1-1-2020

Applying a Newly Learned Second Language Dimension to the Unknown: The Influence of Second Language Mandarin Tones on the Naïve Perception of Thai Tones

Vance Schaefer
University of Mississippi

Isabelle Darcy
TRUSTEES OF INDIANA UNIVERSITY

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

Psychology of Language and Communication 2020, Vol. 24, No. 1

sciendo

DOI: 10.2478/plc-2020-0007

VANCE SCHAEFER[1], ISABELLE DARCY[2]
[1] The University of Mississippi, United States
[2] Indiana University, United States

APPLYING A NEWLY LEARNED SECOND LANGUAGE DIMENSION TO THE UNKNOWN: THE INFLUENCE OF SECOND LANGUAGE MANDARIN TONES ON THE NAÏVE PERCEPTION OF THAI TONES

This study investigates whether L2 Mandarin learners can generalize experience with Mandarin tones to unfamiliar tones (i.e., Thai). Three language groups – L1 English/ L2 Mandarin learners (n=18), L1 Mandarin speakers (n=30), L1 monolingual English speakers (n=23) – were tested on the perception of unfamiliar Thai tones on ABX tasks. L2 Mandarin learners and L1 Mandarin speakers perceived Thai tones more accurately than L1 English non-learners. Mandarin learners L1 speakers showed priming on Mandarin tones on a lexical decision task with repetition priming, suggesting L2 tones had been encoded within lexical representations of L2 Mandarin words. However, results must be interpreted cautiously, with an absence of expected priming and presence of unexpected priming. In sum, learners can transfer L2 tone experience to unfamiliar tones, expanding the Feature Hypothesis (McAllister, Flege, & Piske, 2002) to include L2 influence as well. In addition, results indicate a potential disconnect between perception and encoding.

Keywords: tone, third language, perception, phonolexical encoding

Address for correspondence: Vance Schaefer, Department of Modern Languages, The University of Mississippi, P.O. Box 1848, University, MS 38677 USA
Email: schaefer@olemiss.edu, idarcy@indiana.edu

Native (L1) speakers of a tone language generally perceive another tone language more easily than L1 speakers of a non-tonal language (Gandour & Harshman, 1978; Wayland & Guion, 2004). L1 speakers of a tone language are better able to attend to pitch height and/or direction and/or map non-native tones onto their L1 tones, thereby outperforming L1 speakers of non-tone languages in tone identification tasks (Wayland & Guion, 2004). Native speakers perceive tones as linguistic categories (Van Lancker & Fromkin, 1973), and tonal information constrains lexical access (Lee, 2007). By contrast, speakers of non-tonal languages are not all equal in their sensitivity to non-native tone perception (Burnham et al., 1996; Schaefer & Darcy, 2014; So & Best, 2010) as languages differ in the extent and function to which they use pitch to distinguish words. Naive listeners with various L1s differ in the accuracy of identifying tones due to varying ability to attend to cues as shaped by their L1 (Francis, Ciocca, Ma, & Fenn, 2008; Gandour, 1983). However, for L1 speakers of a non-tonal language, experience (i.e., training) with a tone language improves perception (Francis et al., 2008) and production of L2 tones (Hao, 2012).

Going one step further, the current study investigates whether L2 learners of a tone language can generalize their experience with L2 tones. Specifically, the study compares L2 learners of Mandarin (with L1 English) and L1 speakers of Mandarin on the naive perception (i.e., non-learner) of another tone language: Thai. Furthermore, this study compares these two groups on the phonolexical encoding of Mandarin tones to illuminate the relationship between phonetic perception and phonolexical encoding where perception does not necessarily precede encoding (Darcy et al., 2012).

Perception of Tone

Pitch height and direction are accessed in perceiving tones (Gandour, 1983; Wang, Jongman, & Sereno, 2006). Listeners gauge whether the tone is high, mid, or low in a speaker's voice range (i.e., level tones) or whether the tone goes up or down or in both directions (i.e., contour tones). Languages weight pitch height over pitch direction (Gandour, 1983). Also, speakers of tone languages use direction much more than the speakers of English, a non-tone language, although to varying degrees as pitch height and direction are weighted relative to the usage needed to distinguish lexical pitch (e.g., level tones, contour tones, word stress).

Lexical pitch patterns in English, Mandarin, and Thai vary by pitch height, direction, and length. Stimuli with a similar sequence of segments (i.e., [ma]) plotted in the following three figures allow a focused, clear comparison of pitch duration in milliseconds and pitch movement (i.e., contour) in hertz. The stimuli were produced in isolation by a female speaker of each of the languages, although the two English stimuli were spliced from the word "mama." English word stress exhibits different pitch patterns in stressed versus unstressed syllables (Figure 1). Mandarin has one level tone, three contour tones, and a neutral tone (Figure 2) whereas Thai has three level tones and two contour tones (Figure 3).

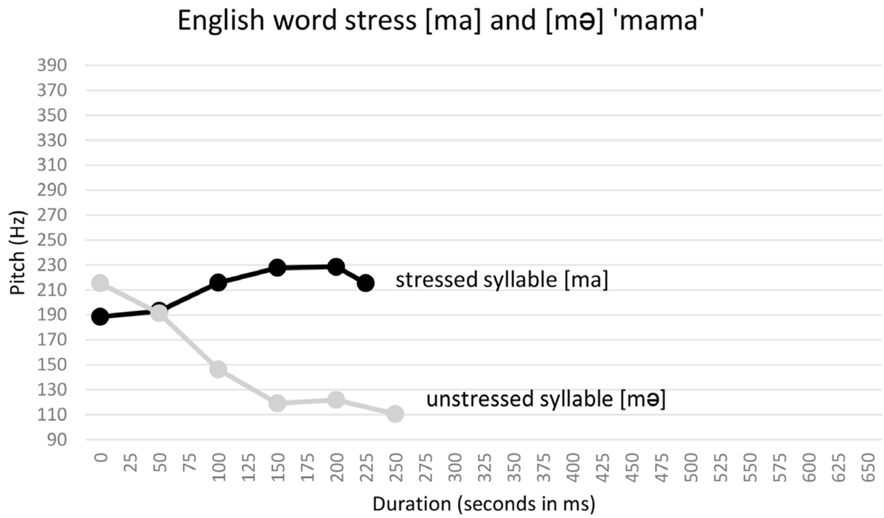

Figure 1 English word stress

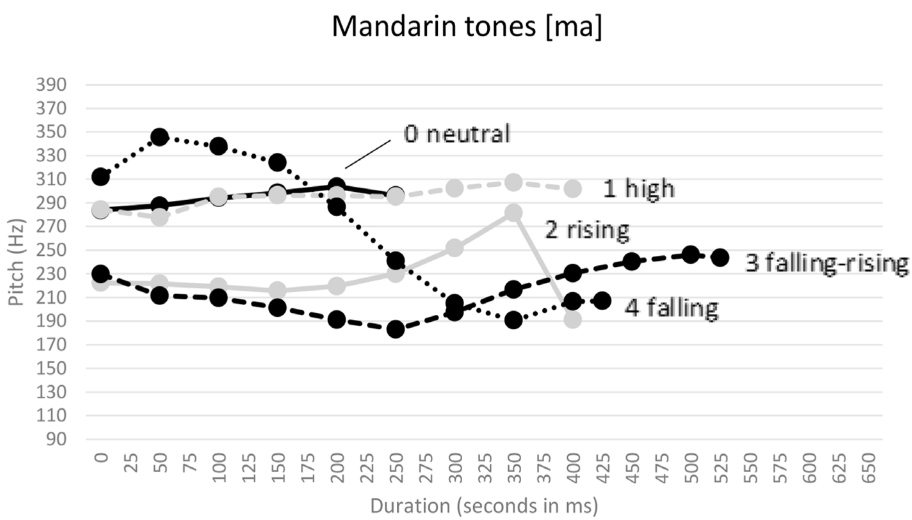

Figure 2 Mandarin tones

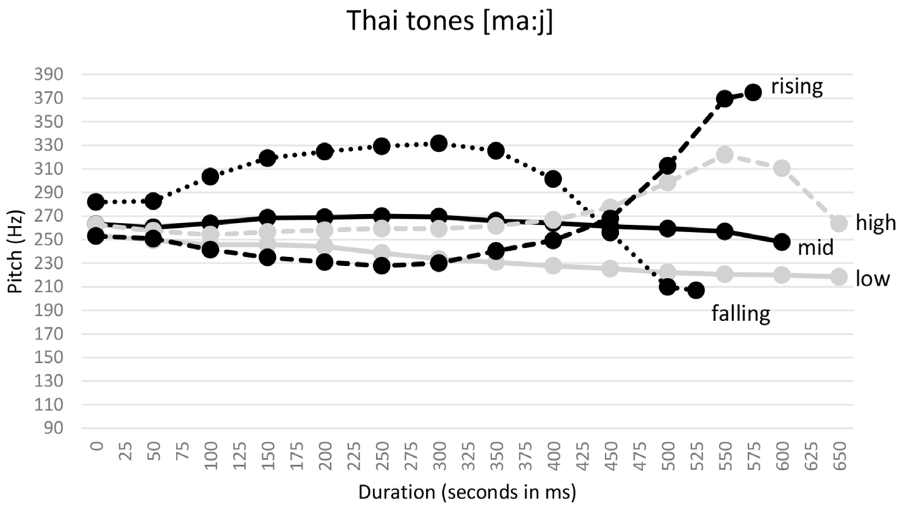

Figure 3 Thai tones

L1 and L2 speakers confuse tones with similar directions or registers, i.e., low vs. high half of the voice range (Leather, 1990) while easily perceiving two tones differing by direction or registers (Abramson, 1975; Burnham, Kirkwood, Luksaneeyanawin, & Pansottee, 1992). L1 Thai listeners (and L1 English speakers) distinguish level from contour tones (e.g., high level vs. falling tones) with the most difficulty but two contour tones with the most ease (Burnham et al., 1992). L1 Thai speakers confuse the low- and mid-level tones (Abramson, 1976), demonstrating difficulty in perceiving the height difference, but do not confuse mid- and high-level tones as the height difference is larger or the shape (i.e., direction) differs, which is indeed the case phonetically.

L1 tone language speakers map L2 tones onto similar L1 tones (So & Best, 2010). They also apply their experience in tracking pitch height and direction (Wayland & Guion, 2004), outperforming L1 non-tone language speakers in L2 tone perception. Nevertheless, L1 tone language speakers confuse L2 tones perceived to be similar to L1 tones (Gandour, 1983; So and Best, 2010; Wayland & Guion, 2004) so that similar L2 tones are mapped onto L1 tone categories, e.g., L2 Mandarin tone 1 [55] and 4 [51] onto L1 Cantonese tone 1 [55] which has an allotone of 53 (So & Best, 2010). Indeed, L1 speakers of a tone language do not necessarily outperform L1 speakers of a non-tone language in the perception/production of L2 tones with similar rates of accuracy and errors (i.e., tone 2 vs. tone 3) (Hao, 2012), but L2 experience with tones does help L1 non-tone language speakers improve discrimination of tone (Hao, 2018).

