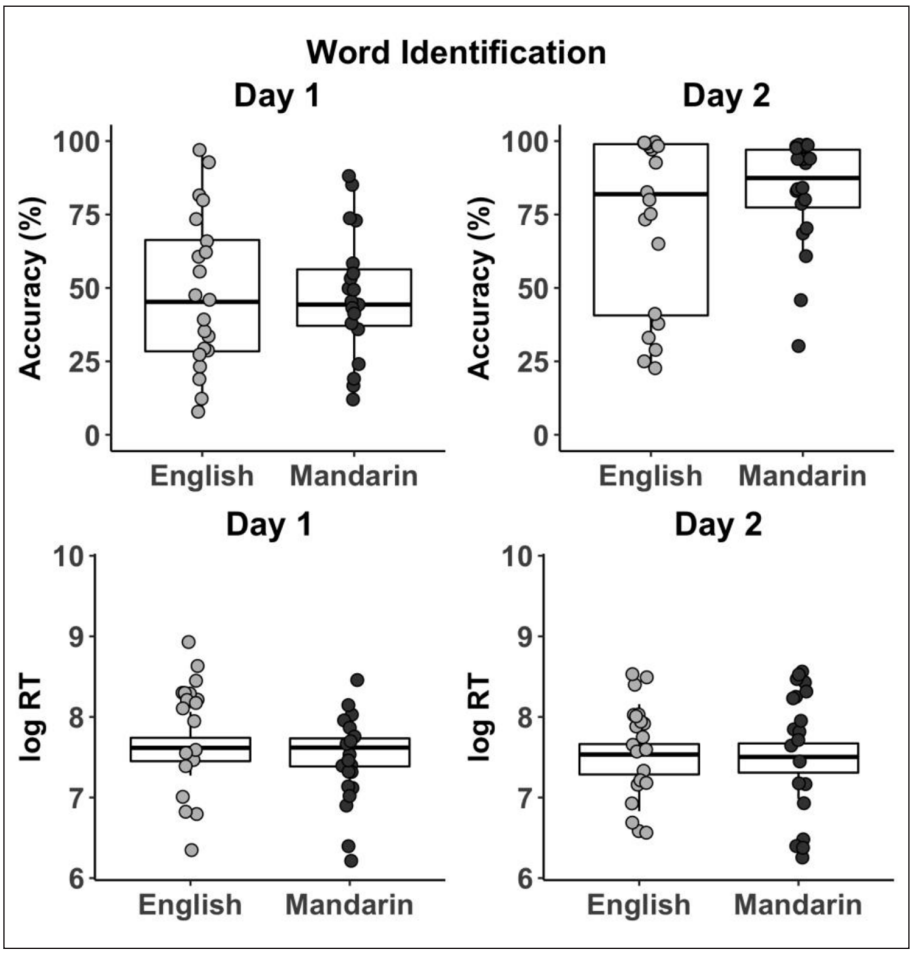


Figure 5. Accuracy and log reaction times (RT) for word identification per group.

we will highlight the effects of *musical experience* and *working memory* in Sections V.II.d–V.II.f.

*a L1*tone interaction.* As shown by the log RTs and accuracy scores in the tone categorization and word identification tasks in Figure 4–5, overall performance between both groups was comparable, and the models revealed no significant main effect of *L1* in either of the tasks. However, in both tasks, there were significant *L1*tone* interactions.

To investigate these interactions in more detail, we first focus on significant multiple comparisons (fully reported in Appendices 5–6).[5] For tone categorization, there were no significant comparisons between groups, nor between tones within the English group. Within the Mandarin group, mid-level ($b = 0.20$, $SE = 0.06$, $p = .027$) and low-level tones ($b = 0.20$, $SE = 0.06$, $p = .047$) were categorized significantly slower in

Table 4. Summary of significant effects and interactions ($p < .05$).

Tone categorization task (logRT)*	Word identification task (accuracy)**
Musical experience	*Tone*
*L1*tone*	*Musical experience*
*L1*musical experience*	*Working memory*
*L1*tone*musical experience*	*L1*tone*
	*L1*musical experience*
	*L1*working memory*
	*Tone*tone categorization*

Notes. *lmer(logRT ~ L1*tone*musical experience + L1*tone*working memory + (1|Subject) + (1|Item)). **glmer(correct ~ L1*tone*musical experience + L1*tone*working memory + L1*tone*tone categorization + (1|subject) + (1|item)).

comparison to falling tones. A visualization of log RT per tone between groups in Figure 6 shows that indeed, log RTs are similar between groups, and similar between tones within the English group, but that within the Mandarin group, mid and low tones were categorized more slowly.

For word identification, multiple comparisons revealed that Mandarin participants were significantly less likely than English participants to identify words carrying a low-level tone ($b = -1.11$, $SE = 0.47$, $p = .018$). There were no significant comparisons between tones within the English group. Within the Mandarin group, words carrying a low-level tone were significantly less likely to be identified than words with a rising ($b = -1.15$, $SE = 0.35$, $p = .005$) and a falling tone ($b = -1.23$, $SE = 0.34$, $p = .002$). A visualization of word identification accuracy per tone between groups in Figure 7 reflects the finding that whereas English participants' word identification accuracy did not vary much between tones, Mandarin participants' accuracy was lower for words carrying a low-level tone.

b Error types in tone categorization. To further investigate how tone type affected tone categorization performance, this section presents error types. Figure 8 displays the count of error types in tone categorization averaged over each participant. For instance, a 'Rise-to-Fall' error indicates that upon hearing a vowel with a rising tone, a participant miscategorized that as a falling tone. A visual inspection of the distribution of all possible 12 error types suggests that English participants miscategorized tones relatively across the board, whereas Mandarin participants predominantly miscategorized mid-level tones as low-level tones and vice versa. Mixed-effects models and multiple comparisons (Appendix 7) revealed that, in the English group, some error types occurred significantly more often than others. Fall-to-mid and low-to-mid errors were more likely to occur in comparison to 5 and 3 other error types, respectively. In the Mandarin group, only the mid-to-low and low-to-mid errors were more likely to occur in comparison to 1 and 3 other error types, respectively.

c Tone-only error types in word identification. It is worth noting that on day 2 of the word identification task, the majority of errors were 'tone-only errors' (Wong and Perrachione,

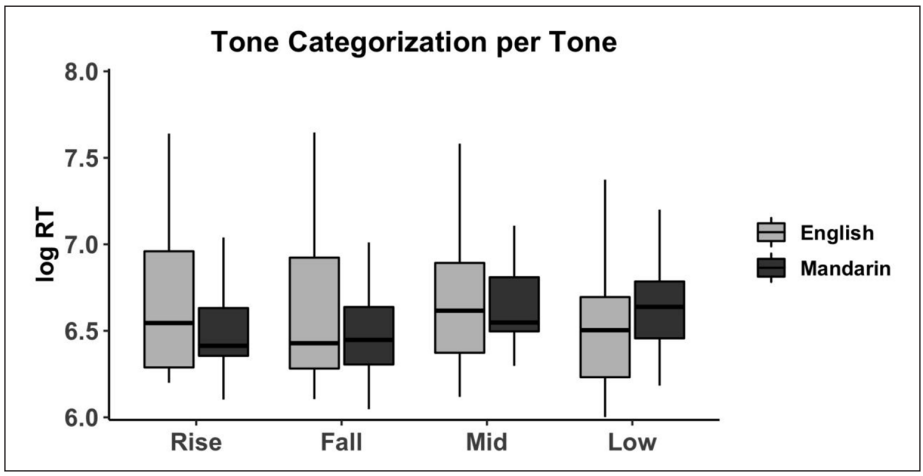


Figure 6. Tone categorization log reaction times (RT) per tone between groups.

