SOUTHERN JOURNAL OF LINGUISTICS
43.1: 101-143 •

The Influence of L1 English Word Stress and L1 Japanese Pitch Accent on the Naïve Perception of Thai Tones in Monosyllabic vs. Disyllabic Stimuli

Vance Schaefer[1]
University of Mississippi

The naïve perception of non-native lexically-contrastive pitch is shaped by the functionality of lexically-contrastive pitch in the native language (Schaefer and Darcy, 2014). Specifically, the naïve perception of non-native tones (Thai) results in a hierarchy of performance shaped by the relative importance of lexically-contrastive pitch in the L1 (from more to less): tone (Mandarin) > pitch accent (Japanese) > word stress (English) > no lexically-contrastive pitch (Korean). However, both L1 English and L1 Japanese listeners may not have fully accessed their respective L1 word stress and L1 pitch accent as the Thai stimuli in the Schaefer and Darcy (2014) study were monosyllabic. That is, "full perception" may have been inhibited without disyllabic stimuli for English speakers or potentially dimoraic stimuli for Japanese speakers to contrast the low vs high pitches reflective of lexically-contrastive pitch in the L1. In response, the current follow-up study employs disyllabic stimuli in an attempt to fully assess the influence of the functionality of L1 lexically-contrastive pitch on the naïve perception of non-native Thai tones. Results of the current study suggest a similar perceptual hierarchy. However, disyllabic stimuli reflective of English word stress or Japanese pitch accent do not appear to aid in perception, but shed light on potential universal and language-specific trends.

1. Introduction

Naïve listeners (=non-learners) with various native languages (L1s) differ in their accuracy in identifying non-native tones. This is attributed to a difference in the ability to attend to the features of pitch height and/or direction as influenced by the L1 (Francis, Ciocca, Ma and Fenn 2008; Gandour 1983). Native (L1) speakers of a tone language, however, generally perceive another tone language more easily than L1 speakers of a non-tonal language (Gandour and Harshman 1978; Hallé, Chang and Best 2004; Wang, Behne, Jongman and Sereno 2004; Wayland and Guion 2004). This superior performance by L1 tone language speakers in identifying non-native tones as compared to L1 speakers of non-tone languages is due to the former group's ability to more accurately perceive pitch height and direction. Additionally, L1 tone language speakers can map non-native tones onto L1 tones (Wayland and Guion 2004) as they perceive tones as linguistic categories (Van Lancker and Fromkin 1973; Wang, Jongman and Sereno 2001) with tonal information constraining lexical access (Lee 2007). Native speakers of non-tonal languages with no exposure to a tone language, on the other hand, are less sensitive in discriminating between different tones than L1 speakers of a tone language (Hallé, Chang and Best 2004, for French listeners; Gandour and Harshman 1978; Wang, Behne, Jongman and Sereno 2004). However, speakers of non-tonal languages are not all equal in their sensitivity to non-native tone perception (Burnham et al. 1996, Schaefer and Darcy 2014, So 2006, So and Best 2010) as languages differ in the extent and function to which they use linguistic pitch in their L1. Understanding cross-linguistic perception of tone is further complicated by the fact that linguistic pitch exists in all languages in the form of intonation, whereas linguistic pitch to distinguish lexical items does not exist in all languages and is employed by varying degrees and manner in those where it does exist.

The current study is a follow-up study to Schaefer and Darcy (2014) which demonstrated that the varying degrees and manner of lexical pitch in the native language (L1) shape the naïve perception of lexical pitch in a non-native language. Specifically, the extent to which lexical pitch in the form of tone (Mandarin), pitch accent (Japanese), word stress (English), or none (Korean; intonation only) influences the naïve perception of non-native Thai tones. That is, the salience of lexical pitch in these languages varies by functional load (i.e. number of words distinguished by lexical pitch), domain (vowel/syllable, word, phrase), exclusivity (other features in addition to pitch), and inventory (number of pitch patterns). Indeed, results indicate a hierarchy of perceptual accuracy determined by the degree and manner of lexical pitch: L1 Mandarin > L1 Japanese > L1 English $\geq$ L1 Korean. In short, the salience of lexically-contrastive pitch in the L1 shapes the naïve perception of non-native lexical pitch (cf. Feature Hypothesis, McAllister, Flege and Piske 2002).

However, L1 Japanese and English listeners in the previous study of Schaefer and Darcy (2014) may not have fully accessed the lexically-contrastive pitch of pitch accent and word stress, respectively, as the study employed only monosyllabic target Thai stimuli. That is, both Japanese pitch accent and English word stress realize full prominence in disyllabic words where a high and low pitch in Japanese or a stressed and unstressed syllable in English are juxtaposed against one another "maximizing prominence" as heard particularly in minimal pairs (e.g. A.me 'rain' vs a.ME 'candy' in Japanese, an IN.sight vs to in.CITE in English where capitalization indicates high pitch or stress, respectively). As such, results in the previous study may not have accurately reflected the influence of L1 lexical pitch in Japanese and English on the naïve perception of Thai tone. Thus, the current study replicates the previous study but employs both monosyllabic and disyllabic stimuli in order to obtain a more accurate assessment of the role of lexically contrastive pitch in these two languages on the naïve perception of a non-native linguistic pitch system.

The current study wishes to answer the following research questions:

RQ1: Does the varying role of lexical pitch in the native language (L1) (i.e., pitch accent, word stress) shape the naïve perception of another non-native lexical pitch system (i.e. Thai)?

RQ2: Do speakers of Japanese and English perceive disyllabic stimuli more accurately and quickly than monosyllabic stimuli?

2. Cross-Linguistic Overview of Lexical Pitch Systems

Languages differ typologically in the maximal and exclusive use of linguistic pitch to distinguish words. The five languages of Mandarin, Thai, Japanese, English and Korean are the target languages in the current study due to the differences in this differing lexical pitch usage.

2.1. Tone Languages

Tone languages such as Thai and Mandarin Chinese use various pitch patterns (referred to as *tone*) maximally since pitch variations contrast words lexically at the syllable level. The pitch patterns in these two languages may be relatively level and spoken at the high, mid or low end of the voice range, or they may have a simple contour, i.e. pitch movement, by falling or rising or both. Each syllable generally has a set pitch pattern, perceived by listeners primarily through pitch height and pitch movement (Gandour 1983). As shown in Table 1, Mandarin has four tones (one level tone and three contour tones): high level (tone 1), rising (tone 2), low falling rising or dipping (tone 3), and falling (tone 4). The classic example of tone is the Mandarin segmental string of /ma/ which depending on the tone has four different meanings. The neutral tone tends to be level at a mid pitch (Chen and Xu 2006) but may display a wide range of contours as shaped by the tone in the preceding syllable with a duration approximately half or a little less than full tones (Lee and Zee 2008). By contrast, Thai has five tones composed of three flat tones (i.e. low, mid, high) and two contour tones (i.e. falling, rising) so that /na:/ has five different meanings depending on the tone (see Table 1).

Table 1. Overview of Tones in Mandarin and Thai[2]

	Mandarin		Thai	
Tone type	Tone	Example word	Tone	Example word
Level	high level (tone 1)	mā *mother*	high	ná: *aunt*
	neutral tone	'marks a question'	mid	na: *rice field*
			low	nà: (nɔ:i nà:) *custard apple*
Contour	rising (tone 2)	má *hemp*		
	low or dipping (tone 3)	mǎ *horse*	rising	nǎ: *thick*
	falling (tone 4)	mà *to scold*	falling	nâ: *face*

Zsiga and Nitisaroj (2007, p. 344) for Thai.

However, there is a discrepancy between these phonological descriptions and phonetic (i.e. physical) descriptions. For example, for Thai tones the high tone (H) is not phonetically level but in fact almost parallels the rising tone (R) in shape (Morén and Zsiga 2006). Nevertheless, listeners must follow the height and/or direction of pitch or weight one more than the other to distinguish tones from one another.

2.2. Pitch-Accent Languages

Pitch-accent languages also use pitch to distinguish words, but not to the same extent and manner as tone languages do. This diverse group of languages includes standard Japanese, Scandinavian languages, and the Kyungsang dialect of Korean (Lee and Ramsey 2000). Pitch accent languages are considered a subclass of tone languages (Yip 2002,) where a language with tone is “one in which an indication of pitch enters into the lexical realization of at least some morphemes" (Hyman 2009: 231).