One possible explanation for the role played by experience (via L1 or L2) in facilitating the discrimination, identification or processing of non-native tones is that experience makes the relevant phonological dimension available for the perceptual system. According to the Feature Hypothesis (McAllister et al., 2002), the extent and manner that a phonological feature or dimension is used in the L1 to contrast lexical items (i.e., functional load) determines the extent that the same feature is perceived and acquired in an L2. The evidence supporting this hypothesis is the hierarchy of perception of L2 Swedish vowel length by L2 learners, relative to the varying degrees of contrastive vowel length in their respective L1s: The groups for whom vowel length played a functionally more important role in the L1 were able to perceive and use vowel length to a greater extent to distinguish words in their L2. Thus, the perception of non-native tones also appears to be shaped by the extent and manner that the pitch is used to contrast L1 words (e.g., tone, pitch accent, word stress, none), independently of the fact that all languages possess the feature of intonation. The perception of L2 lexically-contrastive pitch has been shown to correlate to the functionality of lexically-contrastive pitch in the L1. Accordingly, pitch accent language speakers (e.g., Japanese) perceive non-native tones at similar rates of accuracy as tone language speakers (Burnham et al., 1996) and improve at greater degrees than word stress language speakers (e.g., English) when learning L2 tones (McGinnis, 1997). Indeed, a hierarchy of perceptual accuracy emerges among four language types in the naive perception of non-native tones (i.e., Thai tones), from most accurate to least accurate: tone languages (L1 Mandarin) > pitch accent languages (L1 Japanese pitch accent) > word stress languages (L1 English word stress) > languages without lexically-contrastive pitch (L1 standard Korean), which led to the proposal of the Functional Pitch Hypothesis (Schaefer & Darcy, 2014)

Experience–as mentioned above–is not necessarily restricted to L1 experience. An acquired L2 may also influence the naive perception of another, a third language. Understanding the influence of a L2 on other languages is an area of growing interest, but research in the area of phonology is still scarce. Typological distance, i.e., perceived similarity**,** increases L2 influence over L1 in L3 acquisition (Wrembel, 2010). However, strong L2 influence (positive or negative) is expected only if the level of L2 is fairly proficient, i.e., intermediate level (Fernandes-Boëchat, 2007). Additionally, beginning L3 learners can perceive the subtle phonetic differences between contrasts and may map these L3 sounds onto both L1 and L2 categories, with a partiality to the L2 (Wrembel, Marecka, & Kopečková, 2019). As for lexical pitch, an L2 tone language (Mandarin) may benefit the naive perception of non-native tones (Cantonese) (Qin & Jongman, 2016). Specifically, L2 experience augments L1 experience (English); however, L1 tones (Mandarin) may not always facilitate the perception of unfamiliar tones as the phonetic shape of some tones make them difficult to perceive, and similar L2 tones influence non-native tone perception. Like Qin and Jongman (2016), Wiener and Goss (2019) appear to build on the Feature Hypothesis (McAllister et al., 2002; Schaefer & Darcy, 2014) by determining that L2 experience with tones is additive. Listeners to Japanese pitch accent perform in a hierarchy of perception commensurate to the functional load of lexical pitch in their L1 and/or L2 (least to most accurate): L1 English < L1 English/L2 Japanese < L1 English/L2 Mandarin < L1 Japanese.

Thus, we expect speakers of a language lower in the hierarchy of perceptual accuracy for lexically-contrastive pitch (e.g., English) who learn an L2 language higher in the hierarchy (e.g., Mandarin) to move up the perceptual hierarchy, thanks to L1 lexical pitch augmented by L2 lexical pitch. Proficiency level in the L2 would be determined by the robustness of lexical encoding of the L2 tones in lexical representations, which was not determined in the Qin and Jongman (2016) or Wiener and Goss (2019) study.

Lexical Encoding of Tone

Like segmentals, suprasegmentals are encoded lexically (e.g., vowel length in Japanese, Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; stress in Spanish, Dupoux, Pallier, Sebastian, & Mehler, 1997). Likewise, tones are encoded lexically and can be used to constrain word recognition in the L1. Minimal tone pairs in Mandarin such as *tang* [t^{h}aŋ]: tone 1 'soup' 湯; tone 2 'candy' 糖 are both activated initially and are only disambiguated soon after once tonal information is accessed (Lee, 2007). L1 Mandarin speakers have difficulty ignoring tonal information to the extent that they slow down when categorizing segments that are accompanied by within-category changes to the pitch height of tone (Lee & Nusbaum, 1993). However, L1 tone language speakers show more difficulty in telling non-words from words on a lexical decision task when items differ in tone only than when they differ in segments only (Cutler & Chen, 1997), indicating perhaps a dominance of segmental over tonal information. Furthermore, studies using repetition priming (i.e., faster reaction time upon hearing a word for the second time) revealed differences in the robustness of the encoding of tone versus the encoding of segments as part of the phonological representation of a word in Mandarin: As evidenced by reaction times (from faster to slower), same segment/same tone (i.e., same word) combinations produced faster reactions than same segment/different tone combinations, which in turn were faster than different segment/same tone (Lee, 2007).

Interestingly, word stress in English words appears to be lexically encoded too, but less strongly so than segments. Stimuli pairs differing only in word stress (i.e., 'an insert' vs. 'to insert') activate the two different words to a limited extent due to a strong segmental influence on lexical activation. Of the correlates of stress (vowel quality, vowel length, intensity, pitch), intensity and vowel duration appear to trigger the perception of word stress more strongly (Kochanski, Grabe, Coleman, & Rosner, 2005). Indeed, pitch alone does not constrain lexical activation to the extent of segments or vowel quality and length (Cooper, Cutler, & Wales, 2002).

Learners of a tone language are able to encode L2 tone lexically, but this ability might depend on the functionality of lexical pitch in the L1. The few studies on lexical encoding of L2 tone show that L1 speakers of a language featuring lexical stress can lexicalize L2 tone more efficiently than L1 speakers of a language without lexical stress, tone, or pitch accent (Braun, Galts, & Kabak, 2014). Additionally, advanced learners of a tonal L2 encode tone at the level of the syllable, but the robustness of encoding tone drops in comparison to L1 tone speakers at a more complex level, i.e., disyllabic non-words in isolation or sentences (Pelzl, Lau, Guo, & DeKeyser, 2018). L1 speakers and L2 learners of Thai appear to differ in their encoding of tonal categories with natives having seemingly more robust long-term memory representations (Wayland & Guion, 2003).

Accurate perceptual categorization of minimal pair distinctions is assumed to be a necessary condition to the encoding of phonological dimensions within lexical representations (Flege, 1993). However, L2 lexical encoding of difficult phonological dimensions can occur *independently* of accurate perceptual categorization of these same dimensions. This surprising disconnect between perception and lexical encoding of L2 phonological representations has been shown in learners who could not accurately perceive the minimal pair distinction in question, but were successful at keeping lexical entries (differing in that dimension) separated. This conclusion is based on finding an asymmetry in the activation of words differing in the /æ/-/ɛ/ English minimal segmental pair by L1 Dutch speakers (Weber & Cutler, 2004). The eye-tracking experiment displayed a target and a competitor whose first syllable could be perceptually confusing for Dutch listeners (e.g., “pencil” [pɛn] vs. “panda” [pæn]). Assuming a difficulty in perceiving and encoding this vowel difference, looks to either target or competitor during that first syllable were expected to be symmetrical and to both items equally because /æ/ and /ɛ/ would activate both words. However, in Dutch, the /ɛ/ vowel category is dominant and likely to be clearly represented, whereas the other category may be represented as an ‘odd’ version of that /ɛ/ vowel; indeed, when hearing /æ/-syllables, listeners looked at both target (with /æ/) and competitor (with /ɛ/). However, when hearing /ɛ/-syllables, listeners’ looks were more specific and they looked longer at the target (with /ɛ/) than the competitor containing /æ/. This asymmetry indicates that the distinction is at least partially encoded despite perceptual issues.

This instance has further bolstered findings on the perception and encoding of the French front-back round vowels of /y/-/u/ and /œ/-/ɔ/. In an ABX categorization task, advanced learners with L1 English did not outperform intermediate learners, but behaved like native speakers in encoding on a lexical decision task with repetition priming. These results suggest a disconnect between phonetic perception and lexical encoding (Darcy et al., 2012). A similar but less robust disconnect using the same two tasks was found for advanced learners of

Japanese in their perception and encoding of a suprasegmental dimension, namely geminates (i.e., long consonants) versus singletons (Kojima & Darcy, 2014). Together, these results suggest that learners have lexically encoded something in their phonological representation which may not be an entirely native-like representation of the targeted segment or suprasegmental. In short, such learners display partial phonetic deafness but not lexical deafness by adopting a fuzzy lexical representation that enables them to distinguish the L2 contrasts lexically even if not (yet completely) in perception. Similar disconnects are found in other domains. For example, native-like L2 production can occur without native-like L2 perception via training (Goto, 1971). Thus, the relationship between phonetic categorization and phonological representation remains unclear where experience with tone as an L2 learner does not necessarily equal robust encoding.

In response, the current study uses similar methods as previously cited studies (Darcy et al., 2012; Kojima & Darcy, 2014; Pelzl et al., 2018; Schaefer, 2015), namely ABX tasks and lexical decision task with repetition priming, to investigate whether lexically encoding a suprasegmental dimension such as tone in the L2 can potentially transfer to perception of a similar dimension in a different (unknown) language, and to shed further light on the connection between perception and lexical encoding of tones. Finding that lexically encoding tones in L2 indeed relates to accuracy of tone perception in another language would support but also expand the Feature Hypothesis (McAllister et al., 2002) through establishing the connection between perception and encoding in the application of L2 tones (i.e., Mandarin) to the naive perception of the tones of another unknown language (i.e., Thai).

Research Questions

This study asks three research questions:

ABX Task (Thai Tones) Measured by Accuracy Rates and Reaction Times

RQ1: Does the learning of a tonal L2 expand the extent and manner of lexical pitch in the learners' phonology and thus, aid in the perception of the tones of another, unfamiliar tone language (as compared to L1 English speakers with no tonal experience and L1 Mandarin speakers)? That is, can L2 learners of Mandarin perceive Thai tones more accurately and quickly than monolingual L1 speakers of English and at levels comparable to L1 speakers of Mandarin?

Lexical Decision Task with Priming (Mandarin Tones) Measured by Reaction Times

RQ2: Have the L2 learners in Mandarin with L1 English expanded the extent and manner of lexical pitch in their phonology and thus, allowing them to lexically encode tonal information to constrain lexical access to the same extent and manner as L1 Mandarin speakers? That is, do L2 learners react more quickly upon hearing the same word with the same tone (i.e., presence of repetition priming)? More importantly, do L2 learners *not* react more quickly upon hearing a minimal pair differing by tone (i.e., absence of spurious priming)?

RQ3: Does robust encoding of tones depend on accurate perception? That is, do L2 Mandarin learners both perform more accurately on the perception of Thai tones *and* prime to a certain extent on the same Mandarin words?

We predict the following results:

RQ1. If L2 experience with tones compares to L1 experience with tones and benefits non-native tone perception, L2 Mandarin learners should perform more accurately and quickly than the English monolinguals and in the same or nearly the same manner and extent as the Mandarin natives.