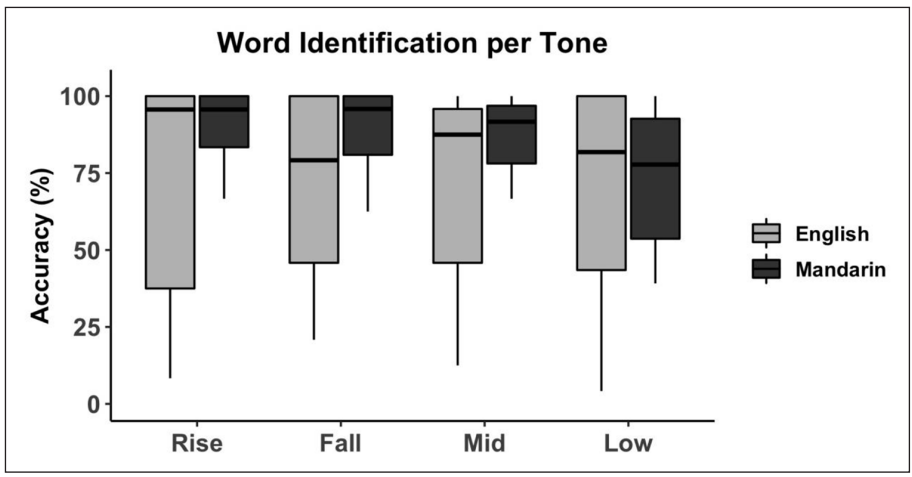


Figure 7. Word identification accuracy per tone between groups.

2007), meaning that participants misidentified a word purely because of its tone, e.g. misidentifying /juɹ15/ as /juɹ22/. Tone-only errors accounted for 73.20% (SD = 34.89) of all errors in the English group and for 64.96% (SD = 31.11) of all errors in the Mandarin group. For visualization, Figure 9 plots the number of word identification errors against the number of tone-only errors. Two simple linear regressions confirmed that the number of tone-only errors significantly predicted the total number of errors and explained a large portion of variance in both the English [$F(1,19) = 91.670$, $p < .001$, $R^2 = .8193$] and the Mandarin group [$F(1,18) = 100.300$, $p < .001$, $R^2 = .8393$]. This

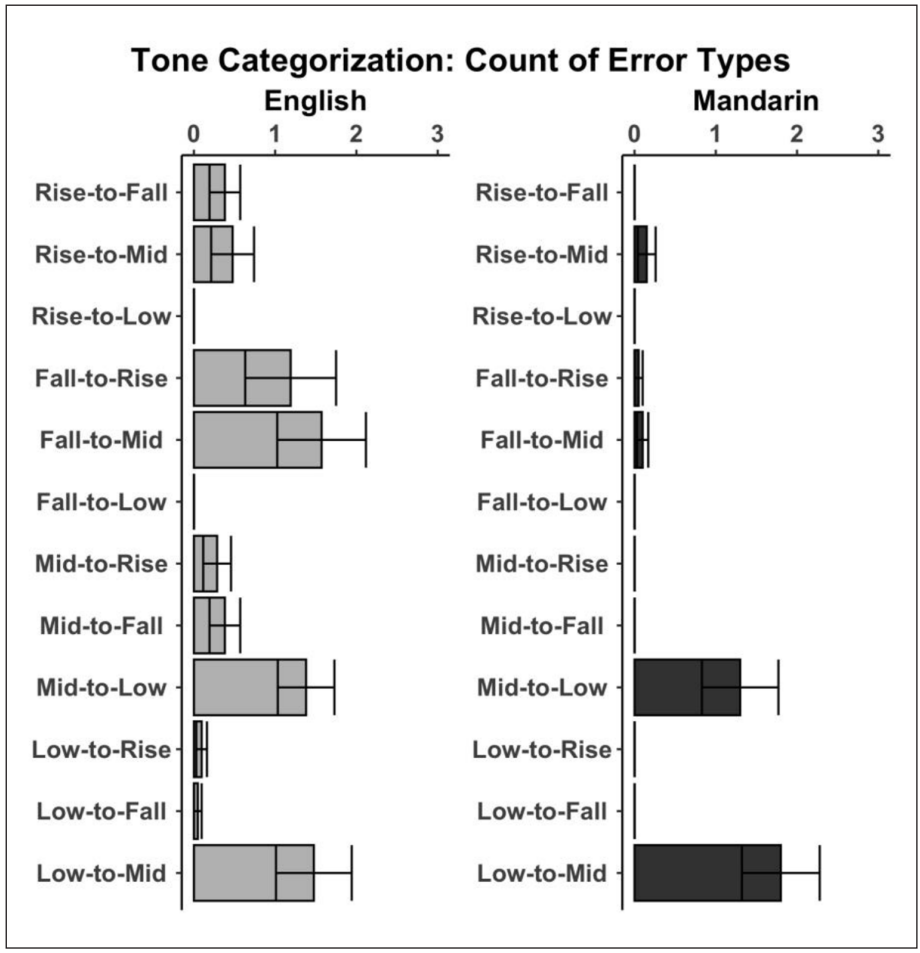


Figure 8. Error types in tone categorization.
Notes. Counts are averaged over subject/participant. Error bars indicate standard error.

suggests that many participants had acquired the segmental, but not the tonal properties of the words at the end of the experiment.

To further investigate the nature of these tone-only errors, Figure 10 displays the distribution of tone-only error types. Similar to the error types in tone categorization (as presented before in Figure 8), it appears that English participants confused tone in words across the board, with no single error type particularly standing out. Mandarin participants however, seem to have made more low-to-mid errors in comparison to other errors. Mixed-effect models and multiple comparisons (Appendix 8) revealed that among the 12 possible error types, there was no indication of one particular error type occurring more often than others in the English group, although it is worth noting that fall-to-mid errors were more likely to occur in comparison to 5 other error types, and that low-to-mid errors

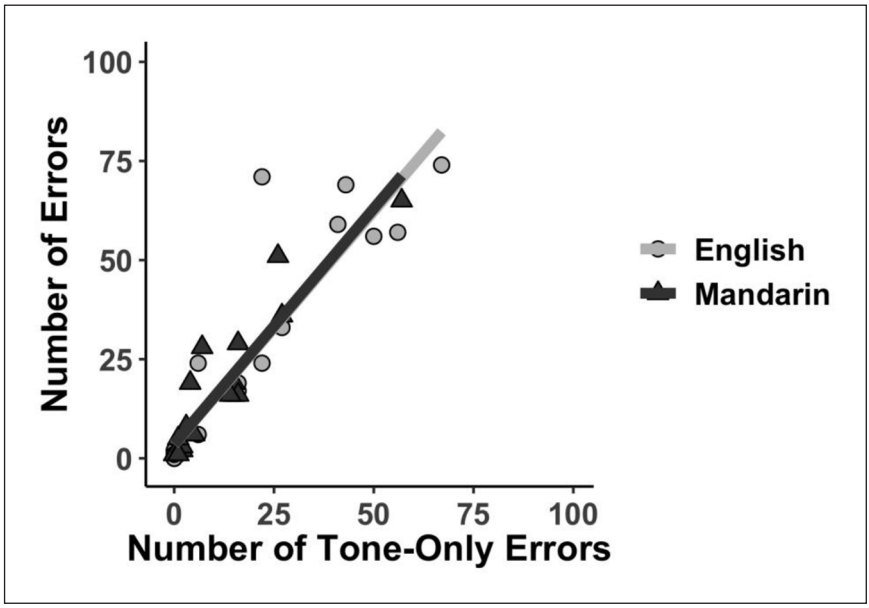


Figure 9. Number of errors against number of tone-only errors in word identification.

were more likely to occur in comparison to 3 other error types. In the Mandarin group, there was a clear indication that the distribution of tone-only errors was skewed toward the low-to-mid type, which was significantly more likely to occur in comparison to almost all other 11 error types, except the mid-to-low error type. The mid-to-low error type was significantly more likely to occur in comparison to 2 other error types.

*d L1*musical experience interaction.* In tone categorization, *musical experience* led to faster log RTs in the English group ($b = -0.28$, $SE = 0.08$, $p = .002$), but not in the Mandarin group ($b = -0.05$, $SE = 0.07$, $p = .699$; full details in Appendix 9). Note that these are trends in the overall tone categorization task averaged over the four different tones: there was also a significant three-way *L1*tone*musical experience* interaction, suggesting that the interaction between L1 and musical experience differed between tones.

To investigate the origin of this interaction, the effect of *musical experience* was analysed per group and per tone. Multiple comparisons in Appendix 10 revealed that the effect for *musical experience* was significantly larger for the English group compared to the Mandarin for rising ($b = -0.25$, $SE = 0.11$, $p = .019$) and falling tones ($b = -0.31$, $SE = 0.11$, $p = .005$), but not for mid-level ($b = -0.17$, $SE = 0.11$, $p = .106$) and low-level ($b = -0.19$, $SE = 0.11$, $p = .076$) tones. A further post-hoc comparison revealed that the effect of *musical experience* was significantly larger for falling tones than for low-level tones within the English group ($b = -0.11$, $SE = 0.03$, $p = .036$).