In Japanese, one mora of each word receives a high pitch (i.e. the "word accent") determining the pitch on the preceding or subsequent moras (Kubozono 1999). One example is /ka.ki.ga/ which can mean 'oyster' (HLL or initial-accented where the first mora is marked by a high tone and the following two moras are marked by low tones), 'fence' (LHL or final-accented), or 'persimmon' (LHH or unaccented) where *ga* is a particle indicating the nominative case. Thus, pitch accent languages resemble tone languages as they employ pitch to contrast words, but with less pitch movement, or a more restricted inventory.

2.3. Stress-Accent Languages

In stress languages like English, pitch is never used alone to distinguish meaning or mark prominence. The stressed syllable as compared to an unstressed syllable generally features higher pitch, greater intensity, lengthened vowels, and unreduced vowel quality (Fry 1958). A typical contrast of stress in English would be the noun "IMpact" and the verb "imPACT" (where capital letters indicate the stressed syllable). Pitch may, however, play a lesser role than loudness and vowel duration in identifying stress among native speakers (Kochanski, Grabe, Coleman and Rosner 2005) or vowel quality due to its strong correlation with stress (Cutler 2015) although pitch has also been shown to be the most salient feature (Fry 1958). Thus, stress-accent languages differ from both tone and pitch-accent languages as pitch is not necessarily used exclusively to mark prominence in words, and in the case of English pitch may not play the most salient role in marking the prominence of a syllable.

2.4. Languages with No Use of Lexical Pitch ("Intonation-Only" Languages)

Some languages do not use tone, pitch accent or stress to distinguish words. Like all languages, they do use pitch (height and contour) and the associated features of duration and intensity at the level of the phrase, i.e. intonation (even if the phrase is just one word long, such as "really?") to communicate questions, statements, and emotional state but not to distinguish words (Hirst and Cristo 1998). As such, these languages can be referred to as "intonation" languages. Such languages include standard Korean (Kim-Renaud 2009) and some dialects of Japanese (Otake and Cutler 1999).

3. Perception of Tone

3.1. Perception of Native Tones

Pitch height and direction are considered the two greatest features to shape the perception of tone whether in the L1 or L2 (Gandour 1983; Wang, Jongman and Sereno 2006). As for height, pitch may be produced at the high end or low end of the voice range or between the high and low ends, e.g. mid. As for direction, pitch may remain relatively level or move up and/or down. Thus, speakers of tone languages must gauge the height and/or direction of pitch in perceiving tone.

However, the L1 shapes the weighting of pitch height and direction. Indeed, it appears that pitch height is weighted more than pitch direction among languages in perceiving tone (Gandour 1983), likely as only height is needed to differentiate certain tones (Tuc

2003; e.g. low and mid-level tones in Thai). Additionally, among tone languages direction is weighted much more (although still less than height) than among non-tone languages (e.g. English).

Additionally, in opposition to the feature of height or direction in a unitary model, an alternative approach views direction as a sequence of high and low pitches aligned to certain moras in a word (cf. Mixdorff, Luksaneeyanawin, Fujisaki and Charnavit 2002) under a compositional model (Moraic Alignment Hypothesis, Morén and Zsiga 2006; Tumtavitikul 1995). This alignment with a specific mora serves as the primary cue in differentiating the five Thai tones as illustrated in Table 2. However, both the unitary and compositional models may be plausible with the unitary model perhaps being "streamlined perceptual shorthand."

Table 2. Moraic Alignment Hypothesis for Thai Tones

	Mid tone		High tone		Low tone		Falling tone		Rising tone	
phrase-final position (incl. citation forms)	μ	μ	μ	H μ	μ	L μ	H μ	L μ	L μ	H μ
nonphrase-final position	μ	μ	μ	H μ	μ	L μ	H μ	μ	L μ	(H) μ

μ = mora; L = low pitch; H = high pitch; blank = no low or high pitch.
(Morén and Zsiga 2006, Zsiga and Nitisaroj 2007: 344-345)

L1 speakers of a tone language can discriminate some tones more easily when comparing tones (Abramson 1975, 1978, Burnham et al. 1992). For example, the comparison between level vs. contour tones (e.g. Low vs. Rising) is the most difficult to discriminate by L1 speakers of Thai, whereas that between contour vs contour tones (Rising vs. Falling) is the easiest (Burnham et al. 1992). Additionally, it is difficult for native speakers to discriminate between the Thai low and mid level tones (Abramson 1976) which have the same shape (i.e. direction) but differ by minimal height. In contrast, all other tone comparisons allow the listener to access the two differing features of direction and height. The difference between the mid and high tones differ by both features, or if only by height, the difference between the mid and high tone is greater than that between low and mid. This possible difference in the magnitude between the low vs mid and mid vs high tones may play a role in either a unitary or compositional model.

3.2. Perception of Second Language Tones

Tone perception in naïve listeners varies by the L1.

3.2.1. Speakers of Tone Languages

Native speakers of a tone language apply their L1 experience with tones to their perception of non-native tones. That is, such speakers appear to map tones from their L1 onto what they perceive to be similar tones in the L2 or they access their experience with using pitch height and direction when perceiving non-native tones (Wayland and Guion 2004). For example, L1 Mandarin speakers might map their falling tone [51] onto the Thai falling tone [51] (where lower numerical values indicate a low pitch and higher values a high pitch on a scale of 1 to 5, Chao, 1948) or employ their ability to follow tone direction, enabling them to outperform non-tone language speakers. Additionally, L1 Mandarin speakers confuse the Thai mid [33] and Thai low [11] tones as Mandarin has no equivalent tone (Gandour 1983), the closest being Mandarin tone 2 [35] and tone 3 [214] (Wayland and Guion 2004).

Additionally, So and Best (2010) conclude that having tones in the L1 does not necessarily aid in the perception of L2 tones since as noted the L1 can also impede L2 perception both phonologically and phonetically. That is, experience in speaking a tone language not only strengthens the categorical perception of tones but also may impede perception of similar but sufficiently different tones as predicted under the Perceptual Assimilation Model (PAM, Best 1995). Indeed, we find that L1 Mandarin speakers are more categorical in their responses than English L1 speakers to Mandarin tones (Leather 1987, Stagray and Downs 1993) as L1 Mandarin speakers need to be less sensitive to pitch differences in order to account for the variation in tone production of a single tone pattern. For example, cross-linguistically we find that L1 Cantonese speakers are confused by similarities between their L1 Cantonese tones and target L2 Mandarin tones. They performed more poorly than the other two groups (i.e. Japanese and English) in distinguishing the target L2 Mandarin tone pairs 1 and 4 and tone pairs 2 and 3. Specifically, they mapped L2 Mandarin tone 1 [55] and 4 [51] onto their L1 Cantonese tone 1 [55] which has the corresponding allotone of 53 as well (i.e. Cantonese tone 1 is realized as either 55 or 53). Also, they mapped both L2 Mandarin tone 2 [35] and tone 3 [214] to Cantonese tone 2 [25 or 35] due to the similarity of their pitch contours.

3.2.2. Speakers of Pitch Accent Languages

L1 speakers of pitch accent languages pattern similarly to L1 speakers of tone languages when it comes to accuracy in the perception of non-native tones (Burnham et al. 1996, So 2006). For example, L1 Swedish speakers mirrored both L1 Cantonese and L1 Thai speakers in their accuracy rates in the perception of Thai tones and only the L1 Cantonese speakers in terms of reaction times but not the faster L1 Thai speakers (Burnham et al. 1996) which would be expected as Thais can access their lexicon. Compared to L1 speakers of English, L1 speakers of pitch accent languages are more accurate in their perception of non-native tones (Burnham et al. 1996). L1 speakers of Japanese (pitch accent) also improved much more than L1 speakers of English (word stress) when learning Mandarin tones (McGinnis 1996). Additionally, speakers of both languages tend to notice pitch height (Guion and Pedersen 2007).

Japanese listeners clearly assimilated Mandarin tones onto "Japanese pitch accent categories" (i.e. HH, LH, HL) in an identification task (So 2010). In contrast, in another study employing an identification task (So and Best 2010), Japanese had difficulty mapping L2 Mandarin tone 2[35] and tone 4 [51] onto Japanese LH and HL pitch patterns, respectively.