RQ2. If L2 experience with tones is of a nature that tonal information can be encoded in lexical representations, and can constrain lexical access, L2 Mandarin learners should perform similarly to Mandarin natives and react more quickly on the second hearing of the same word (i.e., same tone, same segments) but not on hearing the second item of a minimal pair differing by tone (Lexical Decision task with Repetition Priming).

RQ3. If the robustness of lexical encoding of tones (indexed by the *absence* of spurious priming on tonal minimal pairs) is connected to the ability to use this feature in the perception of non-native tones, the magnitude of priming on the tonal minimal pairs should be inversely correlated with ABX accuracy scores (i.e., accurate perception of Thai tones).

Method

Three computer-based tasks were conducted: Thai ABX monosyllabic task, Thai ABX disyllabic task, and Mandarin Lexical Decision task with repetition priming, along with a pronunciation proficiency test.

Participants

Participants included three groups: L1 Mandarin, L1 English who have learned Mandarin (henceforth, Learners), and L1 English who have not learned any tone language (henceforth, English). There were in total 78 participants: 31 L1 Mandarin (female = 25), 23 Learners (female = 10), and 24 L1 English (female = 14). Ages ranged as follows – Mandarin: 18-37 (M = 26.1; SD = 6.3); Learners: 18-53 (M = 24.4; SD = 9.9); and English: 18-50 (M = 31.2; SD = 9.1). Learners had completed at least second-year Mandarin Chinese.

Thai ABX Monosyllabic Task: Stimuli and Conditions

There were two types of stimuli: Target and Control. The target non-word stimuli varied by tone [low (L), mid (M), high (H), rising (R), and falling (F)] but not segment while the control non-word stimuli varied by segment (either a vowel or consonant) but not tone. Stimuli with open-syllable CV structures were used to add cognitive load since tones appear to be more difficult to discriminate in open syllables versus closed stimulus (Wayland and Guion, 2003). Each target tone stimulus featured one of two strings of segments ([no:j], [phuəj]) which, in turn, carried each of the five Thai tones, paired to form ten tone comparisons: F-R, H-F, H-R, L-F, L-H, L-M, L-R, M-F, M-H and M-R. This resulted in 20 tonal pairs (10 with segmental stimulus one, 10 with segmental stimulus two). The control segmental stimulus also carried each of the five tones (both pair members carried the same tone), featuring four pairs of segments ([be:w]~[te:w], [wi:ə]~[thi:ə], [uə]~[iə], [ri:ə]~[rɤ:j]) for a total of 20 pairs. Two female voices of the Central Thai dialect were used for the stimuli (one voice for A and B and another for X), and each repeated stimulus was a different token.

Stimuli were arranged in four triplets: ABA, ABB, BAB, BAA, with the third token as X. The use of all four orderings ensures that a recency or primacy effect (bias to B or A, respectively) does not exert undue influence on the response patterns. This resulted in 40 pairs x 4 orders for 160 experimental trials which were randomly presented. An interstimulus interval of 500ms was inserted between each stimulus within a triplet. A break was inserted every 40 trials for four blocks. Response time-out was set at 2500ms to speed up the task. Reaction times (RTs) were measured from the start of the X token. Participants were given 10 trials in a training session with feedback, featuring only segmental comparisons.

Thai ABX Disyllabic Task: Stimuli and Conditions

The disyllabic task reflected the monosyllabic task. Four disyllabic patterns of LM, LH, MH, HL were selected for fire comparisons: LH-HL, LH-MH, LM-HL, LM-LH, LM-MH, reflecting English pitch patterns and considering, 1) direction: rising/falling pitches (e.g., LH vs. HL), 2) magnitude in pitch change (e.g., LM vs. LH), and 3) different registers (i.e., LM vs. MH). The tone condition contained 10 pairings: two stimuli of [duə.phu:j] and [kiŋ.kɛ:] x five comparisons.

The disyllabic control segment stimuli used the same patterns of HL, LH, LM and MH while adding a fifth pattern, HF. Again, there were 10 pairings (two pairs of [phu:j.wu:j]~ [ru:j.wu:j] and [te:w.fiːŋ]~ [te:w.foːŋ] x five comparisons). In total, there were 20 pairings x four orders for 80 trials. A training session of 10 trials compared only segments using HF, HL, LH, LM, and MH.

Mandarin Lexical Decision Task with Repetition Priming

This task has two components: The participant 1) decides whether a stimulus is a real Mandarin word (yes/no accuracy rates), 2) reacts to two conditions: *repetition* and *minimal pair* (RTs). For *repetition*, a stimulus with the same segment and tone is presented twice in a list of stimuli. Each repetition pair consists of two different recordings of the same word by the same speaker. The second occurrence appears within 8~20 words from the first occurrence. For real words, native speakers and learners should react more quickly upon hearing the same word for the second time (i.e., repetition priming, Zwitserlood, 1996). For *minimal pair*, two stimuli differing either by tone or segment are presented. A minimal pair varying by tone only (e.g., 'mother' [ma] tone 1 vs. 'horse' [ma] tone 3) should not prime native listeners or learners as these are different words. By contrast, if a learner does not encode tonal information for a robust lexical representation of this minimal pair, they may react more quickly on hearing the second word of the pair as they consider both to be 'the same word' (i.e., spurious priming).

Four types of stimuli were created: real-word pairs differing by tone, real-word pairs differing by segments, non-word pairs differing by tone, and non-word pairs differing by segments as show in Table 1. For each type, eight pairs of stimuli were selected, using only tones 1 and 2 to limit the number of trials. Additionally, tones 1 and 2 were chosen because of their comparable lengths as tone 3 is longer and tone 4 is shorter. Also, tones 1 and 2 are spoken in different registers

Table 1 Sample of Mandarin Stimuli Used in the Lexical Decision Task with Repetition Priming

	Test items (tones)			Control items (segments)		
	IPA	**Tone**	**Gloss, character, pinyin**	**IPA**	**Tone**	**Gloss, character, pinyin**
Real words	[tʰaŋ]	1	'soup' 湯, *tang*	[dʒja]	1	'house, home' 家, *jia*
	[tʰaŋ]	2	'candy' 糖, *tang*	[dʒjɛ]	1	'to pick up (the phone)' 接, *jie*
	[tʰjɛn]	1	'heaven, sky' 天, *tian*	[tʃʰəŋ]	2	'to become' 成, *cheng*
	[tʰjɛn]	2	'to be sweet' 甜, *tian*	[tʃʰaŋ]	2	'to be long' 長, *chang*
	[xə]	1	'to drink' 喝, *he*	[ʃau]	3	'to be few' 少, *shao*
	[xə]	2	'and' 和 or 'river'河, *he*	[dʒau]	3	'to search for' 找, *zhao*
Non-words	[hiŋ]	1	*hing*	[sə]	1	*se*
	[hiŋ]	2	*hing*	[so:]	1	*sou*
	[nue]	1	*nui*	[siɛn]	2	*sian*
	[nue]	2	*nui*	[sən]	2	*sen*
	[gi:]	1	*gii*	[bua]	3	*bua*
	[gi:]	2	*gii*	[buɛ]	3	*bue*

(i.e., upper and lower, respectively), facilitating perceptual discrimination. Pairs of non-words carrying tone 1 and 2 were created that sounded like but were not actual words in standard Mandarin.

For the control pairs (differing by segment but having the same tone), eight minimal pairs of words and non-words were created using all four possible tones (i.e., tone 4 = 4 pairs, tone 3 = 2, tone 2 = 1, tone 1 = 1). Four pairs differed by vowel while four pairs differed by the onset consonant. As such, for *repetition*, both tone and control conditions are the same (i.e., exact same word), but not for *minimal pair* (differing by segment and not tone). In total, there were 64 stimuli: 16 real word tone/test items, 16 real word segment/control items, 16 non-word tone/test items, and 16 non-word segment/control items.

There were also 64 distracters: 15 monosyllabic real words, 17 disyllabic real words, 16 monosyllabic non-words, and 16 disyllabic non-words. Half the distracters were repeated to reflect the *repetition/minimal pairs*, resulting in 96 distracters. In total, 160 items were presented. The real words appeared mostly on the beginning list and some on the elementary list of the Chinese Proficiency Test (Hànyǔ Shuǐpíng Kǎoshì). Exposure to the four tones was balanced throughout the task.

Each pair was inserted into a 160-word list. The second item of a pair was manually inserted at a random distance of 8 to 20 items after the first item (cf. Pallier, Colomé, & Sebastián-Gallés, 2001). To counterbalance the pairings across the two conditions (*repetition*: AA, BB; *minimal pair*: AB, BA), four lists were created. Each minimal pair was presented in one of four orders: AB, AA, BA, or BB. One member of the minimal pair was presented as the first word (i.e., A) and then, followed by the other member (i.e., B) or another token of itself (i.e., A), resulting in the AB, AA orders. This was reversed for the other member of the minimal pair (i.e., BA, BB orders). Each list contained just one of the four orders for any pair, preventing bias due to the ordering of stimuli and/or nature of stimuli. Approximately 7-8 participants from the Mandarin group (n = 31) and 5-6 participants from the Learners (n = 23) were tested on each list. Each list was divided into four blocks, creating three breaks. Additionally, to allow use of the dominant hand, versions were created for left- or right-handed participants for each list. A time-out of 2500ms was set for a response to each item (i.e., push computer keyboard keys to answer whether stimuli were real words or not). Also, participants took a training session consisting of 12 items with feedback: CORRECT or INCORRECT.

Pronunciation Proficiency Task

Learners of Mandarin were also recorded on three tasks: 1) answering simple questions as *What did you do last weekend?* 2) counting from 1-20, and 3) reading a short simple passage written in both Chinese characters (simplified and traditional) and *pinyin* (phonetic alphabet). Three samples were spliced from each task from each participant: 1) approximately seven seconds from the first task, 2) counting from 1-10, and 3) reading of the last two sentences. The samples were then rated on a Likert scale from 1 (poor) to 7 (native-like) by three native speakers of Mandarin who were also PhD graduate students: one Taiwanese male studying phonetics and L2 phonology, one PRC (Xi'an) female studying L2 pragmatics, and one Taiwanese female studying phonetics and psycholinguistics. Any pauses beyond 400ms were cut down to 400ms to prevent raters from conflating fluency with pronunciation. Samples were also randomized. Raters were told to consider vowels, consonants, and tones, but to ignore fluency, grammar, and odd meanings. They could replay samples.

Procedure

Tasks were done in the following order for approximately 60 minutes: 1) explanation and signing of consent form, 2) Thai ABX monosyllabic task, 3) language background questionnaire, 4) Thai ABX disyllabic task, 5) Mandarin lexical decision task, 6) pronunciation proficiency task, and 7) debriefing (i.e., asking for participants' comments). See Schaefer, 2015 for complete details.

Results

ABX Tasks

Data for both ABX tasks were treated in the same manner. Two outliers whose performance in the control condition was beyond two SD from the mean of their group were removed. Five Learners were removed due to long-term exposure to a tone language since birth. Furthermore, RTs were examined to check for very fast answers (under 300ms) which may indicate a delayed response to a preceding item, but no datapoint was removed. A log-transformation was applied to RTs (Log-RT) to obtain a normal distribution as RTs were skewed.