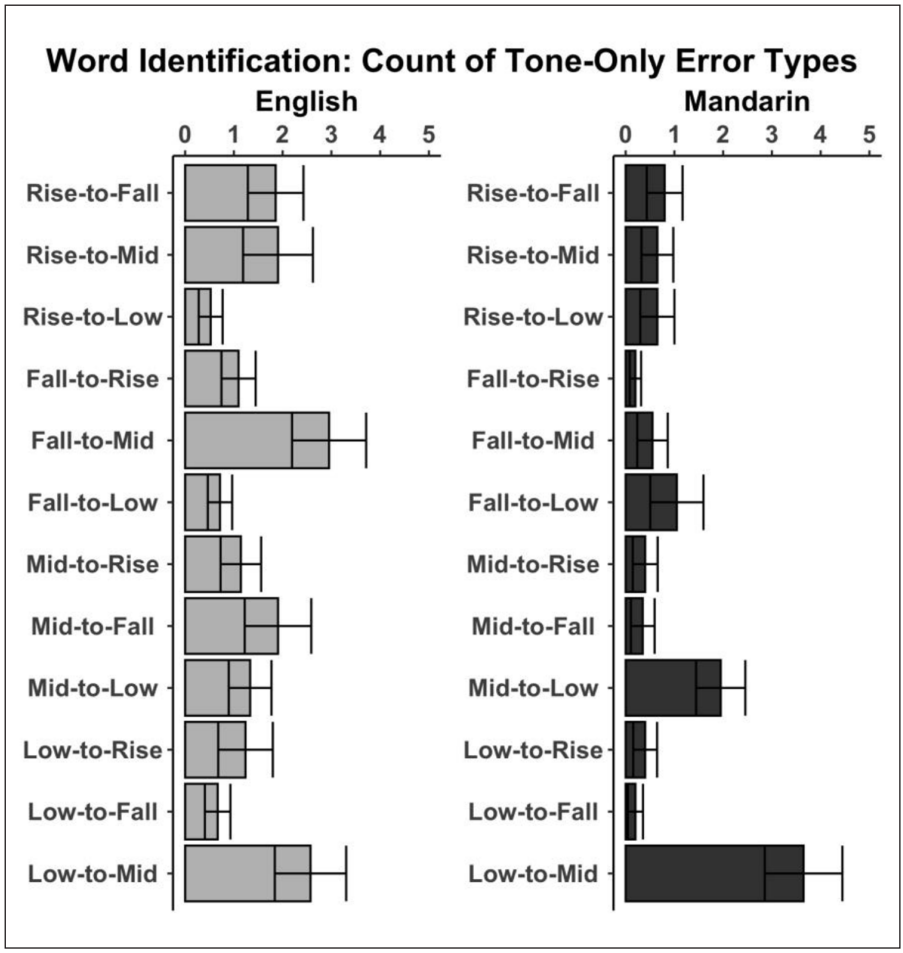


Figure 10. Tone-only error types in word identification.
Notes. Counts are averaged over subject/participant. Error bars indicate standard error.

This is illustrated in Figure 11, which plots tone categorization log RT against musical experience per tone. For the English group, it can be observed that the effect of musical experience is relatively strong (i.e. relatively steeper slopes) for rising and falling tones, and slightly less so for mid-level and low-level tones. For the Mandarin group, the flat slopes indicate that musical experience did not lead to faster log RTs in any of the tones.

In the word identification task, *musical experience* significantly increased the likelihood of correct word identification in the English group ($b = 2.21$, $SE = 0.45$, $p < .001$), but not in the Mandarin group ($b = 0.48$, $SE = 0.29$, $p = .183$; full details in Appendix 11).

For visualization, Figure 12 illustrates the *L1*musical experience* interactions. It can be observed that whereas English participants appear to benefit from musical experience

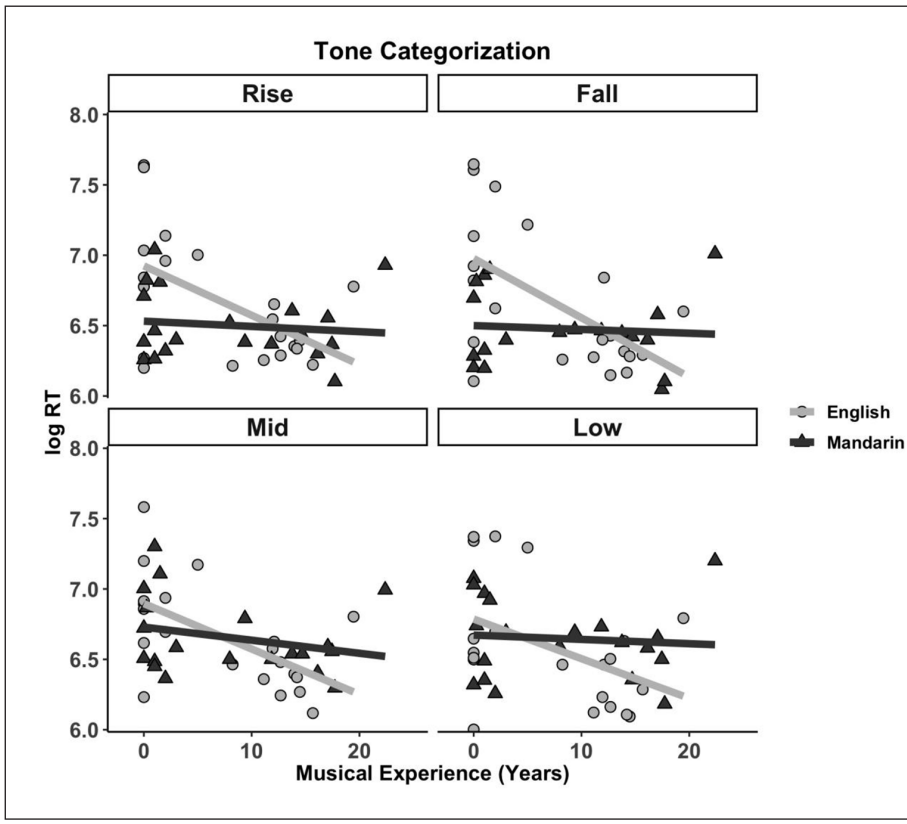


Figure 11. Tone categorization log RT against musical experience per tone.

(yielding to faster RTs in tone categorization and higher accuracies in word identification), this trend is absent in the Mandarin participants.

*e LI*working memory interaction.* *Working memory* did not predict performance in the tone categorization task for either group.

In the word identification task, *working memory* did not significantly increase the likelihood of correct word identification in the English group, but it did in the Mandarin group ($b = 1.91$, $SE = 0.31$, $p < .001$; full details in Appendix 11). This finding is illustrated in Figure 13. Note that although the trend line would suggest otherwise, there was no statistical confirmation that WM, alongside our other predictors of interest, predicted English participants' performance in the word identification task ($b = 0.06$, $SE = 0.35$, $p = .982$, 95% CI [–0.63, 0.75]).

f Tone categorization performance as a predictor of word identification performance. Tone categorization log RTs did not predict word identification performance in neither group in our model, however there was a significant *tone categorization*tone* interaction.

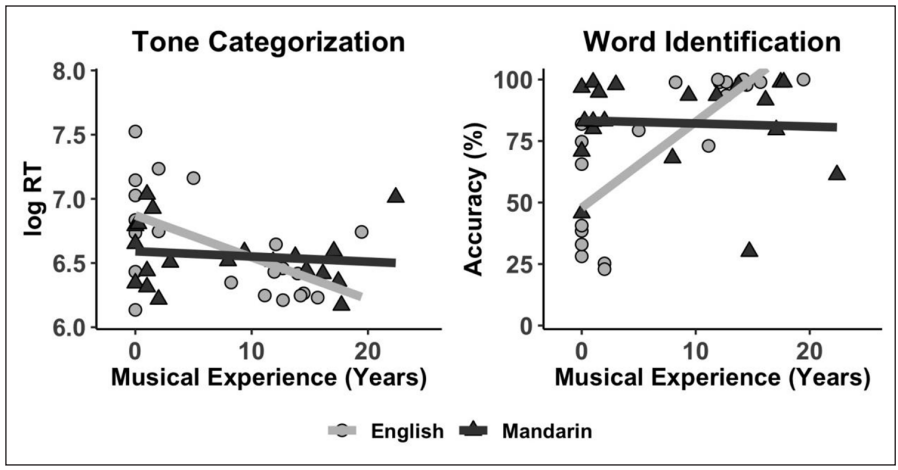


Figure 12. L1*musical experience interaction.