3.2.3. Speakers of Stress Languages

Speakers of stress languages access their experience with intonation in the L1 to the perception of tones (e.g. Mandarin, Francis et al. 2008). Speakers of stress languages may access the lexical pitch associated with word stress in the L1 to L2 tones. For example, L1 lexical pitch in another stress-accent language (i.e. German) benefits the naïve perception of non-native tones (i.e. Mandarin Chinese) as compared to a language without lexical stress or lexical tone (i.e. French) or even pitch accent (i.e. Japanese) (Braun, Galts and Kabak 2014). However, as lexical pitch is one of several features associated with word stress in English and the weakest to mark stress as noted above, it may be difficult to apply only lexical pitch to the perception of tone as seems to be the case in the production of pitch accent in Japanese (Kondo 2007).

Additionally, L1 English listeners tend to confuse similar tones in a comparable manner to L1 speakers of a tone language. When comparing Thai tones in an AX task, L1 English listeners have the most difficulty comparing flat vs. contour tones while the comparison of contour tones with one another is the easiest (Burnham et al. 1992), suggesting that the perception of pitch height is difficult for L1 English listeners as well. L1 English speakers also confuse the rising and dipping tones (tones 2 and 3, respectively) just as L1 Mandarin speakers do (Leather 1983, 1990, Li and Thompson 1977, Wang, Spence, Jongman and Sereno 1999). Guion and Pedersen (2007) also show that English speakers focus much more on pitch height than direction as compared to speakers of tone languages when perceiving synthetic Mandarin tones (as incidentally is the case for Japanese in their study as well).

3.2.4. Speakers of "Intonation-Only" Languages

Speakers of languages which do not employ lexically-contrastive pitch may apply experience with L1 intonation to the perception of tones. Simply put, intonation patterns like tone patterns are stored as categories (cf. Beckman, Hirschberg and Shattuck-Hufnagel 2005, Francis et al. 2008), enabling L1 speakers of intonation-only languages to access intonation patterns when perceiving tones or pitch features. For example, the intonation categories in English for questions appear to have been mapped onto L2 Mandarin tone 2 [35] (Francis et al. 2008, So and Best 2010) while that for statements was mapped onto tone 4 [51] (So and Best 2010).

The difference in domain usage (phrase level for intonation vs word level for tone) must also be considered such that L1 speakers of intonation-only languages would have difficulty mapping intonation patterns onto tone patterns. Additionally, all languages use intonation, potentially cancelling out any resulting difference in the perception of non-native tones influenced by L1 intonation. Moreover, it must be noted that some studies have shown that tone may not be perceived as categories by L1 speakers of intonation-only languages (French, Hallé et al. 2004).

3.2.5. Cross-Linguistic Perception

Comparing all four types of languages within the Feature Hypothesis (McAllister et al. 2002), we would expect differences in the perception of non-native tones by speakers of various non-tonal languages as non-tonal languages vary greatly in the degree and manner of usage of lexically-contrastive pitch while all employ intonation at the phrasal level. This can be attested to by a few studies which report that pitch accent language speakers (e.g. L1 Swedish, L1 Japanese) perform at comparable rates to L1 tone language speakers in their naïve perception of L2 tones (Burnham et al. 1996, So 2006). Again, as noted above, in learning Mandarin tones Japanese show much more improvement than L1 English speakers (McGinnis 1996). Moreover, native speakers of tone languages (i.e. Mandarin) perceive pitch accent patterns of pitch-accent languages (i.e. Swedish) more accurately than speakers of a non-tone language (i.e. Hindi, Eliasson 1997). This study concludes that "the linguistic category of tone is cognitively more salient to learners if it is present in a comparable form in their native language" as "transfer is a viable factor in tonology" (Eliasson 1997: 1274). This finding serves to further reinforce the hypothesis that accuracy in perceiving L2 lexically-contrastive pitch is commensurate to the relative functionality of lexically-contrastive pitch in the L1, i.e. tone language > pitch accent > word stress >

intonation only (i.e. no lexically-contrastive pitch) (Schaefer and Darcy 2014).

In sum, we might consider the functionality of L1 lexically-contrastive pitch in the naïve perception of non-native lexically-contrastive pitch in the form of features (Francis et al. 2008, McAllister et al. 2002) rather than in only terms of tone categories (i.e. PAM). That is, lexically-contrastive pitch usage may be broken down into features (e.g. pitch height, direction, etc.) and the resulting weighting of these features in an attempt to analyze the perception of lexically contrastive pitch. However, we could consider tone patterns as merely streamlined bundles of pitch features used by native speakers in perceiving tones more efficiently.

4. Methodology

4.1. Participants

Participants were recruited from five L1 language groups: Mandarin, Japanese, English, Korean, and Thai. The Thai speakers were recruited as a native-speaker control group. In total, 116 participants were recruited (Mandarin = 31, females = 25; Japanese = 23, females = 11; English = 24, female = 14; Korean = 29, females = 19; Thai = 9, females = 8). None of the non-Thai participants knew Thai nor had any knowledge of a tone language (except the Mandarin participants). The age ranges of each group were as follows: Thai: 32-41 (average age: 31.7; SD = 7), Mandarin: 18-37 (average age: 26.1; SD = 6.3), Japanese: 18-37 (average age: 28.8; SD = 5.2), English: 18-50 (average age: 31.2; SD = 9.1), Korean: 20-54 (average age: 31.3; SD = 7.9). Participants were either undergraduate or graduate students at Indiana University with a few exceptions. Each participant was recruited either through flyers distributed on campus or in courses or by word of mouth. All participants were paid $10 for participation in both the monosyllabic and disyllabic ABX tasks. All procedures were approved by the Indiana University Institutional Review Board.

4.2. Stimuli and Conditions

There were two types of stimuli for both tasks: Target and Control.

The target non-word stimuli varied by tone but not segment while the control non-word stimuli varied by segment but not tone. Stimuli were checked by native speakers of all the languages of the participants, i.e. Thai, Mandarin, Japanese, English and Korean, to ensure that the stimuli were non-words in all these languages. However, a few stimuli were deemed to be words in other Mandarin dialects or Taiwanese in the monosyllabic task. Additional distracters were not included.

4.2.1. Monosyllabic Stimuli and Conditions

For the target stimuli, an equal number of stimuli from each of the five Thai tones was used: low (L), mid (M), high (H), rising (R), and falling (F). The monosyllabic control items differed by only one segment, either a vowel or consonant, and were equally divided among the five possible Thai tones. Open syllable items were created as open-syllable CV words are considered more difficult to perceive than closed syllable words (Wayland and Guion 2003), keeping in consideration that CV structure is actually bimoraic as a glottal stop follows the vowel unless such CV syllables occur in the unstressed non-final position in polymoraic words (Bennett 1994).

Two target tone syllables and eight control syllables [two pairs varying by vowels (including monophthongs, diphthongs, or triphthongs) and two varying by consonants] were selected as stimuli. Each of the two tone target syllables carried each of the five Thai tones; these five tonal items were then paired to form all the ten tone comparisons possible: 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 syllable one, 10 with syllable two). For control items, the four syllable pairs also carried each of the five tones (both members of the pair always carried the same tone), resulting in 20 pairs of items. Table 3 presents an overview of the experimental items.

Each of these pairs was arranged in an experimental trial (a triplet) where one member of the pair (A, B) was repeated. This produced four triplets for each pair: ABA, ABB, BAB, BAA (The third token of the triplet is the X token). All four possible combinations of ABA/ABB/BAA/BAB were used to balance presentation and prevent bias. Thus, 40 pairs of items produced 160 experimental trials. Trial presentation was controlled by DMDX (Forster and Forster 2003).

Table 3. Stimuli, Conditions and Number of Trials in the ABX Monosyllabic Task

Test (tone) (80 trials**)		Control (segments) (80 trials**)		Training (segments only) (10 trials**)	
Tone comparisons	segments	Tones	Segment comparisons	Tones	Segment comparisons
Falling-Rising Low-Falling Low-Rising Mid-Falling Mid-Rising High-Falling High-Rising Low-Mid Low-High Mid-High	[noːj]* [pʰuəj]	Falling Rising Low Mid High	[beːw] - [teːw] [wiːə] - [tʰiːə] [uə] - [iə] [riːə] - [rɤːj]	Falling Rising Low Mid High	[wuːj] - [pʰuːj] [duːə] - [ŋiːn]

*Stimuli were created by two native speakers of Thai who are not linguists, but one reviewer who is a native speaker of Thai points out that some of the stimuli may not conform to tonal distributions/restrictions in Thai.