Accuracy rates and RTs for overall conditions (test vs. control) and individual tone/segment sub conditions for both ABX tasks were treated as follows. A Generalized Estimating Equation (GEE) model for binary responses (1/0) was fitted to the accuracy data. A Linear Mixed Effects model was fitted to the LogRT data. The independent variables Language (Mandarin, Learners, English) and Condition (test vs. control) were declared as fixed factors, as well as the variable Subcondition (individual tonal/segment combinations) for more in-depth analysis; Subjects were declared as a random factor for both GEE and Linear Mixed Effects models. Sidak correction for multiple comparisons was used over Bonferroni. The un-aggregated data were used for statistical analysis.

Thai ABX Monosyllabic Task
Analysis by Condition

The Type III tests of fixed effects, which is an omnibus F statistic test for main effects and interactions and which examines the significance of one effect with the other effects, exhibited no effect of Language ($\chi2(2) = 4.15$, $p = .126$), a significant effect of Condition ($\chi2(1) = 151.31$, $p < .001$), and no significant interaction between the two factors ($\chi2(2) = 4.39$, $p = .111$). That is, there was no between-group difference in accuracy rates for either test, $\chi2(2) = 4.26$, $p = .119$, or control condition, $\chi2(2) = 4.86$, $p = .088$, but, overall, accuracy was lower for the test than for the control condition. Univariate tests further revealed that condition significantly affected performance for all groups: English ($p < .001$), Learners ($p < .001$), and Mandarin ($p < .001$). Accuracy rates in the test condition (tone) were lower than those in the control (segment) condition across all groups (Figure 4).

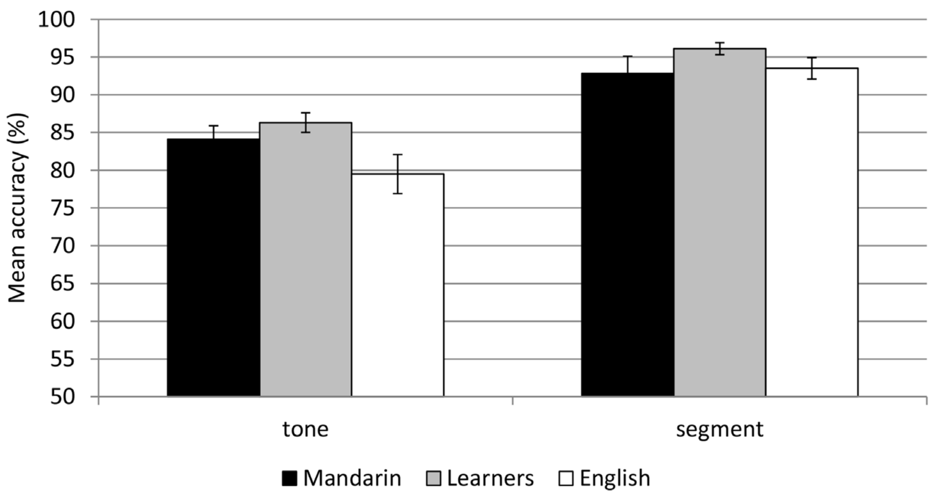

Figure 4 Mean Accuracy Rates (%) for Each Group in Test and Control Conditions

Note. Error bars represent +/-1 SE.

For RTs, when looking at the Type III tests of fixed effects, the F-tests showed: a main effect of Condition ($F(1, 10117) = 845.2$, $p < .001$), a significant effect of Language ($F(2, 68.01) = 3.99$, $p < .023$), and a significant interaction between the two factors ($F(2, 10117) = 13.26$, $p < .001$). The interaction shows that groups differ from one another on either the Test or the Control condition (i.e., between-group difference). Univariate tests demonstrate that groups were significantly different on both, the test ($F(2, 78.82) = 3.45$, $p = .037$) and control condition ($F(2, 77.73) = 3.62$, $p = .031$). In the test (tone) condition, Learners were significantly faster than the English ($p = .037$), but not faster than the

Mandarin (*p* = .874), while the English were slower than the other two groups on both conditions. Further, univariate tests demonstrate that condition significantly affected performance for each group: English ($p < .001$), Learners ($p < .001$), Mandarin ($p < .001$). RTs for tones were overall slower than those for segments, as can be seen in Figure 5. Taken together, the accuracy and RT data show that while groups did not differ in terms of accuracy, a processing difficulty can be observed for the English group, particularly in the tone (test) condition, where they are slower than the Learners.

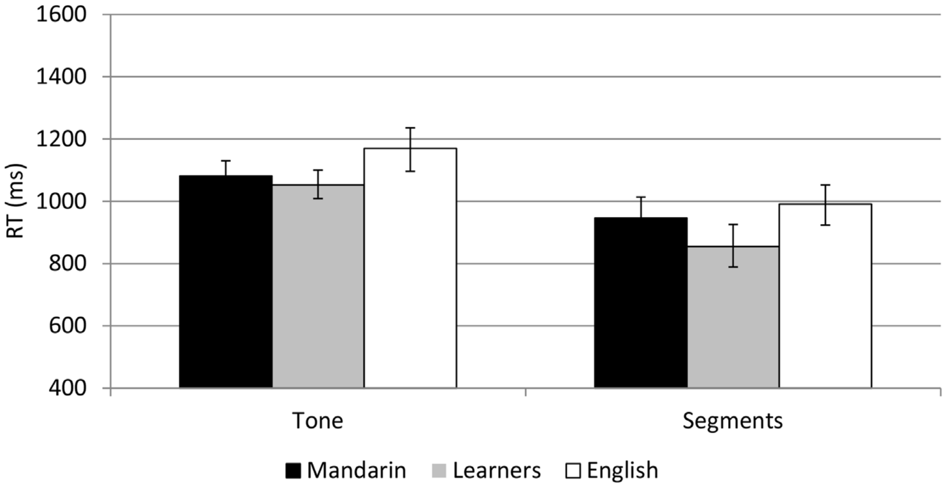

Figure 5 Mean RTs (ms) for Each Group in Test and Control Conditions

Note. Error bars represent the upper and lower CI.

Analysis by Subcondition

For this analysis, we examine each tonal combination as "Subcondition," in addition to Language and Group as above. The Type III tests of fixed effects showed: no effect of Language ($\chi2(2) = 4.67$, $p = .097$), a significant effect of Subcondition ($\chi2(14) = 488.36$, $p < .001$), and a significant interaction between the two factors ($\chi2(28) = 74.34$, $p = .001$). We see no cases of a statistically significant difference in accuracy in the test or control comparisons. All groups displayed significant within-group differences on two accuracy rates: Falling ($p = .018$) and Low ($p = .013$), accounting for the significant interaction. In Figure 6, we see the English group was slower than the other two groups on each test subcondition, although less so in several conditions. In Figure 7, we see all groups performed equally at high levels on control subconditions (see Schaefer, 2015 for detailed statistics).

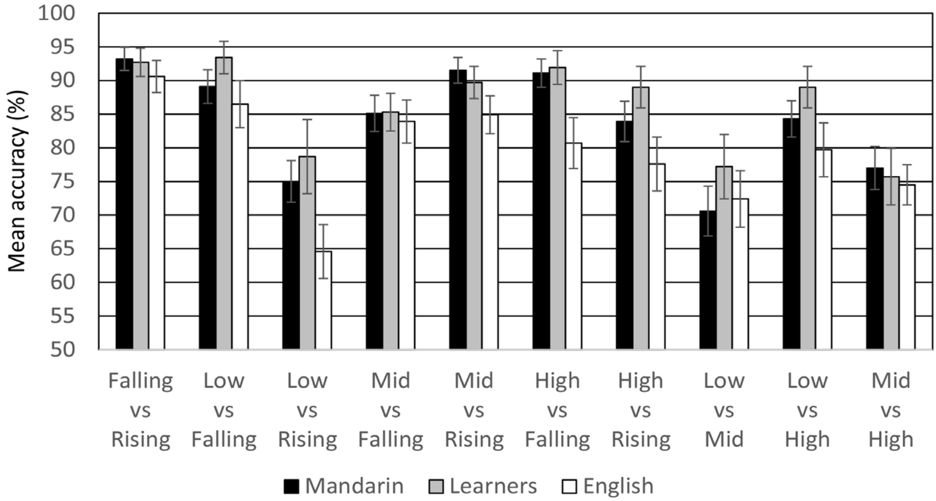

Figure 6. Mean Accuracy Rates (%) for Each Group on Each Test Subcondition

Note. Error bars represent +/–1 SE.

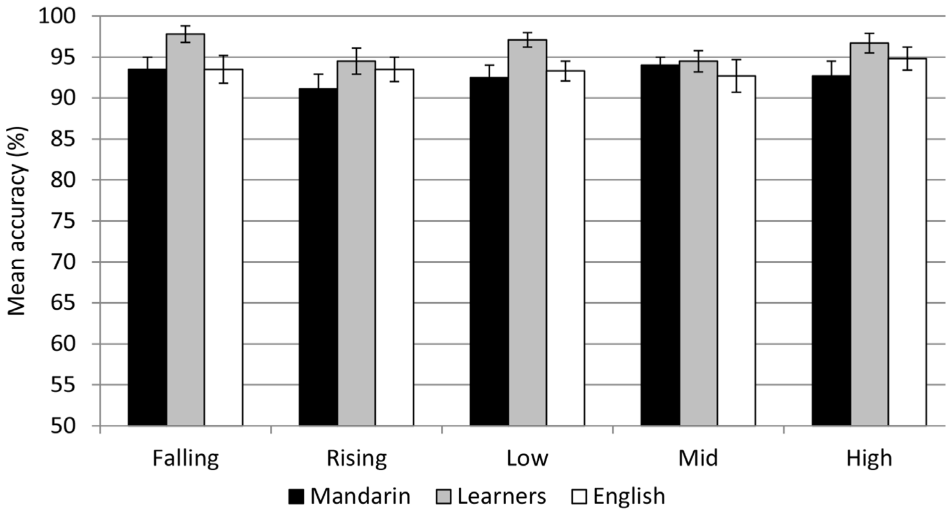

Figure 7. Mean Accuracy Rates (%) for Each Group on Each Control Subcondition

Note. Error bars represent +/-1 SE.

For RTs, when looking at the Type III tests of fixed effects, the F-tests showed: a main effect of Subcondition ($F(14, 9977) = 86.12$, $p < .001$), a significant effect of Language ($F(2, 69.59) = 3.64$, $p < .031$), and a significant interaction between the two factors ($F(28, 9977) = 2.0$, $p = .001$).

In Figure 8, the English group was slower than the other two groups on each test subcondition. Learners were significantly faster than the English in the

Low vs. Rising (p = .037), Low vs. High (p = .049), and Mid vs. High (p = .015) subconditions. The Mandarin group was faster than the English in the Falling vs. Rising (p = .013), Low vs. Rising (p = .037) and Mid vs. High (p = .001) subconditions. Univariate tests demonstrate that all groups display significant within-group differences in the RTs of two test subconditions: Falling vs. Rising ($p < .042$) and Mid vs. High ($p < .007$).