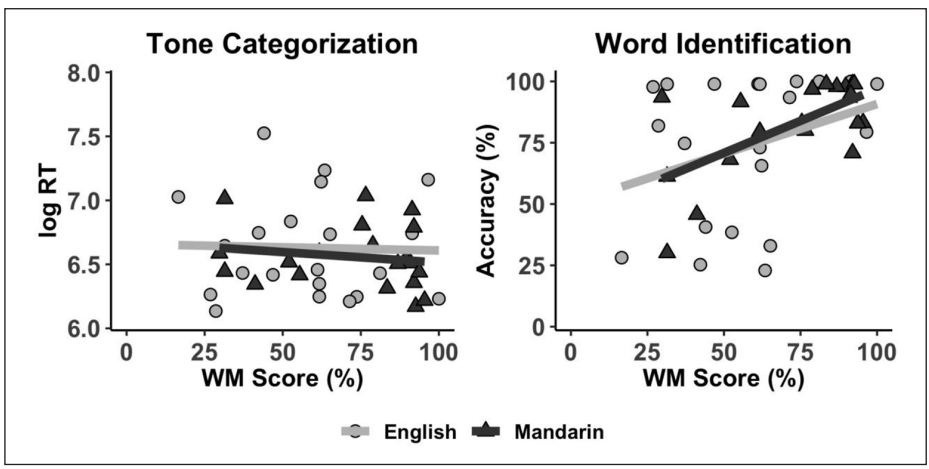


Figure 13. L1*working memory interaction.

Post-hoc multiple comparisons revealed that for both groups together, the effect of *tone categorization* was largest for words with rising tones, however this effect on its own failed to reach significance ($b = -0.63$, $SE = 0.27$, $p = .077$; 95% CI [–1.16, –0.10]).

VI Discussion

This study's aim was to examine the combined effects of individual learners' L1-specific and extralinguistic factors as predictors of L2 tone perception and word learning facility. We will now discuss our findings in light of our research questions and previous research.

1 Effects of L1 tonal status and L1 tone types on tone categorization and word identification

Research question 1 addressed how L1 tonal status and L1 tone type affect individual performance in both pre-lexical and lexical processing of tones. In the tone categorization task, which addressed pre-lexical tone perception, most participants attained near-ceiling performance in terms of accuracy, but they showed more individual variability in reaction times. This variability was not directly attributable to L1 tonal status, as Mandarin listeners were not significantly faster than English listeners in categorizing tones. Instead, as predicted, variability was explained by an interaction between L1 tonal status and L1 tone types.

Specifically, Mandarin participants categorized mid-level and low-level tones slower than falling tones, and the error analysis further revealed that they predominantly miscategorized low-level tones as mid-level tones and vice versa. This suggests that telling apart low-level from mid-level tones constituted the real difficulty for the Mandarin participants in the tone categorization task. This finding is interpretable when considering Mandarin L1 tone types: in phonological-categorical terms, Mandarin listeners may have assimilated our low-level and mid-level tones to their L1 high-level tone, making the level distinction difficult. As pointed out by Francis et al. (2008, p. 284), any claims regarding categorical assimilation can only be 'speculative in nature'. This is especially the case in our study since we did not ask our participants to explicitly rate the similarity between target and L1 tones (J. Chen et al., 2020; Reid et al., 2015). Nevertheless, it is worth noting that, although purely anecdotal, many Mandarin participants did indicate that the mid-level and low-level tones were particularly difficult to categorize because they had no clear equivalents in Mandarin, unlike the rising and falling tones.

Alternatively, an acoustic-phonetic interpretation as to why Mandarin participants appeared to struggle with quickly categorizing level tone contrasts would be that they put relatively more weight on differences in f_0 direction rather than in f_0 height (Francis et al., 2008; Gandour and Harshman, 1978; Qin and Jongman, 2016). It is additionally possible that the categorization of low-level tones was complicated because of absence of phonation cues (creaky voice), which contributes to native speakers' perception of the low-dipping tone in Mandarin (Yang, 2015). Indeed, in real tone languages, acoustic cues such as phonation (Tsukada and Kondo, 2019) and duration (Liu and Samuel, 2004) can contribute to the overall salience of different tone types.

As to the English speakers, log RTs did not significantly differ across tones. The error analysis further revealed that English participants tended to confuse tone types with one another in every direction, incorrectly categorizing both contour as level tones (fall-to-mid) and level as level tones (low-to-mid) relatively often. Although again we cannot ascertain whether English listeners relied on L1 f_0-based categories in their tone categorization, whatever reliance on intonational categories English participants may have had, it appears that these did not affect performance, as performance on individual tones was equal across the board. This resonates with Best's (2019) conclusion that assimilations of L2 tones to intonational distinctions may be 'less categorical than are assimilations to another lexical tone system' (p. 5). Although we had tentatively predicted that English speakers would categorize level tones faster than contour tones

based on a phonetic-acoustic approach of tone type, this was not borne out by our data. Rather than being affected by tone type, English participants' performance appeared to be largely guided by their musical experience, as will be discussed in the next section.

Our findings from the word identification task suggest that L1 tonal status and L1 tone type modulated performance in a similar way as in the tone categorization task: differences between the English and Mandarin groups were not seen in overall performance (against our predictions), but in performance per tone. The error analysis showed that in both groups, most word identification errors were tone-only errors, suggesting that tonal rather than segmental distinctions were the hardest feature to memorize in the pseudolanguage words. However, *which* tonal distinctions were hardest to learn appeared to be strongly influenced by L1 tone type, as Mandarin participants were less likely to identify words with low-level tones compared to words with rising and falling tones, and even compared to English participants. Mandarin participants predominantly misidentified low-level tone words as mid-level tone words, whereas the English participants confused tones on words across the board.

In sum, our findings addressing research question 1 show that L1 tone type not only interferes in pre-lexical tone processing, as has been shown widely in previous studies (Cooper and Wang, 2012; Hao, 2012; Qin and Jongman, 2016; So and Best, 2010; X. Wu et al., 2014), but also in lexical processing, and in remarkably similar ways. It is crucial to note that in our study, this effect appeared to be strong enough that Mandarin participants, who by virtue of their L1 tonal status would be expected to overall outperform non-tonal peers in L2 tone word learning (Chan and Leung, 2020; Poltrock et al., 2018), were in fact less likely to recall low-level tone words than non-tonal English participants. This highlights that L1 tonal status alone cannot fully account for individual differences in neither tone perception nor tone word learning facility, and that it is crucial to simultaneously factor in the effect of L1 tone type. It is worth noting that if our pseudolanguage had contained the exact same tone types as in Mandarin, we would have expected Mandarin participants to outperform the English speakers, thereby indirectly showing an overall facilitative effect of L1 tonal status.

2 Combined effects of L1-specific and extralinguistic factors

In research question 2, we asked how musical experience and working memory affect individual performance in tone perception and tone word learning, and whether the effects of these extralinguistic factors are modulated by L1-specific factors.

We found that, in line with our predictions, musical experience significantly predicted tone categorization performance for English but not for Mandarin participants. Even for mid-level and low-level tones, which were relatively difficult for Mandarin participants, musical experience did not lead to faster RTs. The absence of a facilitative effect of musical experience on tone perception for Mandarin speakers in our study chimes in with earlier findings (Tang et al., 2016; Wong et al., 2020; H. Wu et al., 2015), although it is worth noting that finding such a facilitative effect may be task-dependent (D. Chang et al., 2016). For instance, Qin et al. (2021) tentatively suggest that musical *ability* (a different measure of musicianship) may in fact enhance perception (as measured by discrimination and identification accuracy) of Cantonese level tone contrasts for

Mandarin-L1 speakers. We interpret however that in our tone categorization task, Mandarin participants' performance was largely guided by the effect of L1 tone type, and that this may have overridden any facilitative effect of musical experience on tone perception.