** Test condition = 10 tone types x 2 segments pairs x 4 orders (AAB, ABB, BBA, BAA) = 80 trials; Control condition = 5 tone types x 4 segment pairs x 4 orders (AAB, ABB, BBA, BAA) = 80 trials; training = 5 tone types x 2 segment pairs = 10 trials.

An interstimulus interval (ISI) of 500ms was inserted between each stimulus within a triplet (cf. Wayland and Guion 2003). A break was inserted every 40 trials, resulting in four blocks of 40 trials. The 40 trials were randomized within each block. Blocks were randomized with each other as well. Response time-out was set at 2500ms to speed up the task. Reaction times were measured from the start of the X token which is the last of the three tokens. In total, duration of the monosyllabic task was approximately 20 minutes.

Participants were given 10 trials as training. Each trial contained two syllable pairs carrying the five tones, resulting in 10 pairs of items, arranged as above into 10 triplets. The members of each pair in each trial carried the same tone. Thus, only differences in segments were used in the trials and no differences in tones to ensure that the participants understand how to do the task. Feedback was only provided during the training session. Each tone was heard 6 times each in the training session, and 96 times in the experimental trials. Therefore, in total in this task, each tone was heard 102 times.

4.2.2. Disyllabic Stimuli and Conditions

For disyllabic stimuli, each syllable carries a tone with a vowel being a monophthong, diphthong, or triphthong. The full matrix of pairings would be a possible 20 pairs of tones on each disyllabic item (i.e. LM, LH, LF, LR, ML, MH, etc.). Thus, considering the large number of possible combinations of pairings of the five tones, only five disyllabic combinations were selected: LM, LH, MH, HF, HL. These patterns were selected to test the potential mapping of L1 English word stress patterns onto the target stimuli reflecting English stress pitch patterns.

- The LH-HL comparison reflects word stress patterns in English as in inSERT vs INsert, differing in pitch direction.
- The LH-MH and LM-LH comparisons also reflect word stress with the same pitch direction but a differing magnitude of pitch change, i.e. low vs mid or mid vs high.
- The LM-HL comparison combines a change in both pitch direction and magnitude.
- The LM-MH comparison differs in the register of tones, i.e. LM in the lower register of the voice while MH in the higher.

Each disyllabic non-word had five pairings. In total, for two non-words and four presentation orders, there were 40 trials in the test (tone) condition. Forty trials for this test condition were created by using the five tone comparisons on two syllables (i.e. [duːə.pʰuːj], [kiŋ.kɛː]) with four different orderings (i.e. ABA, ABB, BAA, and BAB). Thus, for one of the LH-HL trials the disyllabic ABX task looks as follows with the tones broadly assigned to the syllables in these stimuli:

[duːə.p^{h}uːj]	[duːə.p^{h}uːj]	[duːə.p^{h}uːj]
L H	H L	L H
A	B	A

For the disyllabic control stimuli, the same patterns of HF, HL, LH, LM and MH were used. The stimuli varied in segments, either by only one vowel or one consonant as seen in Table 4 below. Again, the total number of trials included two pairs of disyllabic stimuli presented in the four possible ABX orderings of ABA, ABB, BAA, and BAB for five comparisons. This resulted in 40 possible trials (2 pairs of items x 4 orderings x 5 tone combinations). In total, there were 80 experimental trials in the disyllabic task with a break inserted between two blocks of 40. The disyllabic task mirrored the monosyllabic task in all other characteristics as note above. The disyllabic task required approximately 10 minutes. A trial for the control stimuli on the disyllabic ABX task looks as follows with again the tones broadly assigned to the syllables in these stimuli:

[teːw.fiːŋ]	[teːw.foːŋ]	[teːw.fiːŋ]
M H	M H	M H
A	B	A

In the training session, the trials compared segments with the same tone combination for the A, B and X stimuli. The training items basically mirrored those of the control condition using the same tone combinations of HF, HL, LH, LM and MH, but with the A and B items differing greatly in terms of segments, e.g., [no:beːw] vs [luəj.p^{h}uəj]. Each of the five tone combinations was heard 6 times each in a total of 10 trials. In sum, the total number of times a tone combination was heard in the disyllabic task was LM x 66, LH x 66, MH x 54, HL x 54, HF x 30 (total = 270). Thus, participants heard 540 tone samples. They heard the low tone 186 times; the mid tone for 120 times; the high tone for 204 times; and the falling tone for 30 times. The times the listeners heard each tone was controlled as best as possible to prevent possible bias.

Table 4. Stimuli, Conditions and Number of Trials in the ABX Disyllabic Task

Test (=tone) (40 trials**)		Control (=segments) (40 trials**)		Training (=segments only) (10 trials**)	
Tone comparisons	Segments	Tone combinations	Segment comparisons	Tone combinations	Segment comparisons
Low+High - High+Low		High+ Falling		High+ Falling	
Low+Mid - High+Low	[duːə.pʰuːj]	High+Low	[pʰuːj.wuːj] - [ruːj.wuːj]	High+Low	[no:beːw] - [luːj.pʰuəj]
Low+High - Mid+High	[kiŋ.kɛː]*	Low+High		Low+High	
Low+Mid - Low+High		Low+Mid	[teːw.fiːŋ] - [teːw.foːŋ]	Low+Mid	[ro:beːw] - [luəj.pʰuəj]
Low+Mid - Mid+High		Mid+High		Mid+High	

* Stimuli were created by two native speakers of Thai who are not linguists, but one reviewer who is a native speaker of Thai points out that some of the stimuli may not conform to tonal distributions/restrictions in Thai.
** The tone test condition = 5 tone types x 2 segments x 4 orders (AAB, ABB, BBA, BAA) = 40 trials; the control test condition = 5 tone types x 2 segments x 4 orders (AAB, ABB, BBA, BAA) = 40 trials; training = 5 tone types x 2 segments = 10 trials.

4.3. Speakers for Thai Stimuli and Elicitation Method

Two female voices of the Central Thai dialect recorded the stimuli. The speaker for the A and B tokens was a 28-year old speaker from the Bang Phlat (บางพลัด) district of Bangkok, Thailand, while the speaker for the X tokens was a 25-year old speaker from the Min Buri (มีนบุรี) district of Bangkok, Thailand. Both self-reported being speakers of standard Thai and were graduate students at prominent universities in Bangkok. One voice was used for the X token and the other voice for the A and B tokens in order to add more difficulty to the task than using one voice for all three tokens. Two female voices were used to prevent any possible difficulty in comparing voice height between a male (low) and female (high) voice.

Recordings were made in a soundproof room with the aid of a professional sound technician. The Thai stimuli were both recorded using a Shure SM7B microphone. A Shure M267 4-channel mixer was used to sum all the signals or in this case amplify signals from a dynamic microphone. A Motu 828 (audio interface) converter was used to convert analog to digital signals. A G5 Mac using Peak LE 5.2 digital audio workstation (i.e. a computer plus Peak LE audio software) was used to record and make edits on the stimuli. All files were recorded at 48kHz/24bit and left unprocessed. Thai stimuli were recorded on a 2-track stereo channel with 2 microphones, and then the two tracks were split into separate mono files. A Peak LE 6 audio editor was used to export dual mono tracks from the original stereo file. They then were spliced into individual wav. files using a program in Praat (Boersma and Weenink 2015). Finally, the wav. soundfiles from both speakers were normalized for amplitude so that all items were comparable in loudness. The stimuli were normalized using Audition software at 13.3Hz for the monosyllabic items and 13Hz for the disyllabic items.

Tone patterns and their labels were checked using Praat to determine by visual inspection and by ear whether the tones were indeed the targeted tones. Soundfiles were also checked for clear splicing to ensure that final or initial sounds of the items were not cut off.

4.4. Procedure

Participants were tested individually in a quiet psycholinguistics laboratory on the campus of Indiana University. After arriving in the lab, participants first read and signed the consent form. Each participant sat in front of a personal computer, wearing high-quality headphones. They listened to the auditory stimuli at a self-selected comfortable volume level.