In Figure 9, in the control comparisons, the English were again the slowest (see Schaefer, 2015 for detailed statistics).

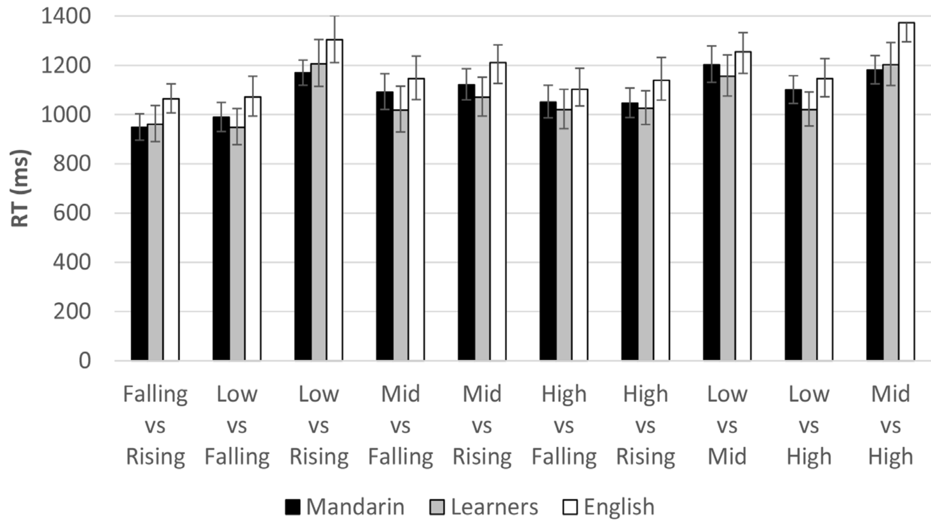

Figure 8. RTs (ms) for Each Group on Each Test Subcondition

Note. Error bars represent the upper and lower CI.

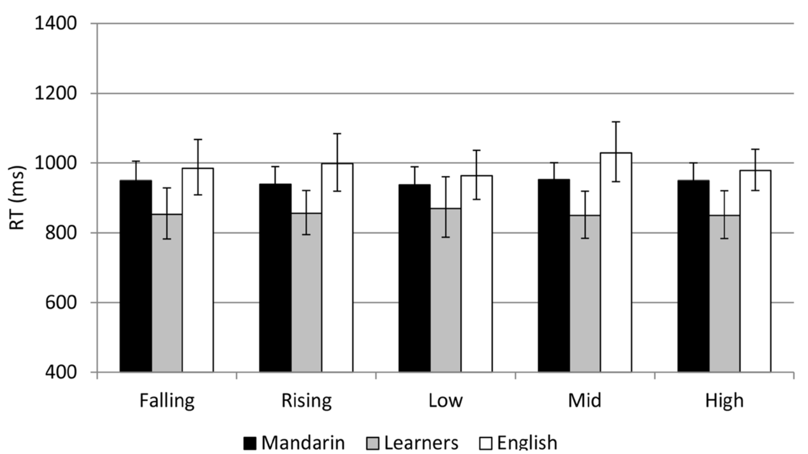

Figure 9. RTs (ms) for Each Group on Each Control Subcondition

Note. Error bars represent the upper and lower CI.

Thai ABX Disyllabic Task
Analysis by Condition

The Type III tests of fixed effects showed: no main effect of Language ($\chi2(2) = 3.52$, $p = .172$), a significant effect of Condition ($\chi2(1) = 210.86$, $p < .001$), and no interaction between the two factors ($\chi2(2) = 4.84$, $p = .089$). Performance differed significantly between groups in the test condition ($\chi2(2) = 7.88$ $p < .019$) but not in the control condition ($\chi2(2) = 4.31$, $p < .806$). There is only one difference between groups in the test condition: Learners were more accurate overall than the English ($p = .016$). Accuracy rates for the overall test condition were lower than those for the control condition across the three groups, and the English performed less accurately in the test condition (Figure 10).

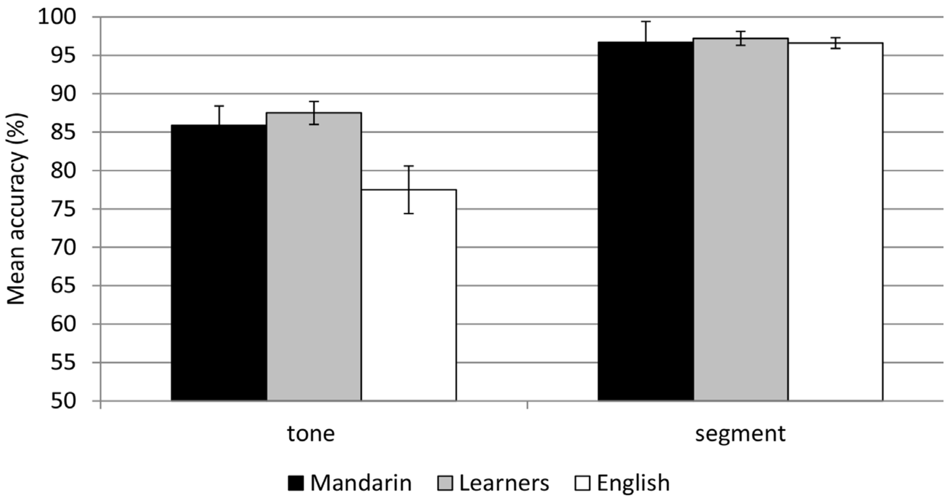

Figure 10. Mean Accuracy Rates (%) for Each Group in Test and Control

Note. Error bars represent +/–1 SE.

For RTs, when looking at the Type III tests of fixed effects, the F-tests showed: a main effect of Condition ($F(1, 4974) = 550.67$, $p < .001$), a significant effect of Language ($F(2, 67.09) = 3.43$, $p = .038$), and a significant interaction between the two factors ($F(2, 4974) = 24.06$, $p < .001$). There is a main effect of condition between the mean RTs on the Test vs. Control condition. Univariate tests revealed that performance did not differ significantly between groups in the test condition ($F(3, 72.53) = 2.54$, $p < .086$) but did vary in the control condition ($F(2, 72.48) = 6.15$, $p < .003$). In Figure 11, all groups were slower in the test condition with the English group the slowest on both conditions.

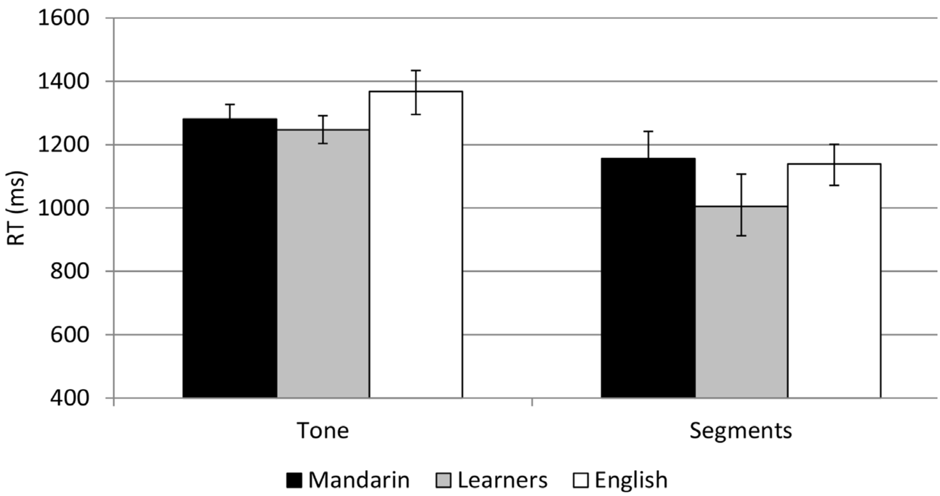

Figure 11. RTs (ms) for Each Group in Test and Control Conditions

Note. Error bars represent the upper and lower CI.

Analysis by Subcondition

The Type III tests of fixed effects showed: no main effect of Language ($\chi2(2) = 4.35$, $p = .114$), a significant effect of Subcondition ($\chi2(9) = 362.91$, $p < .001$), and a significant interaction between the two factors ($\chi2(18) = 59.45$, $p < .001$). The interaction is likely driven by the differences in performance between groups in the Test condition. Learners were more accurate than the English in Low + High vs. High + Low (p > .001) subconditions while the Mandarin group was more accurate than the English group in both, the Low + High vs. High + Low (p = .021) and Low + Mid vs. High + Low comparisons ($p < .001$). All three groups display significant within-group differences between accuracy rates on some subconditions: Low + High vs. High + Low ($p = .001$) and Low + Mid vs. High + Low ($p = .001$).

In Figure 12, we see that the English performed less accurately than the two other groups. The English group exhibited flat accuracy rates on four of the five comparisons (lower on the Low + High vs. Mid + High comparison). This trend is repeated but less robustly by the other two groups. In Figure 13, all three groups performed equally at high levels of accuracy in all subconditions (see Schaefer, 2015 for detailed statistics).

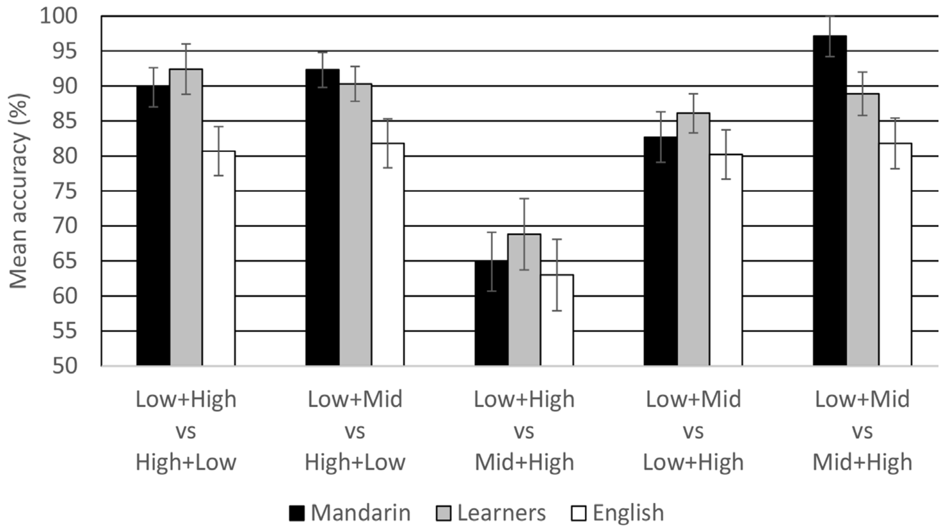

Figure 12. Mean Accuracy Rates (%) for Each Group in Each Test Subcondition

Note. Error bars represent +/-1 SE.