English participants did appear to benefit from musical experience, as musical experience led to significantly faster reaction times. In addition, the *L1*tone*musical experience* interaction revealed that musical experience particularly facilitated categorization of falling tones as opposed to low-level tones. This suggests that English listeners, who have been found to pay less attention to f_0 contour differences than to f_0 height differences, particularly in falling contours (Jongman et al., 2017), may have benefited from additional pitch acuity derived from musical experience to quickly categorize 'difficult' falling tones.

In the word Identification task, we similarly found that musical experience predicted performance for English but not for Mandarin participants. Our interpretation is similar to that of Cooper and Wang (2012, pp. 4765–4766), who suggest a 'differential in relevance of musicality depending on linguistic background' in tone word learning. Namely, Mandarin participants, who are already familiar with the use of pitch for lexical purposes, may not benefit as much from enhanced pitch acuity gained through musical experience as English participants do.

In sum, these findings suggest a dynamic interplay of musical experience and L1 tonal status in L2 tone perception and word learning. We note that we only measured musical experience in terms of years of musical practice, and that more refined measures of musicality (Wallentin et al., 2010) might reveal different results.

As predicted, we did not find a significant facilitative effect of working memory on pre-lexical pitch processing in the tone categorization task for neither English nor Mandarin participants. Although this finding falls in line with existing literature that suggests that WM has a null, or limited effect on performance in relatively undemanding pre-lexical pitch perception tasks (Bidelman et al., 2013, p. 8; Goss, 2020; Goss and Tamaoka, 2019), we are aware that we only measured backwards digit span as a rough proxy of WM, and future studies could assess whether other cognitive measures, such as attentional resources or executive function, are linked to tone perception.

To the best of our knowledge, our study is the first of its kind that incorporates a measure of WM in assessing the combined effects of L1 tonal status, L1 tone type, and musical experience in tone word learning. We found that when considering all these factors together, WM significantly predicted word recall of tonal pseudolanguage words for Mandarin but, unexpectedly, not for English participants, for whom musical experience was the only significant extralinguistic predictor. The finding for English participants resembles that of Bowles et al. (2016), who found that variance in English learners' performance in Mandarin tone word learning was only partially explained by domain-general memory skills, and most strongly by pitch-specific skills, suggesting that 'mastery of a feature of a target language known to be particularly challenging for L2 learners – as a necessary component of learning the language at large – is predicted most successfully by behavioral measures that are most relevant to that feature' (Bowles et al., 2016, p. 775). In other words, our word identification task may have been particularly challenging for English participants because it involved *tone* words, and therefore individual participants with better pitch acuity (assumed to be derived from musical experience)

would benefit from these skills to memorize words based on tonal distinctions. Mandarin participants, by virtue of their L1 tonal status, may not have found recalling our pseudolanguage words particularly challenging because they contrasted in tone per se (except for the distinction between level tone words). This could explain why their ability to recall our pseudolanguage words was mainly guided by WM capacity as a general predictor of L2 vocabulary recall (Cheung, 1996; Kormos and Sáfár, 2008) rather than pitch-specific skills.

Finally, our models revealed that, when also accounting for other L1-specific and extralinguistic factors, pitch perception ability in the tone categorization task (as measured by log RTs) did not independently predict performance in word identification. However, this does not imply that performance in the pre-lexical tone categorization task was completely unrelated to performance in the lexical word identification task. For instance, the tone error patterns largely mirrored one another across both tasks. It is also worth noting that in our alternative model of word identification in which we used tone categorization accuracy instead of log RT as a proxy of pitch perception ability, we did find a main effect of tone categorization accuracy on word identification likelihood, and post-hoc analyses showed that tone categorization accuracy predicted word identification accuracy for rising tone words for English participants and for mid-level tone words for Mandarin participants (supplemental material 5). Although we are cautious to derive strong conclusions from this alternative analysis given the near-ceiling accuracy scores in the tone categorization task, this may suggest a link between performance in pitch perception and lexical pitch processing in our tasks. Our general findings, in which we used log RTs as a proxy of pitch perception ability, reveal that tone word learning performance in English participants was mainly facilitated by musical experience, and in Mandarin participants mainly by WM capacity, which may fill the gap when neither musical experience nor pitch perception ability strongly facilitate tone word recall.

Thus, addressing research question 2, it appears that any facilitative effect of musical experience and working memory on pre-lexical and lexical tone processing is indeed modulated by L1 tonal status: for non-tonal English learners, musical experience appears to be facilitative for tone perception and word learning, whereas for tonal Mandarin learners, individual performance is guided by L1 tone type and working memory (the latter only for word identification). The findings from our study thus suggest that the ease with which L2 tones are perceived and learned depends on a dynamic interplay between L1 tonal status, L1 tone type, musical experience, and working memory. This provides a more refined account of the several factors that determine an individual learner's aptitude to explain the large variability observed in L2 tone perception and word learning facility, beyond what has been described in previous studies that separately assessed the factors included in this study.

Future studies should examine the combined effect of L1-specific and extralinguistic factors in tone word learning in more naturalistic settings than our pseudolanguage word identification task, for instance in tasks in which learners process tones in sentence contexts or multi-speaker environments. As pointed out by a reviewer, the fact that we only modified f_0 and kept other acoustic parameters constant may limit the applicability of our findings to real tone languages, in which secondary acoustic cues can play a role in tone processing. Future studies should thus include a wider range of native and non-native tone systems to further refine our understanding of a dynamic interplay between L1-specific and extralinguistic factors in L2 tone learning.

VII Conclusions

This study aimed to account for individual differences in L2 tone perception and tone word learning by assessing the combined effects of L1-specific and extralinguistic factors, testing a combination of factors that were only addressed separately in earlier studies. We argue that none of the L1-specific and extralinguistic factors determine learning outcomes in and of themselves, but that both go hand-in-hand and dynamically affect tone perception and tone word learning performance in the individual and thereby shape the profile of learners who are expected to do relatively well, and learners who are expected to do relatively poorly in early-stage tone learning. Our findings suggest that a complete theoretical model of tone learning would ideally acknowledge this 'dynamic' and 'multi-systemic' nature of L2 speech-learning (A. Li and Post, 2014). That is, our study shows that a comprehensive theory of L2 tone learning facility should not only be able to account for extralinguistic factors that shape individual performance in early-stage tone learning – such musical experience and working memory – but it should also be able to account for any L1-specific factors – L1 tonal status and L1 tone type, here – which interact with extralinguistic factors to modulate individual performance in complex ways.

Acknowledgements

We thank three anonymous reviewers and the associate editor for their valuable comments on earlier versions of this article. Special thanks go out to the members from the Cambridge University Phonetics Lab for their valuable feedback throughout the course of this study, and to the volunteer participants.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was conducted as part of T.J. Laméris' doctoral research, funded by the Economic and Social Research Council (Grant 2117864) and a St John's College Learning and Research Fund.

ORCID iD

Tim Joris Laméris https://orcid.org/0000-0002-1365-3022

Supplemental material

Supplemental material for this article is available online. Supplemental material comprises:

1. Raw data.
2. Codes for figures and models.
3. Stimulus sound files.
4. Details on performance by Mandarin participants with knowledge of other Chinese languages.
5. Alternative analysis of tone categorization task based on accuracy instead of reaction time.