For both the monosyllabic and disyllabic task, the experimental session started with the instructions presented on the screen. Participants were instructed to listen to the triplets to decide whether the last token (X) was more similar to the first (A) or the second one (B). They indicated their response by pressing a clearly labelled key (A or B) on the computer keyboard. The researcher sat near the participant as they did the training session, and asked whether they understood what was expected of them. Participants were allowed to ask questions to clarify. They then started the experimental blocks.

After finishing the monosyllabic ABX task, the participant was asked to fill out a questionnaire. After this, the participant then proceeded to the disyllabic ABX task. When the participants finished the disyllabic AXB task, they were debriefed, i.e. asked if they had any comments or questions about the two AXB tasks. They were then thanked and paid a small fee for their participation. The entire procedure generally required between 45-60 minutes in total.

5. Results

Participants were screened. Two were removed from the analysis as they had extensive exposure to Chinese (i.e. Mandarin, Cantonese) while another two were removed as they performed beyond two standard deviations of the accuracy mean on the control condition. The final number of participants in each group is as follows: total (n = 112) Thai (n = 9, female = 8), Mandarin (n = 30, female = 24), Japanese (n = 23, female = 11), English (n = 23, female = 14) and Korean (n = 27, female = 18).

After cleaning the data, language groups were compared in terms of accuracy and reaction times (using SPSS) on overall performance on the tone (test) condition versus the segmental (control) condition and the ten individual tonal comparisons.

For accuracy data, since the data structure is categorical (1 vs 0 for correct vs incorrect answer), a Generalized Estimating Equations (GEE) model for a binary response was fitted to the data, declaring subjects as a random factor. For most analyses, Language (Thai, Mandarin, Japanese, English, Korean) and Condition (test vs control) or Subcondition (10 comparisons) are declared as fixed factors, and Subjects are declared as a random factor. Sidak correction for multiple comparisons was used over Bonferroni as it is a less conservative method to calculate significance when conducting multiple comparisons; otherwise, considering the multiple comparisons needed over several groups and many tonal comparisons, the chances of revealing true significance would be greatly diminished and therefore, not accurately reflect the true situation. Because item variability was constrained by the experiment design and stimulus construction rather than randomly sampled, and in an effort to analyze the data with the simplest possible model, only subject variability was declared as random effect (Raaijmakers, Schrijnemakers and Gremmen 1999).

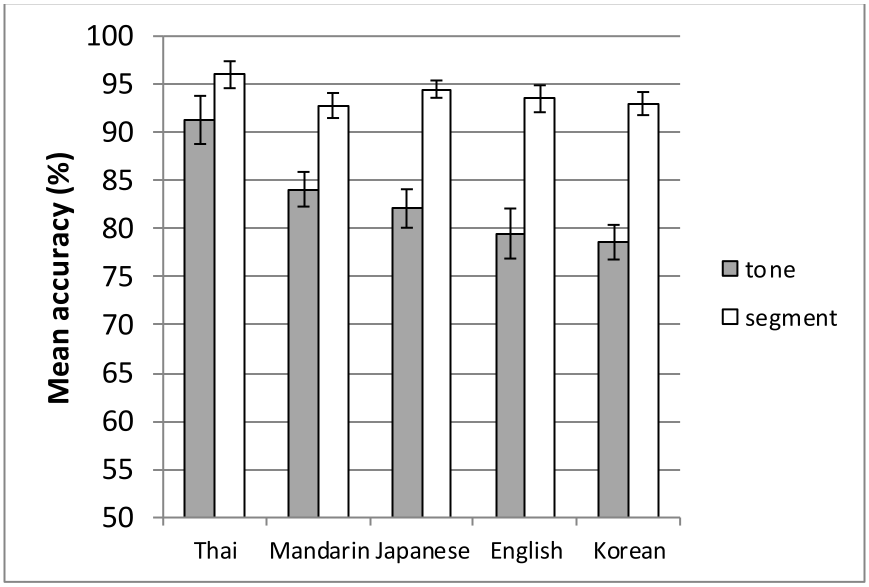
Reaction times were also measured and analyzed but will not be discussed in this paper as they seemingly contribute little to answering the research questions.

5.1. Analysis by Condition for Monosyllabic Task: Test and Control

Raw scores for mean accuracy rates on the tone test condition of the monosyllabic ABX task in Figure 1 show that the Thai group perform the most accurately (M = 91.3%) and the Korean group perform the least accurately (M = 78.6%). The three language groups for Mandarin, Japanese and English perform at "in-between" scores (M = 84.1%, M = 82.1%, M = 79.5%, respectively). In contrast, accuracy scores for the segment control condition range from the Thai group at the highest (M = 96%) to the English group at the lowest (M = 93.5%). The focus of results, however, is on the four groups other than the Thai group as the Thai are expected to perform well given that the target tones are native. Additionally, the Thai group also serves as a type of control for comparison and to validate the tasks.

The non-aggregated data for accuracy were used for statistical analysis. For accuracy data, a GEE model was fitted with the fixed factors Language (Thai, Mandarin, Japanese, English, Korean) and Condition (test vs control) and Subjects as a random factor. The Type III tests of fixed effects exhibited no main effect of Language (χ^2 (4) = 5.34, p = 0.25), a significant effect of Condition (χ^2 (1) = 268.7, p < 0.001), and a significant interaction between the two factors (χ^2 (4) = 9.85, p = 0.043) [where a p-value equal to or below 0.05 signifies statistical significance, cf. Fisher 1925].

Figure 1. Mean Accuracy Rates (%) for Each Language Group on Test and Control Conditions.



Error bars represent +/-1 SE.

The interaction shows that while groups do not differ on the control condition, their performance varied on the test condition (see Table 5). There were no statistical differences for the control condition, i.e. perception of segments. However, overall test results revealed that performance differed significantly between groups on the test condition ($\chi^2(4) = 19.24$, $p = 0.001$). On the test condition for tones, the native Thai speakers performed significantly better than the English ($p = 0.011$), Korean ($p < 0.001$) and Japanese ($p = 0.038$). The Thai group did not perform statistically better than the Mandarin group ($p = 0.278$). In turn the Mandarin group did not perform statistically better than the other language groups.

5.2. Analysis by Subcondition for Monosyllabic Task

Mean scores on accuracy rates for all the test tone and control segment subconditions across the language groups are as follows. Mean accuracy rates across all language groups for the control segment subconditions range from a low of 89.9% by the Korean group on the control subcondition using a rising tone on all three items (i.e. A, B, X) to a high of 97.9% by the Thai group on the control subcondition using a high tone on all three items. On all the segment control subconditions, accuracy rates were relatively similar across groups.

In comparison, the test tone subconditions showed greater variety: The accuracy rates varied from 62.5% by the Korean group on the Low vs Rising comparison to a high of 97.2% by the Thai group on the Mid vs Rising comparison. The overall trend for the tone test subconditions across all groups was as follows: The Thai group performed at the highest accuracy rates for almost all the tone test subconditions while the English and Korean groups performed with the least accuracy for almost all the tone test subconditions with the Koreans generally less accurate than the English group.

Looking at the accuracy rates for each tone subcondition for each language group in Table 5 and Figure 2, we are able to discern two possible trends. First, perceptual accuracy appears to be influenced by the type of the two tones compared. For example, "direction" tone comparisons (i.e. comparing two direction tones) appear to be the easiest (e.g. Falling vs Rising) while "height" tone comparisons appear to be the most difficult (e.g. Low vs Mid). The "mixed" tone comparisons comparing tone height (e.g. low tone, mid tone) with tone direction (e.g. falling, rising) appear to fall between the "direction" and "height" tone comparisons in terms of accuracy rates. In contrast, accuracy rates on the control segment subconditions (Figure 3) do not show any hierarchical trend, i.e. flat across all subconditions and languages. Second, certain tone comparisons were less accurate universally, e.g. Low vs. Mid. At this point, it should also be noted that while height is relative being influenced by gender and voice quality, native speakers have demonstrated the ability to quickly and accurately gauge the voice range of multiple unknown speakers of both genders (Lee 2009).

To examine the accuracy data on individual tone comparisons, a GEE model was fitted with the fixed factors Language (Thai, Mandarin, Japanese, English, Korean) and Subcondition (individual tonal comparisons for both the tone and segment subconditions), with Subjects as a random factor. Examining individual tonal comparisons, the Type III tests of fixed effects showed a marginal effect of Language ($\chi^2(4) = 7.95$, $p = 0.094$), a significant effect of subcondition ($\chi^2(14) = 689.56$, $p < 0.001$), and a significant interaction between the two factors ($\chi^2(56) = 262.69$, $p < 0.001$).