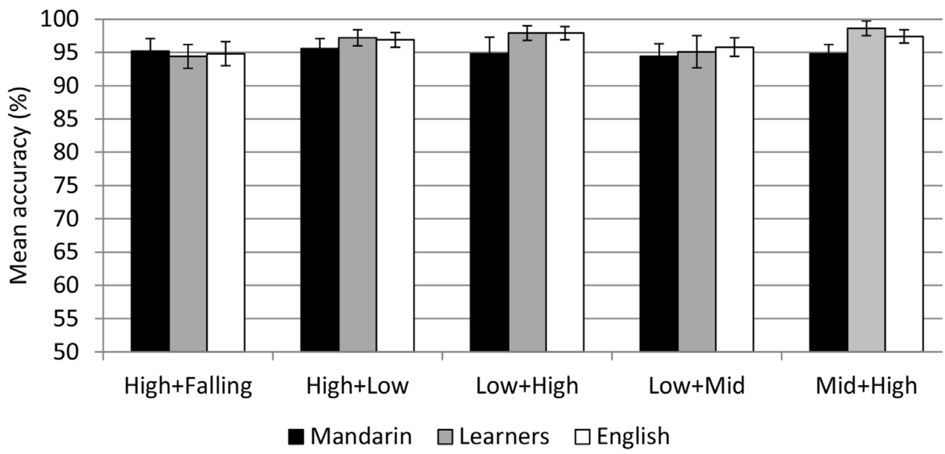

Figure 13. Mean Accuracy Rates (%) for Each Group in Each Control Subcondition

Note. Error bars represent +/-1 SE.

When looking at the Type III tests of fixed effects, the F-tests showed: a main effect of Condition ($F(9, 4949) = 84.40$, $p < .001$), a significant effect of Language ($F(2, 67.13) = 3.456$, $p = .037$), and a significant interaction between the two factors ($F(18, 4949) = 3.34$, $p < .001$). Univariate tests show that all three groups patterned alike in RTs: Low + High vs. High + Low ($p < 0.048$) and marginally in Low + Mid vs. High + Low RT ($p = .053$) subconditions. In Figure 14, the English performed at the slowest RTs with all three groups exhibiting a fairly similar pattern of performance on individual comparisons.

In Figure 15, RTs were flat for all three groups with Learners being faster than the other two groups or the Mandarin group being slower than expected (see Schaefer, 2015 for detailed statistics).

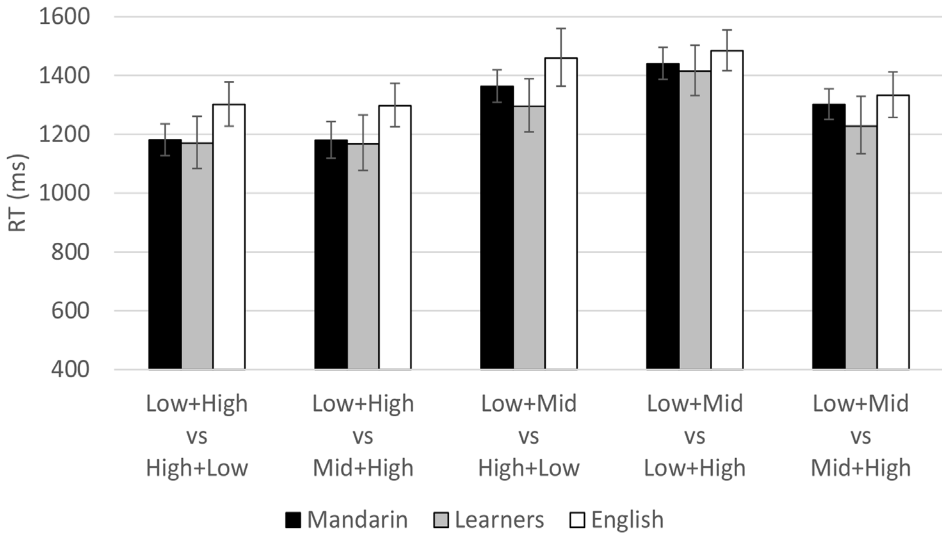

Figure 14. RTs (ms) for Each Group on Each Test Subcondition

Note. Error bars represent the upper and lower CI.

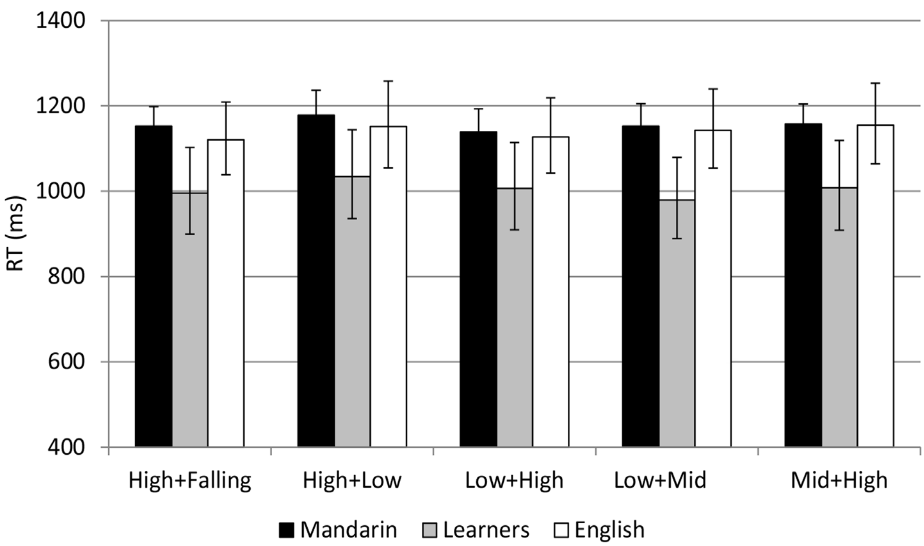

Figure 15. RTs (ms) for Each Group on Each Control Subcondition

Note. Error bars represent the upper and lower CI.

Mandarin Lexical Decision Task with Repetition Priming

For each participant, a mean accuracy score and mean RT were computed for both stimuli types (i.e., words and non-words). No participant was excluded for low performance on the distractors, and no item was excluded due to low performance by Mandarin natives. No outliers among the Mandarin or the Learner group were found, i.e., scored less than 70% on the real word vs. non-word accuracy response. Statistical analyses were run for both, overall accuracy and RTs for each condition (repetition and minimal-pair), for words and non-words. Sidak was used over the more conservative Bonferroni to estimate significance. The magnitude of priming effects is obtained by subtracting the mean RT during the second presentation of an item from the mean RT during the first presentation of the item. In "repetition" pairs, these two items are identical and are expected to trigger shorter RTs during the second presentation, resulting in a facilitation (priming) effect (a negative number). In "minimal pair" pairs, the two items are different, and are not expected to trigger shorter RTs during the second presentation if they activate a different lexical representation. Little priming or even inhibition is expected.

RTs were generally similar between groups (Table 2, Figure 16): Robust priming effects are obtained in most conditions for the "repetition" pairs, except for the Mandarin group on the segment condition. This unexpected lack of priming is difficult to explain and may be due to varying individual patterns of RT.

Table 2. Mean RTs (ms) and 95% CI for Words in Each Condition and Each Occurrence, for Both Groups

Language Group	Occurrence	Tone		Segment	
		Minimal Pair	Repetition	Minimal Pair	Repetition
Learners	first	1164 (1076;1259)	1117 (1033; 1208)	1140 (1054; 1233)	1094 (1012; 1186)
	second	1143 (1057; 1236)	1019 (942; 1102)	1109 (1026; 1199)	1023 (946; 1107)
	difference ($2^{nd} – 1^{st}$)	-21	-98	-31	-71
Mandarin	first	1143 (1074; 1213)	1122 (1057;1191)	1067 (1105; 1132)	1072 (1109; 1138)
	second	1146 (1079; 1216)	1033 (973; 1096)	1052 (991; 1117)	1064 (1102; 1130)
	difference ($2^{nd} – 1^{st}$)	2	-89	-15	-8

Note. CI = confidence interval (lower; upper).

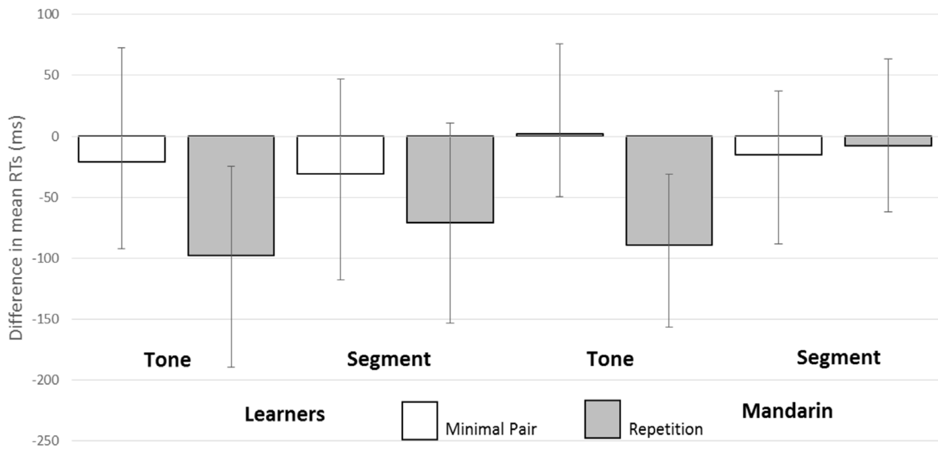

Figure 16. Difference in Mean RTs (ms) on Words for Each Condition and Group (RT2-RT1)

Note. Error bars represent 95% CIs.

At the non-words, the Mandarin group reacted as expected with little to no priming across both conditions and both, tone and segment items. In contrast, the Learner group exhibited an unexpected pattern: They showed priming in both conditions (-61 and -77 ms) of the tone non-words (Table 3, Figure 17) and also for the repeated condition of the segment non-words. While this could be due to the Learners' more variable lexical knowledge, it is important to recall that only RTs for correct responses were used in this analysis.

Table 3. Mean RTs (ms) and 95% CI for Non-Word Pairs in Each Condition and Each Occurrence, for Both Groups

Language Group		**Tone**		**Segment**	
	Occurrence	**Minimal Pair**	**Repetition**	**Minimal Pair**	**Repetition**
Learners	first	1186 (1084; 1294)	1233 (1130; 1346)	1167 (1069; 1271)	1178 (1079; 1282)
	second	1125 (1033; 1225)	1156 (1162; 1262)	1167 (1069; 1274)	1148 (1054; 1253)
	difference ($2^{nd} – 1^{st}$)	–61	–77	0	–30
Mandarin	first	1135 (1062; 1211)	1104 (1033; 1178)	1094 (1023; 1169)	1072 (1002; 1143)
	second	1135 (1064; 1213)	1122 (1050; 1197)	1089 (1021; 1161)	1091 (1021; 1167)
	difference ($2^{nd} – 1^{st}$)	0	18	-5	19

Note. CI = confidence interval (lower; upper).