Notes

1. Chan and Leung (2020) investigated the effect of tonal status (L1 Cantonese and L1 English) and musical experience on 'phonological learning' (in between pre-lexical and lexical learning) of Thai tones. Chang et al (2016) investigated the effect of tonal status (L1 Mandarin and L1 English) and musical experience on Mandarin and musical tone perception. Chen et al. (2020) investigated the effect of tonal status (L1 Mandarin and L1 English) and musical experience on tone perception of meaningless syllables. Cooper and Wang (2012) investigated the effect of tonal status (L1 Thai and L1 English) and musical experience on Cantonese tone perception and word learning.
2. We note that some of the Chinese L2s reported by our Mandarin speakers have level tone contrasts unlike Mandarin, which may have affected performance on our mid- and low-level tones. However, a visual inspection of performance by participants who reported a L2 with level tone contrasts versus participants who did not, did not reveal notable differences (see supplemental material 4). In addition to the fact that all participants reported that Mandarin was their L1 and the language they used the most, we therefore deemed it fit to group these participants together.
3. We take note of empirical evidence that suggests that production during training may disrupt perceptual learning of the non-native sound to be learned, at least in certain pre-lexical tasks and when production and perception are required within the same trial (Baese-Berk and Samuel, 2016). Although our study did not investigate the effect of different training paradigms, it is worth noting that our participants reached relatively high word identification scores after only two training sessions (involving both mimicry and word identification with feedback) in comparison to similar tone word learning studies that only involved feedbacked word identification trials: Participants in Cooper and Wang (2012) completed seven 30-minute training sessions spread out over two weeks to learn 15 Cantonese tone words (3 syllables $\times$ 5 tones), and mean word identification of accuracy was 67%. In addition, the mimicry task was included in our study because – although not reported in this article for brevity – participants were also tested on their word production, which was expected to benefit from training in the same modality (Baese-Berk, 2019; M. Li and Dekeyser, 2017).
4. The observed power in our models – using the *simr* package (Green and MacLeod, 2016), following Wiener et al. (2020) – for 100 simulations was 92.00% (CI: 84.84, 96.48) for *musical experience* and 77.00% (67.51, 84.83) for the *L1*musical experience* interaction in the tone categorization model. In the word identification model, it was 100.0% (96.38, 100.00) for *musical experience* and 95.00% (88.72, 98.36) for the *L1*musical experience* interaction. We acknowledge the limitations of post-hoc power analyses (Hoenig and Heisey, 2001).
5. Note that in the tables, multiple pairwise comparisons are made with reference to the latter element in a pair, as obtained by the list(pairwise~) command in *emmeans*. For instance, in Appendix 5, the 'Fall-mid' comparison with a negative b-estimate of –0.20 indicates that, compared to mid-level tones, falling tones were identified with smaller (faster) reaction times. Changing the reference to falling tones by using the list(revpairwise~) command yields the exact same output, but reverses the sign of the b-estimate and z-score or t-score. For ease of reading, we report the estimate with the sign as relevant to the comparison mentioned in the main text, which may in some cases differ from the sign mentioned in the output table.

References

Baddeley AD (2003) Working memory and language: An overview. *Journal of Communication Disorders* 36: 189–208.

Baese-Berk MM (2019) Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, and Psychophysics* 81: 981–1005.

Baese-Berk MM and Samuel AG (2016) Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language* 89: 23–36.

Bailles F, Suárez-González N, González-Fuente S, and Prieto P (2019) Ibserving and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition* 41: 33–58.

Barcroft J and Sommers MS (2014) Effects of variability in fundamental frequency on L2 vocabulary learning: A comparison between learners who do and do not speak a tone language. *Studies in Second Language Acquisition* 36: 423–49.

Bates D, Mächler M, Bolker B, and Walker S (2015) Fitting Linear Mixed-Effects models using {lme4}. *Journal of Statistical Software* 67: 1–48.

Battig WF and Montague WE (1969) Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology* 80: 1–46.

Best CT (1995) A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* 15: 167–200.

Best CT (2019) The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research. *Frontiers in Psychology* 10: 00364.

Best CT and Tyler MD (2007) Nonnative and second-language speech perception. In Munro MJ and Bohn O-S (eds) *Second language speech learning: The role of language experience in speech and production* (pp. 13–34). Amsterdam: John Benjamins.

Bidelman GM, Hutka S, and Moreno S (2013) Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS ONE* 8: 60676.

Boersma P and Weenink D (2019) *Praat: Doing phonetics by computer: 6.0.48* [computer program]. Available at: http://www.praat.org (accessed March 2022).

Bowles AR, Chang CB, and Karuzis VP (2016) Pitch ability as an aptitude for tone learning. *Language Learning* 66: 774–808.

Braun B and Johnson EK (2011) Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics* 39: 585–94.

Braun B, Galts T, and Kabak B (2014) Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research* 30: 323–50.

Brooks ME, Kristensen K, van Benthem KJ et al. (2017) {glmmTMB} balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9: 378–400.

Chan RKW and Leung JHC (2020) Why are lexical tones difficult to learn? *Studies in Second Language Acquisition* 42: 33–59.

Chang D, Hedberg N, and Wang Y (2016) Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America* 139: 2432–47.

Chang YS, Yao Y, and Huang BH (2017) Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America* 141(2).

Chao YR (1968) *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen A, Peter V, Wijnen F, Schnack H, and Burnham D (2018) Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language* 180–82: 31–41.

Chen J, Best CT, and Antoniou M (2020) Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics* 83: 101013.

Chen S, Zhu Y, and Wayland R (2017) Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PLoS ONE* 12: 180656.
Chen S, Zhu Y, Wayland R, and Yang Y (2020) How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers? *PLoS ONE* 15: e0232514.
Cheung H (1996) Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology* 32: 867–73.
Cooper A and Wang Y (2012) The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America* 131: 4756–69.
Dumay N and Gaskell MG (2007) Sleep-associated changes in the mental representation of spoken words: Research report. *Psychological Science* 18: 35–39.
Francis AL, Ciocca V, Ma L, and Fenn K (2008) Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics* 36: 268–94.
Gandour JT and Harshman RA (1978) Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech* 21: 1–33.
Goss S (2020) Exploring variation in nonnative Japanese learners' perception of lexical pitch accent: The roles of processing resources and learning context. *Applied Psycholinguistics* 41: 25–49.
Goss S and Tamaoka K (2019) Lexical accent perception in highly-proficient L2 Japanese learners: The roles of language-specific experience and domain-general resources. *Second Language Research* 35: 351–76.
Green P and MacLeod CJ (2016) <scp>SIMR</scp>: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* 7: 493–98.
Hallé PA, Chang YC, and Best CT (2004) Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 32: 395–421.
Hao Y-C (2012) Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics* 40: 269–79.
Hartig F (2020) *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models*. Available at https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html (accessed March 2022).
Hoenig JM and Heisey DM (2001) The abuse of power. *The American Statistician* 55: 19–24.
Hutka S, Bidelman GM, and Moreno S (2015) Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia* 71: 52–63.
Jongman A, Qin Z, Zhang J, and Sereno JA (2017) Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America* 142(2).
Kachlicka M, Saito K, and Tierney A (2019) Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language* 192: 15–24.
Klein D, Zatorre RJ, Milner B, and Zhao V (2001) A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage* 13: 646–53.
Kormos J and Sáfár A (2008) Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism* 11: 261–71.
Lakens D, Scheel AM, and Isager PM (2018) Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science* 1: 259–69.
Lenth R (2020) *emmeans: Estimated marginal means, aka least-squares means* [software]. Available at https://cran.r-project.org/web/packages/emmeans/index.html (accessed March 2022).

Li A and Post B (2014) L2 acquisition of prosodic properties of rhythm. *Studies in Second Language Acquisition* 36: 223–55.
Li M and Dekeyser R (2017) Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition* 39: 593–620.
Ling W and Grüter T (2020) From sounds to words: The relation between phonological and lexical processing of tone in L2 Mandarin. *Second Language Research* 38: 289–313.
Liu S and Samuel AG (2004) Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech* 47: 109–38.
Mathôt S, Schreij D, and Theeuwes J (2012) OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44: 314–24.
Nation P (2000) Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal* 9: 6–10.
O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity* 41: 673–90.
Pelzl E, Lau EF, Guo T, and DeKeyser R (2019) Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition* 41: 59–86.
Pelzl E, Lau EF, Guo T, and DeKeyser R (2020) Even in the best-case scenario L2 learners have persistent difficulty perceiving and utilizing tones in Mandarin. *Studies in Second Language* 43: 268–96.
Peng G, Zheng HY, Gong T et al. (2010) The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics* 38: 616–24.
Perrachione TK, Lee J, Ha LYY, and Wong PCM (2011) Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America* 130: 461–72.
Perrachione TK, Fedorenko EG, Vinke L, Gibson E, and Dilley LC (2013) Evidence for shared cognitive processing of pitch in music and language. *PLoS ONE* 8: 73372.
Poltrock S, Chen H, Kwok C, Cheung H, and Nazzi T (2018) Adult learning of novel words in a non-native language: Consonants, vowels, and tones. *Frontiers in Psychology* 9: 01211.
Qin Z and Jongman A (2016) Does second language experience modulate perception of tones in a third language? *Language and Speech* 59: 318–38.
Qin Z, Zhang C, and Wang WS (2021) The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America* 149: 435–46.
R Core Team (2020) *R version 4.0.1 (2020-06-06): 'See things now'* [software]. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/
Reid A, Burnham D, Kasisopa B et al. (2015) Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, and Psychophysics* 77: 571–91.
Rossion B and Pourtois G (2004) Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception* 33: 217–36.
Sadakata M, Weidema JL, and Honing H (2020) Parallel pitch processing in speech and melody: A study of the interference of musical melody on lexical pitch perception in speakers of Mandarin. *PLoS ONE* 15: e0229109.
Schaefer V and Darcy I (2014) Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology* 5: 489–522.
So CK and Best CT (2010) Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53: 273–93.
So CK and Best CT (2014) Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition* 36: 195–221.