The interaction suggests that the performance of the groups varies on the different subconditions (see Table 6). Overall test results revealed that performance differed significantly between groups on the some test subconditions ($\chi^2(4) = 11.622 \sim 32.555$, $p < 0.05$), i.e. High vs Falling ($p = 0.004$), Low vs Mid ($p = 0.002$), Low vs Rising ($p < 0.001$), Mid vs High ($p < 0.001$), Mid vs Rising ($p = 0.002$) and the segment condition bearing the High tone ($p = 0.02$). The Thai group performed significantly better than other groups on only four of the ten individual tone comparisons as seen in Table 6.

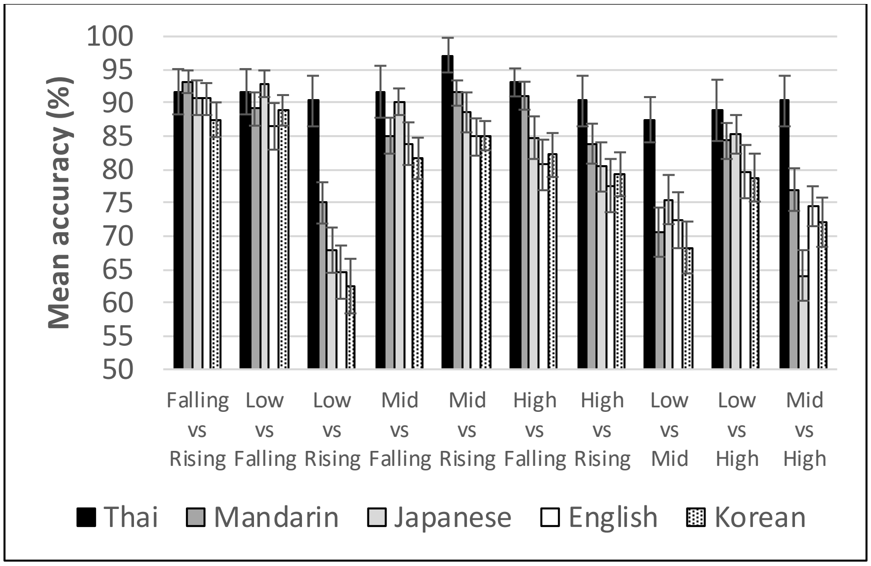
Table 5. Mean Accuracy Rates for Each Group on Each Tonal Comparison

Test sub-conditions	Mean accuracy (SE)				
	T	M	J	E	K
F vs R	91.7 (3.4)	93.2 (1.7)	90.8 (2.6)	90.6 (2.6)	87.5 (2.6)
L vs F	91.7 (3.4)	89.1 (2.5)	92.9 (2.0)	86.5 (3.7)	88.9 (2.3)
L vs R	90.3 (3.8)	75.0 (3.1)	67.9 (3.4)	64.6 (4.3)	62.5 (4.1)
M vs F	91.7 (3.9)	85.1 (2.7)	90.2 (2.0)	83.9 (3.4)	81.7 (3.1)
M vs R	97.2 (2.6)	91.5 (1.9)	88.6 (3.0)	84.9 (3.0)	85.1 (2.2)
H vs F	93.1 (2.1)	91.1 (2.1)	84.8 (3.2)	80.7 (4.1)	82.2 (3.3)
H vs R	90.3 (3.8)	83.9 (3.0)	80.4 (3.7)	77.6 (4.2)	79.3 (3.3)
L vs M	87.5 (3.4)	70.6 (3.7)	75.5 (3.7)	72.4 (4.4)	66.3 (3.9)
L vs H	88.9 (4.6)	84.3 (2.7)	85.3 (2.9)	79.7 (4.4)	78.8 (3.6)
M vs H	90.3 (3.8)	77.0 (3.2)	64.1 (3.8)	74.5 (3.2)	72.1 (3.7)
Control sub-conditions					
F	95.1 (1.3)	93.5 (1.5)	94.6 (1.3)	93.5 (1.7)	95.4 (1.4)
R	96.5 (2.2)	91.1 (1.8)	93.2 (1.4)	93.5 (1.5)	89.9 (1.8)
L	94.4 (2.5)	92.5 (1.5)	95.1 (0.9)	93.3 (1.2)	93.3 (1.2)
M	95.8 (1.7)	94.0 (1.0)	94.8 (1.2)	92.7 (2.0)	95.2 (1.3)
H	97.9 (1.0)	92.7 (1.8)	94.8 (1.4)	94.8 (1.4)	91.1 (2.0)

Note: SE = standard error.

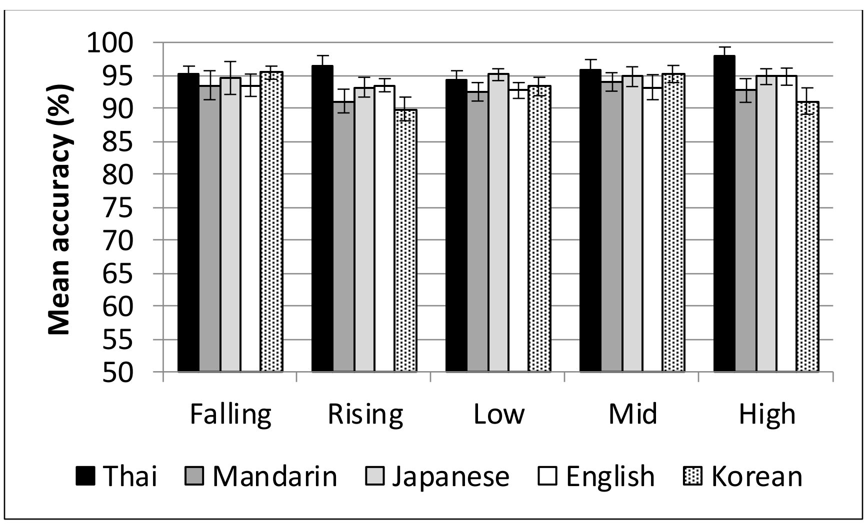
T = Thai; M = Mandarin, J = Japanese, E = English, K = Korean; L = low tone, M = mid tone, H = high tone, F = falling tone, R = rising tone.

Figure 2. Accuracy Rates (%) for Each Language Group on Each Test Subcondition.



Error bars represent +/- 1 SE.

Figure 3. Accuracy Rates (%) for Each Language Group on Each Control Subcondition.



Error bars represent +/- 1 SE.

Among the control subconditions, the Thai group is more accurate than the Korean group on the segment trials using the High tone (p = 0.033). Otherwise, there is no significant difference in accuracy on the control subconditions between the language groups and so, are not reported in Table 6.

Table 6. Statistically Significant Accuracy Rates on Each Tonal Comparison (i.e. subcondition) for Each Group

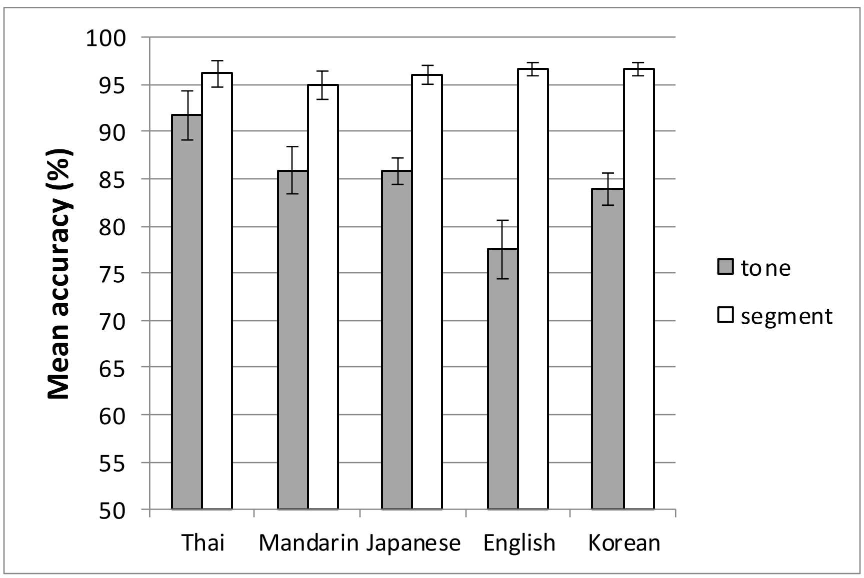
Comparison	Accuracy
Tone overall	Thai > English (p = 0.011) Thai > Japanese (p = 0.038) Thai > Korean (p < 0.001)
FR	--
LF	--
LR	Thai > English (p < 0.001) Thai > Korean (p < 0.001) Thai > Japanese (p < 0.001) Thai > Mandarin (p = 0.022)
MF	--
MR	Thai > English (p = 0.021) Thai > Korean (p = 0.007)
HF	--
HR	--
LM	Thai > Korean (p = 0.004) Thai > Mandarin (p = 0.014)
LH	--
MH	Thai > English (p = 0.014) Thai > Japanese (p < 0.001) Thai > Korean (p = 0.012)