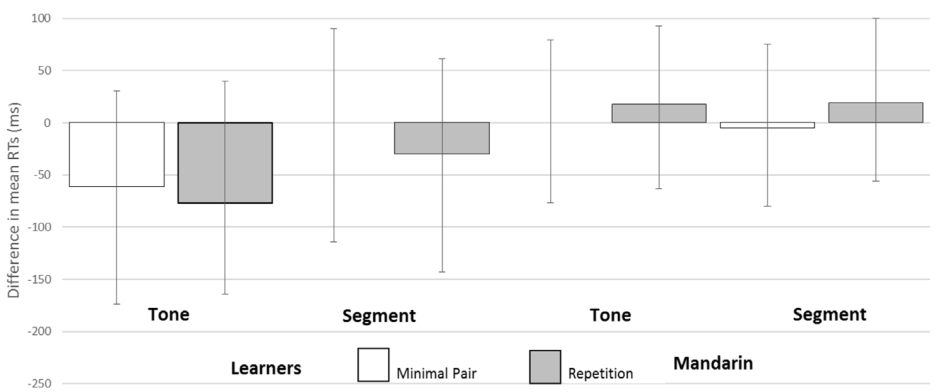

Figure 17. Difference in Mean RTs (in ms) to Non-Words for Each Condition and Each Group (RT2-RT1)

Note. Error bars represent 95% CIs.

For the real word items, a Linear Mixed Effects model was run on LogRTs. A log-transformation was applied to RTs (Log-RT) to obtain a normal distribution as RTs were skewed. Condition (tone vs. segments) and Language (Mandarin, Learners) were declared as fixed factors and Subjects as a random factor. When looking at the Type III tests of fixed effects in Table 4, the F-tests show a main effect of Condition ($F(1, 1444.2) = 7.108, p = .008$). There is no significant effect of Language ($F(1, 46.8) = 0.102, p = .751$). We see a main effect of "Minimal pair_same" ($F(1, 1444) = 18.22, p < .001$), which compares the RTs on minimal pairs vs. the RTs on repeated word pairs. Also, a marginal interaction was evident between Language and Minimalpair_same ($F(1, 1444) = 3.63, p = .057$) and Condition and Minimalpair_same ($F(1, 1444) = 3.74, p = .053$).

Table 4. Type III Tests of Fixed Effects for overall RTs on the Lexical Decision Task with Repetition Priming

Type III Tests of Fixed Effects[a, b]				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	46.812	120601.650	.000
Language	1	46.812	.102	.751
condition	1	1444.212	7.108	.008
minpair_same	1	1444.190	18.219	.000
Language * condition	1	1444.212	1.255	.263
Language * minpair_same	1	1444.190	3.633	.057
condition * minpair_same	1	1444.266	3.744	.053
Language * condition * minpair_same	1	1444.266	1.259	.262
a. lexical = Word b. Dependent Variable: RT_log.				

For the repetition of same tone words, Learners and Mandarin natives show a significant difference between the first and second RTs (p = .011 and p = .004, respectively), indicating significant priming. For the repetition of same segment words, Learners exhibit a marginal difference between the first and second RTs (p = .090) while the Mandarin group does not display any difference (p = .979), confirming the observed pattern in Figure 16. There were no statistical differences for either group on the minimal pair words, either tone or segments, as determined by a two-tailed test. In sum, as a group, the learners seem to have encoded tones in lexical representations to the same extent as native Mandarin speakers, as evidenced by the lack of spurious priming on the minimal pair tone condition. However, within-group variance was large, and priming effects ranged from -243 ms (large facilitation) to +255 ms (inhibition/no facilitation). Eight Learners displayed a facilitation, one Learner's priming was at 0, and 8 Learners displayed inhibition. In the next section, we examine whether the magnitude of the priming on this minimal pair tonal condition correlates with tone discrimination accuracy in an unfamiliar tone language.

Correlation of Thai ABX tasks and Mandarin Lexical Decision Task with Repetition Priming

To examine whether the robustness of tonal acquisition in Mandarin relates to the accuracy in the Thai tonal discrimination task, we use the Learners' performance in the lexical decision task. We reason that learners' acquisition of Mandarin tones can be indexed by their ability to *not* have priming in the tonal minimal pair condition of the lexical decision task. So, the absence of spurious priming–or the presence of inhibition–in this condition would be a good index of how robustly the tone is encoded in lexical representation for a word. The priming effect, if negative, indicates facilitation (spurious priming); if the priming is positive, it means that there is no facilitation, even inhibition. If this is related to their accuracy of Thai tone perception, then we expect to find a correlation where higher accuracy on ABX correlates with *less* priming (that is, more positive numbers). This relationship is presented in Figure 18.

A one-tailed Pearson correlation test was run between the accuracy rates of each participant (N = 17; any participant who was an outlier in one of the tasks was not included) in the Thai ABX monosyllabic task and each participant's priming effect in the tonal minimal pair word condition (in the lexical decision task). The association between the two variables was medium to strong (r = .593, $p < .006$), indicating a relationship in the direction we expected.

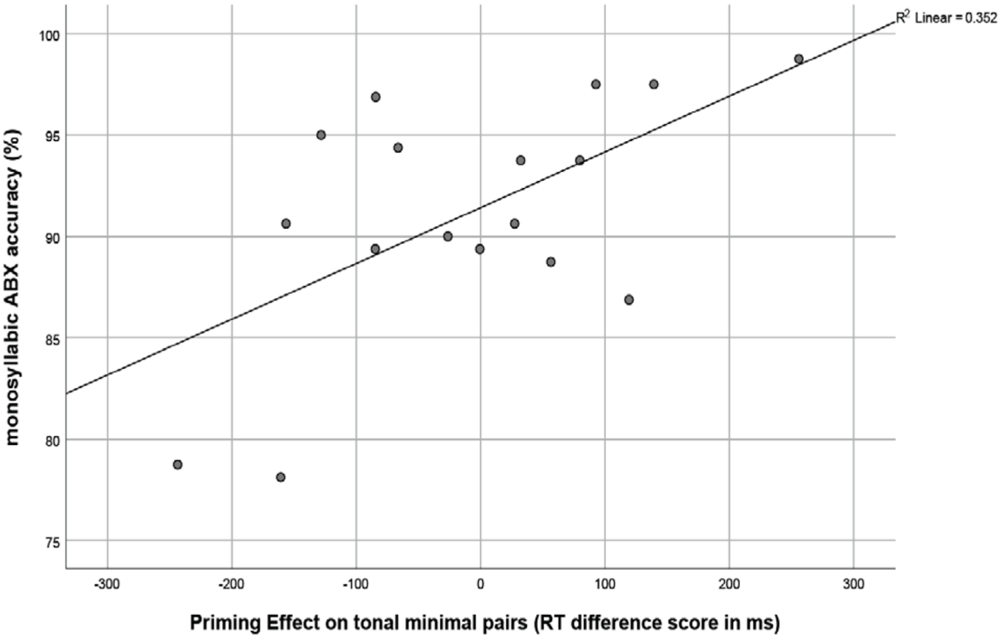

Figure 18. Scatterplot of Accuracy on ABX (y-axis) by Priming Effect Size on Tonal Minimal Pairs (x-axis)

Pronunciation Proficiency Test

In the pronunciation proficiency test, learners who were rated as having a more native-like pronunciation did not automatically perform more accurately in the Thai perception task and/or have faster RTs during the second occurrence of the tonal minimal pair condition of the lexical decision task. Some learners who were rated with lower native-like pronunciation performed with relatively high accuracy rates in the Thai perception task and faster RTs during the second occurrence of a tonal minimal pair. However, a one-tailed Pearson correlation test was run between the accuracy rates of each participant (N = 17) in the Thai ABX monosyllabic task and their pronunciation score, and between each participant's priming effect in the tonal minimal pair word condition and their pronunciation score. The association between the ABX and pronunciation was medium to strong ($r = .556$, $p = .010$) in the expected direction (see Figure 19, higher pronunciation scores correlate with higher ABX scores), whereas the one between pronunciation and lexical encoding was low ($r = .155$, *n.s.*).

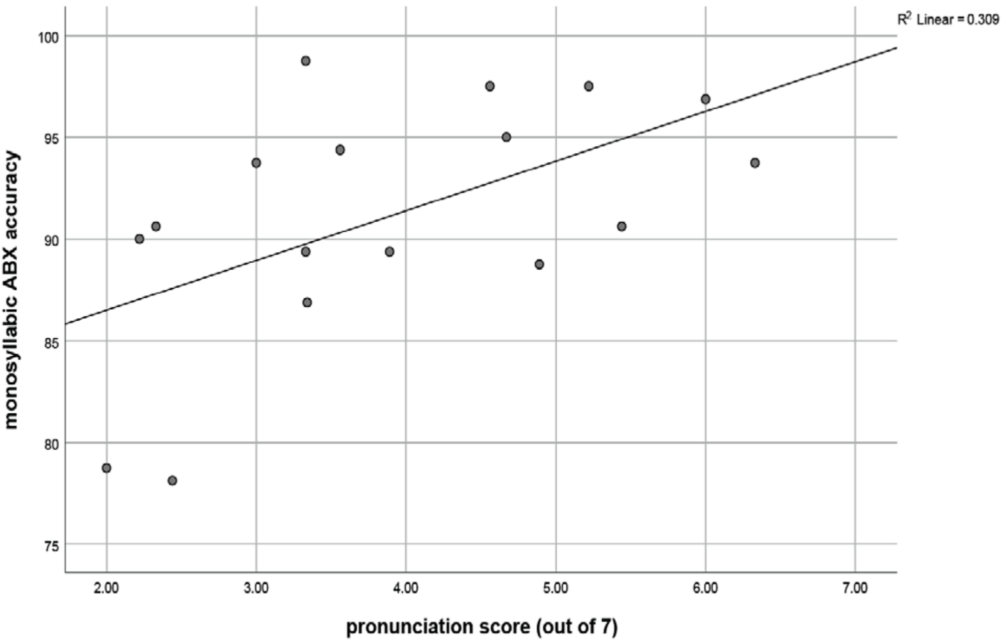

Figure 19. Scatterplot of Accuracy on ABX (y-axis) by Pronunciation Score (x-axis)

Discussion

In this paper, we compared the effect of L1 vs. L2 experience with Mandarin tones on perceiving novel tonal contrasts in Thai. Our results show that Learners were able to transfer their L2 experience with Mandarin tones to the perception of Thai tones: They were more accurate and faster than English participants who did not have experience with tones. In some conditions, the Learners' performance was not different from that of native Mandarin listeners. In particular, results of the subcondition analyses suggest that Learners may have increased their sensitivity to pitch direction, becoming more similar to L1 Mandarin participants, as shown, for instance, by the fact that both L2 Learners, and L1 Mandarin had faster RTs in the Low vs. Rising comparisons than the L1 English participants (p = .037 for both comparisons).

In order to claim that L2 experience with tones can indeed be transferred, it is necessary to establish that this knowledge is in fact acquired. To that end, we used a lexical decision task with repetition priming to examine the extent to which the learners demonstrated the ability to lexically represent tonal information in their lexical representations for Mandarin words. The results show that as a group, the Learners behaved as would be expected if the tones were encoded: They experienced facilitation (repetition priming) on items that were repeated, but did not experience significant (spurious) priming on words that minimally differed from other words in tones only (no significant repetition priming in the minimal pair tonal condition). This pattern parallels the one obtained by Mandarin native speakers for tones, and suggests that as a group, the Learners are able to lexically represent tonal information. However, individual variance was large in the tonal minimal pair condition. We finally correlated performance between tonal perception in the discrimination task (ABX monosyllables) and the size of the repetition priming in the tonal minimal pairs, reasoning that if L2 experience can indeed transfer, this may be visible in a relationship between having acquired tonal representations (as evidenced by the absence of facilitative priming in the minimal pair tonal condition) and accuracy in perceiving novel tonal contrasts. This relationship was borne out in the data with a significant correlation between the two dimensions. These results, therefore, indicate that as a dimension is robustly acquired (via the lens of lexical encoding), it can also be applied to similar, even if unknown, phonological dimensions.