Tang W, Xiong W, Zhang Y-X, Dong Q, and Nan Y (2016) Musical experience facilitates lexical tone processing among Mandarin speakers: Behavioral and neural evidence. *Neuropsychologia* 91: 247–53.

Tsukada K and Kondo M (2019) The perception of Mandarin lexical tones by native speakers of Burmese. *Language and Speech* 62: 625–40.

Van Overschelde JP, Rawson KA, and Dunlosky J (2004) Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language* 50: 289–335.

Wallentin M, Nielsen AH, Friis-Olivarius M, Vuust C, and Vuust P (2010) The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences* 20: 188–96.

Wang X (2013) Perception of Mandarin tones: The effect of L1 background and training. *Modern Language Journal* 97: 144–60.

Wang Y, Behne DM, Jongman A, and Sereno JA (2004) The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics* 25: 449–66.

Wayland RP and Guion SG (2004) Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54: 681–712.

Wickham H (2016) *ggplot2: Elegant graphics for data analysis* [software]. Springer. Available at: https://ggplot2.tidyverse.org (accessed March 2022).

Wiener S and Goss S (2019) Second and third language learners' sensitivity to Japanese pitch accent is additive. *Studies in Second Language Acquisition* 41: 897–910.

Wiener S, Ito K, and Speer SR (2020) Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinations. *Studies in Second Language Acquisition* 43: 155–80.

Wong PCM and Perrachione TK (2007) Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28: 565–85.

Wong PCM, Kang X, Wong KHY et al. (2020) ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances* 6: eaba5090.

Wu H, Ma X, Zhang L et al. (2015) Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Frontiers in Psychology* 6: 00436.

Wu X, Munro MJ, and Wang Y (2014) Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics* 46: 86–100.

Yang R (2015) The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics* 43: 453–72.

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

Yu K, Li L, Chen Y et al. (2019) Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology* 56: e13448.

Yu K, Zhou Y, Li L et al. (2017) The interaction between phonological information and pitch type at pre-attentive stage: An ERP study of lexical tones. *Language, Cognition and Neuroscience* 32: 1164–75.

Yu KM and Lam HW (2014) The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America* 136: 1320–33.

Zhu M, Chen X, and 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: 4214–27.

Appendix 1. Detailed participant demographics (English group) ME: Musical experience; WM: Working memory.

ID	Age	L2s and self-reported level (0–10)	Currently practicing	ME	WM
EN-MU-F-1	21	German 3	Keyboard/piano; woodwind; singing	14	47
EN-MU-F-2	19	Spanish 7, Portuguese 6	Drums; keyboard/piano; singing	14	27
EN-MU-F-3	20	–	Keyboard/piano; strings; woodwind	14	74
EN-MU-F-4	19	French 5, Spanish 4	Woodwind; choral singing	11	62
EN-MU-F-5	20	–	Guitar; choral singing; singing	12	31
EN-MU-F-6	20	Gujarati* 5, Spanish 4, French 1	Keyboard/piano; strings; choral singing	13	71
EN-MU-M-1	20	–	Strings; singing	13	61
EN-MU-M-2	20	–	Keyboard/piano; strings; brass; choral singing	16	100
EN-MU-M-3	20	–	Keyboard/piano; woodwind; choral singing	12	81
EN-MU-M-4	25	Russian* 7, French 7, German 7, Spanish 7	Keyboard/piano	19	91
EN-MU-M-5	22	Italian* 7, Spanish 4, French 1	Guitar; keyboard/piano; singing	8	62
EN-NM-F-1	19	French 7, Spanish 2	–	–	17
EN-NM-F-2	22	French 5, Hindi 3	–	2	42
EN-NM-F-3	23	Spanish 4	–	–	53
EN-NM-F-4	21	–	–	–	65
EN-NM-F-5	21	Spanish 3	–	–	29
EN-NM-M-1	21	German 7	–	5	97
EN-NM-M-2	20	Hindi* 7	–	–	44
EN-NM-M-3	23	German 6	–	2	63
EN-NM-M-4	22	–	–	–	37
EN-NM-M-5	22	–	–	–	62

Note. * exposure to a (heritage) language before the age of 12 year at home or in other surroundings.

Appendix 2. Detailed participant demographics (Mandarin group) ME: Musical experience; WM: Working memory.

ID	Age	L2s and self-reported level (0–10)	Currently practicing	ME	WM
MA-MU-F-1	20	English 8, Wu* 7, Japanese 6	Keyboard/piano; woodwind; choral singing	16	55
MA-MU-F-2	20	English 8, Italian 1, French 1	Strings	17	92
MA-MU-F-3	19	English 6	Guitar; keyboard/piano; singing; guzheng	17	62
MA-MU-F-4	29	English 8, Cantonese* 7, French 2	Guzheng	22	31
MA-MU-F-5	24	English 8, Cantonese* 7, French 1	Keyboard/piano	14	90
MA-MU-M-1	24	English 10, French 1	Erhu	18	93
MA-MU-M-2	23	English 8, Cantonese* 7	Guitar	8	52
MA-MU-M-3	28	English 8, Wu* 7, German 5, Cantonese 2	Keyboard/piano; strings; choral singing	15	31
MA-MU-M-4	19	English 8	Strings; singing	9	30
MA-MU-M-5	19	English 8, French 1	Guitar; keyboard/piano	12	91
MA-NM-F-1	29	Italian 10, English 8, Wu* 7, French 7, Japanese 2, Persian 2	–	3	87
MA-NM-F-2	23	English 8	–	2	95
MA-NM-F-3	25	English 8	–	–	41
MA-NM-F-4	22	English 8	–	–	92
MA-NM-F-5	24	English 8, Japanese 7	–	–	79
MA-NM-M-1	18	English 7	–	1	77
MA-NM-M-2	19	English 8	–	1	94
MA-NM-M-3	20	English 8	–	1	83
MA-NM-M-4	23	English 8	–	2	91
MA-NM-M-5	23	Kunming Chinese* 7, English 7	–	–	75

Note. * exposure to a (heritage) language before the age of 12 at home or in other surroundings.

Appendix 3. Tone categorization: Mixed model ANOVA table for logRT results (Type III Wald Chisquare tests).

Tone categorization
Formula: lmer(logRT ~ L1*tone*musical experience + L1*tone*working memory + (1|subject) + (1|item))

Effect	χ^2	df	*p*
L1	0.516	1	0.087
Tone	4.356	3	0.226
Musical experience	11.749	1	< 0.001
Working memory	0.266	1	0.606
L1*tone	61.021	3	0.000
L1*musical experience	5.979	1	0.014
L1*working memory	2.755	1	0.097
Tone*musical experience	6.528	3	0.089
Tone*working memory	3.417	3	0.332
L1*tone*musical experience	12.118	3	0.007
L1*tone*working memory	4.885	3	0.180

Appendix 4. Word identification: Mixed model ANOVA table for accuracy results (Type III Wald Chisquare tests).