5.3. Analysis by condition for disyllabic task: Test and control

Mean accuracy scores for each group on the disyllabic ABX task are presented in Figure 4 and Table 7 below. They show the Thai performed the most accurately on the test tone condition (M = 91.7%) while the English group performed at the least accurate level

(*M* = 77.5%). The Japanese, Mandarin and Korean groups performed at levels in between the Thai and English groups at *M* = 85.8%, *M* = 85.9%, *M* = 83.9%, respectively. On the control (segment) condition, accuracy rates are similarly high for all the language groups (*M* = 90 percentile). The range of accuracy rates for the control condition (all scores were approximately 96%) is smaller than that for the test tone condition (from a low of 77.5% to a high of 91.7%).

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



Error bars represent +/-1 SE.

The non-aggregated data for accuracy were used for statistical analysis. To examine the accuracy data, a GEE model was fitted, declaring the fixed factors Language (Thai, Mandarin, Japanese, English, Korean) and Condition (test vs control), with Subjects as a random factor. The Type III tests of fixed effects exhibited no main effect of Language ($\chi^2(4) = 3.02$, $p = 0.55$), a significant effect of Condition ($\chi^2(1) = 165.2$, $p < 0.001$), and a significant interaction between the two factors ($\chi^2(4) = 10.01$, $p = 0.040$).

The interaction shows that while groups do not differ on the control condition, their performance varies on the test condition where there is a main effect of Language (See Table 7). Overall test results revealed that performance differed significantly between groups on the test condition ($\chi^2(4) = 13.33$, $p < 0.010$), where the native Thai speakers performed significantly better than only the English group ($p = 0.004$) but not significantly better than the other groups. In contrast, all groups performed at statistically comparable levels on the control condition.

5.4. Analysis by Subcondition for Disyllabic Task

Mean scores on accuracy rates for all the test tone and control segment subconditions across the language groups are as follows. Table 7 and Figures 5 and 6 present the individual accuracy means for the test and control items in each subcondition for each group. Accuracy rates for the test condition range from a low of 63% to a high of 97.2%. There are some universal tendencies. For example, the Low+Mid vs High+Low comparison is the highest score for almost all the groups. On the other hand, the Low+High vs Mid+High comparison is the lowest score for all the groups. In contrast, some accuracy rates are language specific. For example, the Japanese performed least accurately among the groups on the Low+Mid vs Low+High comparison.

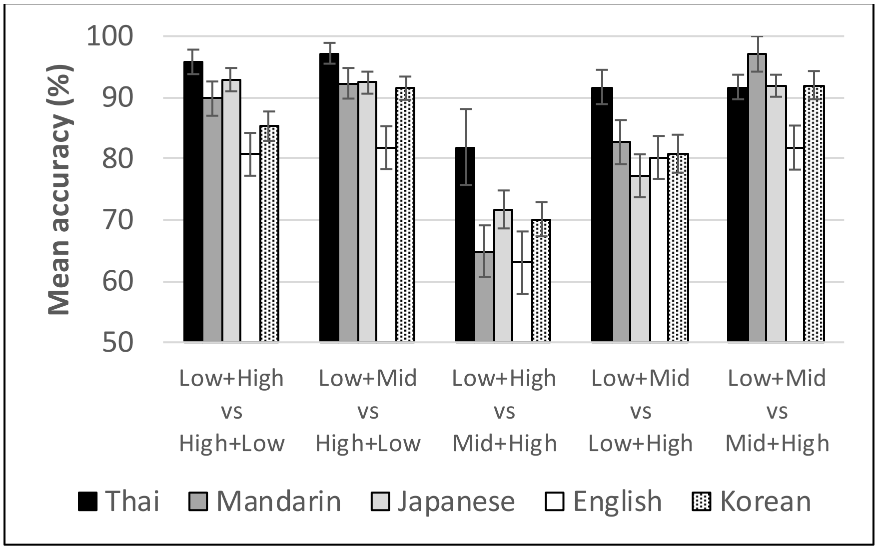
To examine the accuracy data on individual tone comparisons, a GEE model for a binary response was fitted with the fixed factors Language (Thai, Mandarin, Japanese, English, Korean) and Subcondition (individual tonal comparisons for both the tone and segment subconditions), with Subjects as a random factor. Examining individual tonal comparisons, we see that the Type III tests of fixed effects exhibited no main effect of Language ($\chi^2(4) = 3.884$, $p = 0.422$), a significant effect of Subcondition ($\chi^2(9) = 443.771$, $p < 0.001$), and a significant interaction between the two factors ($\chi^2(36) = 155.482$, $p < 0.001$). In Table 8 we can see in which subconditions there was a statistically significant difference among the groups.

Table 7. Mean Accuracy Rates (%) for Each Group on Each Tonal Comparison

Test subconditions	Mean accuracy (SE)				
	T	M	J	E	K
LH vs HL	95.8 (2.0)	89.8 (2.8)	92.9 (1.9)	80.7 (3.5)	85.3 (2.4)
LM vs HL	97.2 (1.7)	92.3 (2.5)	92.4 (1.8)	81.8 (3.5)	91.5 (1.9)
LH vs MH	81.9 (6.2)	64.9 (4.2)	71.7 (3.1)	63.0 (5.1)	70.1 (2.8)
LM vs LH	91.7 (2.8)	82.7 (3.6)	77.2 (3.5)	80.2 (3.5)	80.8 (3.1)
LM vs MH	91.7 (2.0)	97.1 (2.9)	91.9 (1.8)	81.8 (3.6)	92.0 (2.3)
Control subconditions					
HF	95.8 (2.0)	95.2 (1.9)	96.2 (1.2)	94.8 (1.8)	99.6 (0.4)
HL	97.2 (2.6)	95.6 (1.5)	97.8 (1.0)	96.9 (1.1)	94.2 (1.3)
LH	95.8 (2.0)	94.8 (2.5)	92.9 (2.2)	97.9 (1.0)	96.4 (1.3)
LM	94.4 (2.1)	94.4 (1.9)	96.2 (1.6)	95.8 (1.4)	96.9 (1.2)
MH	97.2 (1.7)	94.8 (1.4)	96.7 (1.6)	97.4 (1.0)	95.5 (1.6)

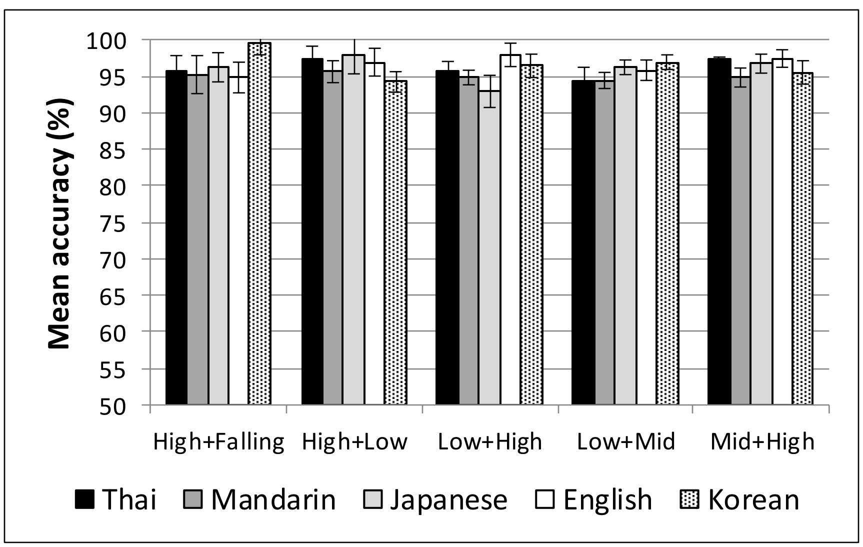
Note: SE=standard error.
T = Thai; M = Mandarin, J = Japanese, E = English, K = Korean; L = low tone, M = mid tone, H = high tone, F = falling tone, R = rising tone. Hence, LM = Low tone + Mid tone disyllabic word, LH = Low tone + High tone disyllabic word, MH = Mid tone + High tone disyllabic word, HL = High tone + Low tone disyllabic word, HF = High tone + Falling tone disyllabic word.