This conclusion should however remain somewhat tentative, as several unexpected results emerged from the priming patterns in the lexical decision task, rendering interpretation difficult: It is unclear why Learners show priming patterns on non-words; this requires further research. Additionally, L1 Mandarin speakers unexpectedly did not show priming in the segmental (control) repetition word condition, a pattern that is equally difficult to explain. The Learners' priming pattern in this condition is also only marginal, suggesting that possibly superficial, subphonemic differences in realizations between stimuli used for the first vs. the second presentation of the words may have been present. These could be sufficiently perceptible to prevent the two instances to be perceived as repetition from one another. Mandarin native speakers may have been more attuned to such differences than the Learners, for instance, if the tones, vowels or consonants were realized slightly differently by the one speaker across the two repetitions. Future research is clearly warranted here. Of note, this unanticipated result aligns with Qin and Jongman (2016) who found differences in perception due to varying segments: Unlike Learners, L1 Mandarin speakers performed less accurately on Cantonese tones carried by [jɐu] due to phonetic variation, namely level tones for /jau/ which are closer than level tones for [se]. However, Qin and Jongman (2016) did not determine the encoding of L2 Mandarin tones by the Learners. Another potential reason for the absence of repetition priming in the segmental same condition could be due to higher variability overall. The segment condition is identical to the tone condition, except that it contains all four tones while the tone condition contains only tones 1 and 2. As such, the segment condition is more difficult (even for the "repeated" pairs), particularly as native speakers have been noted to confuse tone 2 with tone 3 and tone 1 with tone 4 (Lee, Tao, & Bond, 2008). Since paired items are separated by a number of other items (between 8 and 20), native speakers may perceive the second instance differently from the first, which would prevent them from activating the same lexical representation, and trigger no facilitation (repetition) priming.

This could be compounded by the potential influence of dialectal differences in Mandarin on tone perception (cf. Li, Xiong, & Wang, 2006).

As to why Learners show priming in the repetition and minimal pair conditions for the non-words (which should show no priming across the board, as seen in the Mandarin participants), this could possibly be due to the fact that they "misinterpret" the non-words as words in terms of lexical activation. Even if they correctly reject them as non-words, it is conceivable that the items have activated similar words – which would then possibly trigger repetition priming in either condition. A few other limitations to the current study should be noted. In this paper, we decided to index "robustness" of acquisition of tonal contrasts by using the amount of inhibition/priming effect in one condition of the lexical decision task. This choice can be criticized on several accounts. Firstly, it may not actually be a good estimation of participants' true acquisition of tones. More research using an expanded battery of tasks should be used to obtain correlations with performance in a task such as the lexical decision we used. Secondly, given the unexpected findings in other conditions of this task, using performance in the condition we chose (tonal minimal pairs) may not be warranted. Ideally, a very robust task should be used in future studies. Finally, determining the relationship between overall language proficiency to pronunciation is an important future endeavor, since higher or lower levels in the former do not equal higher or lower levels in the latter. An evaluation of each participant's overall pronunciation might be necessary. At the very least, testing learners of Mandarin at early stages (i.e., a few months to less than two years) may be required to determine how early in the learning process they encode tone.

It is interesting that on the one hand, the lexical decision priming magnitude was related to ABX discrimination accuracy, but our proficiency test was not. There can be several reasons for this: The lexical and the ABX tasks are both based on perceptual performance, whereas the proficiency task is based on production. On the other hand, both the lexical and the proficiency task involve real words. Previous research has found that discrimination accuracy was related to differences in executive control such as inhibition (Darcy, Mora & Daidone, 2016), where more accurate discrimination was accompanied by a greater ability to inhibit the first language from interfering, but this relationship was only obtained in perception. It did not hold for production. Future research will be needed in this area as well to determine whether the lexical decision task we used can truly serve as a measure of "robustness" of acquisition, or whether it is "merely" another measure of tonal perception accuracy. If so, the relationship we found may not actually be an instance of transfer of L2 experience but a simpler claim that accurate perception helps across all languages someone is exposed to. If, however, it indexes robustness of acquisition, our claim that this acquired knowledge can then transfer to an unknown language will need to be confirmed with other tasks.

To conclude, our results show that L2 experience with tones can benefit learners in perceiving novel, unknown tonal distinctions, and therefore expand and confirm previous reports that in order to be beneficial, experience is not necessarily restricted to L1 experience with specific phonological dimensions. Whether or not this is truly an instance of L2 transfer remains tentative. If it is indeed transfer, the Feature Hypothesis provides a useful framework for this phenomenon and, therefore, could be expanded to include L2 features which, when learned or acquired in that L2, can be applied to perception of a non-native or a third language. If this scenario is confirmed, the good news is that a newly learned second language dimension could be applied to the unknown.

Acknowledgements

We would like to express our gratitude to the following people for their guidance and support of this research: Professors Rex Sprouse, Laurent Dekydtspotter, Öner Özçelik, Charles Lin, Kathleen Bardovi-Harlig, and the Department of Second Language Studies at Indiana University. We further thank Wannaporn Tongtako along with Pawanrat Rattana and Janjarus Karachot for checking and recording the Thai stimuli. We would like to thank Yunwen Su for checking and recording the Mandarin stimuli and for evaluating the Mandarin pronunciation samples along with Chung-Lin Martin Yang and Yueh Tu. Additionally, we thank Natasha Branch for her expertise in recording all the stimuli and Stephanie Dickinson and the Indiana Statistical Consulting Center (ISCC) for their patient and meticulous assistance with the statistics. Also, we would like to thank Dr. Yea-Fen Chen and the staff at the Indiana University Chinese Flagship Program for allowing us to recruit Mandarin learners. The research was supported by a generous dissertation grant from the journal of *Language Learning* and a Louise McNutt Graduate Dissertation Year Research Fellowship from the College of Arts and Sciences at Indiana University, for which we are grateful. We also thank the many participants. Additionally, we would like to thank two anonymous reviewers and Dr. Marcin Szczerbinski for their many immensely helpful suggestions and insightful comments to improve this article.

Funding source

Language Learning (Journal) Dissertation Grant (2014) * INTERNAL FUNDING: Indiana University College of Arts and Sciences Dissertation Year Research Fellowship (2014-2015)

References

Abramson, A. S. (1975). The tones of Central Thai: Some perceptual experiments. In J. G. Harris & J. R. Chamberlain (Eds.), *Studies in Thai linguistics in honor of William J. Gedney* (pp. 1-16). Bangkok: Central Institute of English Language.

Abramson, A. S. (1976). Thai tones as a reference system. In T. W. Gething, J. G. Harris, & P. Kullavanijaya (Eds.), *Tai linguistics in honor of Fang-Kuei Li* (pp. 1-12). Bangkok: Chulalongkorn University Press.

Braun, B., Galts, T. & Kabak, B. (2014). Lexical encoding of L2 tones: the role of L1 stress, pitch accent and intonation. *Second Language Research 30*(3), 323-350.

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics,* 39, 585-594.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the fourth international conference on spoken language processing 1* (pp. 2514-2517). Wilmington, DE: Applied Science and Engineering Laboratories.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of Central Thai tones and segments by Thai and Australian adults. In *Pan-Asiatic Linguistics: Proceedings of the third international symposium on language and linguistics* (pp. 546-560). Bangkok: Chulalongkorn University Press.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*(3)*,* 207-228.

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

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front round vowels in L1 English-L2 French acquisition. *Second Language Research, 28,* 5-40.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, 66(4), 741-773.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568-1578.

Dupoux, E., Pallier, C., Sebastain, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*, 406-421.

Fernandes-Boëchat, M. H. (2007). The CCR theory: A cognitive strategy research proposal for individual multilingualism. *Revista Luminária*, 8(1), 93-97.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of Acoustic Society of America*, 93(3), 1589-1608.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. M. (2008). Perceptual learning of Cantonese lexical tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *36*(2), 268-294.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 49-175.

Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1-33.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r"s. *Neuropsychologia, 9*, 317–323.

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*, 269-279.

Hao, Y.-C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, 61(1), 135-152.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America, 118*(2), 1038-1054.

Kojima, C., & Darcy, I. (2014). Learners' proficiency and lexical encoding of the geminate/non-geminate contrast in Japanese. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. M. Tseng, A. Tuninetti, & D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges Between Disciplines* (pp. 30-38). Somerville, MA: Cascadilla Proceedings Project.

Leather, J. H. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (Eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second-Language Speech* (pp. 72-89). Amsterdam: University of Amsterdam.

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*(1), 101-123.

Lee, C.-Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36*(4), 537-563

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics, 53,* 157-165.

Li, A., Xiong, Z., & Wang, X. (2006). Contrastive study on tonal patterns between accented andstandard Chinese. In Q. Huo, B. Ma, E.-S. Chng, & H. Li (Eds.). Chinese Spoken Language Processing: 5th International

Symposium, ISCSLP 2006, Singapore, December 13-16, 2006, Proceedings (pp. 157-168). Berlin: Springer-Verlag.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229-258.

McGinnis, S. (1997). Tonal distinction errors by beginning Chinese language students: A comparative study of American English and Japanese native speakers. In S. McGinnis (Ed.), *Chinese pedagogy: An emerging field* (pp. 81-91). Columbus, OH: Ohio State University Foreign Language Publications.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries. *Psychological Science*, *12*(6), 445-449.

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2018). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition, 34*, 187-214.

Qin, Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech, 59*(3), 318-338.

Schaefer, V. (2015). *Influence of the first and second language on the perception of Thai tones* (Publication No. 3737731) [Doctoral dissertation, Indiana University]. ProQuest Dissertations Publishing.

Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology, 5*(4), 489-522.

So, C. K., & Best, C. K. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*, 273-293.

Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and 'tone': Evidence from Thai. *Journal of Phonetics, 1*, 101-109.

Wang, Y., Jongman, A., & Sereno, J. A. (2006). L2 acquisition and processing of Mandarin Chinese tones. In P. Li, E. Bates, L.H. Tan, & O. Tseng (Eds.), *The handbook of East Asian psycholinguistics* (pp.250-256). Cambridge, UK: Cambridge University Press.

Wayland, R. P., & Guion, S. G. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics 24*, 113-129.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: *A preliminary report. Language Learning, 54*, 681-712.

Weber, A. & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language,* 50, 1–25.

Wiener, S., & Goss, S. (2019). Second and third language learners' sensitivity to Japanese pitch accent is additive: An information-based model of pitch perception. *Studies in Second Language Acquisition, 41*, 897-910.

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism* *7*(1), 75-90.
Wrembel, M., Marecka, H., & Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism.* Advance online publication. doi:10.1080/14790718.2019.1583233
Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes, 11*(6), 589-596.