Word identification
Formula: glmer(correct ~ L1*tone*musical experience + L1*tone*working memory + L1*tone*tone categorization + (1|subject) + (1|item))

Effect	χ^2	df	*p*
L1	1.594	1	0.207
Tone	8.876	3	0.031
Musical experience	25.181	1	0.000
Working memory	7.016	1	0.008
Tone categorization	1.644	1	0.200
L1*tone	11.107	3	0.011
L1*musical experience	10.492	1	0.001
L1*working memory	5.766	1	0.016
L1*tone categorization	0.017	1	0.896
Tone*musical experience	2.013	3	0.570
Tone*working memory	4.449	3	0.217
Tone*tone categorization	11.204	3	0.011
L1*tone*musical experience	2.013	3	0.570
L1*tone*working memory	6.302	3	0.097
L1*tone*tone categorization	1.418	3	0.701

Appendix 5. Tone categorization: Significant multiple comparisons for tone (Bonferroni-corrected).

Predictors	Estimates	Standard error	*t*	*p*
English				
(No significant comparisons)				
Mandarin				
Fall-mid	–0.20	0.06	–3.32	0.027
Fall-low	–0.18	0.06	–2.92	0.047

Note. For brevity, only significant comparisons are listed.

Appendix 6. Word identification: Significant multiple comparisons for tone (Bonferroni-corrected).

Predictors	Estimates	Standard error	*t*	*p*
English–Mandarin | Low	1.11	0.47	2.36	0.018
English				
(No significant comparisons)	–	–	–	–
Mandarin				
Rise-low	1.15	0.35	3.31	0.005
Fall-low	1.23	0.34	3.59	0.002

Note. For brevity, only significant comparisons are listed.

Appendix 7. Tone categorization: Multiple comparisons between error types.

Contrast		Estimates	Standard error	*t*	*p*
English:					
Rise-to-fall	Fall-to-mid	–1.42	0.39	–3.60	0.026
Fall-to-mid	Mid-to-rise	1.70	0.44	3.84	0.010
	Mid-to-fall	1.42	0.39	3.60	0.026
	Low-to-rise	2.80	0.73	3.85	0.010
	Low-to-fall	3.50	1.02	3.45	0.045
Mid-to-rise	Mid-to-low	–1.58	0.45	–3.51	0.035
	Low-to-mid	–1.64	0.45	–3.68	0.019
Mid-to-fall	Low-to-mid	–1.35	0.40	–3.42	0.049
Mid-to-low	Low-to-rise	2.67	0.73	3.66	0.021
Low-to-rise	Low-to-mid	–2.74	0.73	–3.76	0.014
Mandarin:					
Rise-to-mid	Low-to-mid	–2.44	0.60	–3.79	0.013
Fall-to-rise	Low-to-mid	–3.58	1.00	–3.47	0.042
Fall-to-mid	Mid-to-low	–2.59	0.80	–3.44	0.046
	Low-to-mid	–2.90	0.70	–3.87	0.009

Notes. For brevity, only significant comparisons are listed. The counts of error types were subjected to a zero-inflated general linear mixed effect model (Brooks et al., 2017), with confusion type (12 levels: Rise-to-fall, rise-to-mid, etc.) as fixed factor, and subject as a random intercept. Because not all models would converge on the full data sets, the models were fitted on data subsets per group. glmmTMB(count ~ errortype + (1|subject), ziformula=~1, family=poisson)

Appendix 8. Word identification: Multiple comparisons between tone-only error types.

Contrast		Estimates	Standard error	*t*	*p*
English:					
Rise-to-mid	Rise-to-low	1.34	0.36	3.76	0.014
	Fall-to-low	1.09	0.31	3.51	0.035
Rise-to-low	Fall-to-mid	−1.72	0.34	−5.04	0.001
	Mid-to-fall	−1.26	0.35	−3.56	0.029
	Low-to-mid	−1.51	0.34	−4.40	0.001
Fall-to-rise	Fall-to-mid	−1.03	0.25	−4.15	0.003
Fall-to-mid	Fall-to-low	1.47	0.29	5.06	< 0.001
	Mid-to-rise	1.01	0.24	4.13	0.003
	Low-to-fall	1.45	0.31	4.64	< 0.001
Fall-to-low	Low-to-mid	−1.26	0.29	−4.31	0.002
Low-to-fall	Low-to-mid	−1.25	0.32	−3.94	0.007
Mandarin:					
Rise-to-Fall	Low-to-mid	−1.32	0.32	−4.17	0.003
Rise-to-Mid	Low-to-mid	−1.38	0.38	−3.63	0.023
Rise-to-Low	Low-to-mid	−1.54	0.34	−4.51	0.001
Fall-to-Rise	Mid-to-low	−2.11	0.55	−3.86	0.009
	Low-to-mid	−2.76	0.54	−5.13	< 0.001
Fall-to-mid	Low-to-mid	−1.69	0.37	−4.63	< 0.001
Fall-to-low	Low-to-mid	−1.08	0.28	−3.86	0.009
Mid-to-rise	Low-to-mid	−2.03	0.41	−4.93	0.001
Mid-to-fall	Low-to-mid	−2.10	0.45	−4.69	< 0.001
Mid-to-low	Low-to-fall	2.06	0.56	3.70	0.018
Low-to-rise	Low-to-mid	−2.04	0.41	−5.03	< 0.001
Low-to-fall	Low-to-mid	−2.71	0.55	−4.95	< 0.001

Notes. For brevity, only significant comparisons are listed. The counts of error types were subjected to a zero-inflated general linear mixed effect model (Brooks et al., 2017), with *confusion type* (12 levels: Rise-to-fall, rise-to-mid, etc.) as fixed factor, and *subject* as a random intercept. Because not all models would converge on the full data sets, the models were fitted on data subsets per group. glmmTMB(count ~ errortype + (1|subject), ziformula=~1, family=poisson).

Appendix 9. Tone categorization: Multiple comparisons and estimates per L1 for extralinguistic factors.

Predictors	Estimate	Standard error	*t*	*p*	95% CI
English–Mandarin | Musical experience	−0.23	0.10	−2.26	0.028	–
English–Mandarin | Working memory	0.16	0.11	1.53	0.131	–
English:					
Musical experience	−0.28	0.08	−3.58	0.002	[−0.43, −0.12]
Working memory	0.11	0.08	1.40	0.307	[−0.05, 0.26]
Mandarin:					
Musical experience	−0.05	0.07	−0.70	0.979	[−0.18, 0.09]
Working memory	−0.06	0.08	−0.76	0.699	[−0.21, 0.09]

Appendix 10. Tone Categorization: Estimates of musical experience per L1 and per tone.

Predictors	Estimates	Standard error	*t*	*p*	95% CI
English–Mandarin (rise)	–0.25	0.11	–2.41	0.019	–
English–Mandarin (fall)	–0.31	0.11	–2.91	0.005	–
English–Mandarin (mid)	–0.17	0.11	–1.64	0.106	–
English–Mandarin (low)	–0.19	0.10	–1.80	0.076	–
English:					
Rise	–0.29	0.08	–3.60	0.001	[–0.45, –0.13]
Fall	–0.33	0.08	–4.14	< 0.001	[–0.50, –0.17]
Mid	–0.26	0.08	–3.29	0.002	[–0.42, –0.10]
Low	–0.23	0.08	–2.89	0.007	[–0.39, –0.07]
Mandarin:					
Rise	–0.03	0.07	–0.50	0.616	[–0.17, 0.10]
Fall	–0.03	0.07	–0.37	0.711	[–0.16, 0.11]
Mid	–0.09	0.07	–1.31	0.196	[–0.23, 0.05]
Low	–0.04	0.07	–0.52	0.601	[–0.17, 0.10]

Appendix 11. Word identification: Multiple comparisons and estimates per L1 for extralinguistic factors.

Predictors	Estimate	Standard error	z	*p*	95% CI
English–Mandarin | Musical experience	1.73	0.53	3.24	0.001	–
English–Mandarin | Working memory	–1.13	0.47	–2.40	0.016	–
English–Mandarin | Tone categorization	–0.06	0.48	–0.13	0.896	–
English:					
Musical experience	2.21	0.45	4.90	< 0.001	[1.32, 3.09]
Working memory	0.06	0.35	0.17	0.982	[–0.63, 0.75]
Tone categorization	–0.34	0.29	–1.17	0.424	[–0.91, 0.23]
Mandarin:					
Musical experience	0.48	0.28	1.66	0.193	[–0.09, 1.04]
Working memory	1.19	0.31	3.79	< 0.001	[0.58, 1.81]
Tone categorization	–0.28	0.38	–0.72	0.720	[–1.03, 0.47]