Figure 5. Mean Accuracy Rates (%) for Each Language Group on Each Test Subcondition



Error bars represent +/-1 SE.

Figure 6. Mean Accuracy Rates (%) for Each Language Group on Each Control Condition.



Error bars represent +/-1 SE.

The interaction again demonstrates that the performance of the groups varies on the test subconditions but not the control subconditions. For accuracy rates on individual tone comparisons, we see in Table 7 that the Low+High vs High+Low and the Low+Mid vs High+Low show the greatest differences between language groups. On the control segment conditions, none of the groups performed more accurately than the other groups in terms of statistical significance and therefore, no statistical results for the control condition are listed in Table 8.

Table 8. Statistically Significant Accuracy Rates on Each Tonal Comparison (i.e. subcondition) for Each Group

Comparison	Accuracy
Tone overall	Thai > English (p = 0.004)
LH-HL	Thai > English (p = 0.002) Japanese > English (p = 0.03) Thai > Korean (p = 0.008)
LM-HL	Thai > English (p = 0.001) Mandarin > English (p = 0.004) Japanese > English (p = 0.031)
LH-MH	--
LM-LH	Thai > Japanese (p = 0.023)
LM-MH	--

6. Discussion/Conclusions

A hierarchy of the naïve perception of non-native tones on monosyllabic stimuli as shaped by lexical pitch in the L1 did not bear out as clearly in the current study as defined in the previous study (Schaefer and Darcy 2014). However, the following hierarchy is suggested by the results of the current study: L1 Mandarin > L1 Japanese > L1 Korean ≥ L1 English. That is, the Thai group's overall performance exceeded with statistical significance that of the three groups of Japanese, English and Korean, but not that of the Mandarin group. This indirectly suggests that L1 speakers of tone languages (as strictly defined, i.e. excluding speakers of a pitch accent language like Japanese) performed more accurately than the non-tonal language speakers. Furthermore, when looking at the individual results on the same task for the 10 tone comparisons in the monosyllabic task, we see that the Thai group outperformed the other groups to varying degrees with statistical significance: two comparisons against the Mandarin group, two against the Japanese group, three against the English group, and four against the Korean group (See Table 9). .This may also indirectly suggest a hierarchy of perception as shaped by lexical pitch in the L1. In contrast, as expected all groups performed at higher rates of accuracy on the overall and individual control segmental comparisons with no statistical differences. The difference between results for the target and control stimuli demonstrate that the naïve perception of non-native tone differs and that this difference likely emerges from the varying functionality of lexical pitch in the L1.

Table 9. Statistically Significant Performances on Monosyllabic Task

	Thai	Mandarin	Japanese	English	Korean
Overall			T	T	T
Falling vs Rising					
Falling vs Low					
Falling vs Mid					
Falling vs High					
Rising vs Low		T	T	T	T
Rising vs Mid				T	T
Rising vs High					
Low vs High					
Low vs Mid		T			T
Mid vs High			T	T	T

Single capital letters indicate the L1 language group that outperformed the language at the top of the column. T = L1 Thai group.

Nevertheless, there may be several reasons as to why the hierarchy does not bear out so clearly in the current study. The target tonal stimuli in the current study used two strings of segments to carry the tones, namely, the non-words of [no:j]] and [p^{h}uəj] while the previous Schaefer and Darcy (2014) study was less conventional in using more segmentally varied target stimuli, i.e. 16 strings of segments: [ba:], [pu:], [tʃʰa:], [di:], [hu:], [kʰa:j], [ma:], [ma;j], [mi:], [na:], [pʰi:], [ru:], [su:], [tʰa:], [wa:], and [ja:], half of which were real words in Thai. Thus, the previous study may have added a greater cognitive load to perception, potentially teasing the performance of groups apart to a greater degree.

Additionally, the L1 Korean speakers in the current study spoke English to varying degrees which might be presumed to allow them to perform at the levels of L1 English speakers. However, the previous study (Schaefer and Darcy, 2014) ruled out this possible influence: This previous study showed that the presence of L1 speakers of the pitch accent Kyungsang dialect of Korean and not English proficiency boosted the overall performance of L1 Korean group to the level of the L1 English group. An analysis of the L1 Korean group in the current study by pitch accent/non-pitch accent dialects (but not reported in this article) did not result in the predicted hierarchy of English > Korean (Schaefer 2015). In sum, in regard to the first research question as to whether the varying role of lexical pitch in the native language (i.e. pitch accent, word stress) shapes the naïve perception of another non-native lexical pitch system (i.e. Thai), the answer is affirmative although not as robust as the previous study (Schaefer and Darcy 2014).

In regard to the second research question, disyllabic stimuli did not appear to aid L1 English and L1 Japanese speakers in their perception of non-native lexical pitch. Indeed, the results on the disyllabic task are less uniform than those on the monosyllabic task. On overall results, the Thai group only outperformed the English group. On individual comparison results, the Thai group outperformed the English group on two comparisons, the Japanese group on one, and the Korean group on one. Additionally, the Japanese group outperformed the English group on two comparisons while the Mandarin group outperformed the English group on one comparison (see Table 10). As such, we see that the English group performed less accurately than the other groups in the perception of disyllabic stimuli. Raw overall scores indicate that unlike performance on monosyllabic stimuli the English group was outperformed by the Korean group which performed at approximately the same accuracy level as that of the Mandarin group. Thus, disyllabic stimuli do not appear to aid the English group in better accessing the pitch correlate of word stress in their perception of non-native lexical pitch. Disyllabic stimuli do not clearly appear to aid the Japanese group either in accessing pitch accent in their perception. Nevertheless, disyllabic stimuli appear to shed light on perception or categorization of the level tones of low, mid, and high tones as discussed below.

Table 10. Statistically Significant Performances on Disyllabic Task

	Mandarin	Japanese	English	Korean
Overall			T	
Low+High vs High+Low			T, J	T
Low+Mid vs High+Low			T, M, J	
Low+High vs Mid+High				
Low+Mid vs Low+High		T		
Low+Mid vs Mid+High				

Single capital letters indicate the L1 language group that outperformed the language at the top of the column. T = L1 Thai group, M = L1 Mandarin group, J = L1 Japanese group.

There are several universal tendencies regardless of L1. As noted, all groups perform better on segmental comparisons than tonal comparisons. The falling tone seems salient to all groups such that any comparison involving a falling tone was perceived with relatively high accuracy. We would expect this for L1 Mandarin speakers as this tone maps onto tone 4 in Mandarin (i.e. falling tone) and for L1 Japanese as the fall in pitch is said to mark accent in Japanese (Sugiyama 2012). Yet, all groups performed well. There was not an issue with the high vs rising comparison despite the two appearing to have the same physical shape under a unitary model and therefore, differing only by height.

Language-specific tendencies also appear. For example, all groups but the Thai performed poorly on Low vs. Rising, Low vs. Mid, and Mid vs. High comparisons, but at varying levels of performance. The overall poor performance for all groups in perceiving the Low vs Mid tone comparison was expected as this has been established in research (Abramson 1976). Namely, the difference in height appears difficult to gauge. However, the Low vs Mid tone comparison posed more difficulty for the Mandarin and Korean L1 groups than other speaker groups. The L1 Mandarin speakers were expected to have difficulty with the mid level and low level tones as there are no counterparts to these tones among Mandarin tones unlike the Thai High, Rising, and Falling tones which could be mapped onto Mandarin counterpart tones (cf. PAM, Best 1995) as nearly "complete tones" (cf. the Unitary Model). The overall performance by L2 Mandarin speakers is expected as Mandarin speakers weight direction over height to distinguish native Mandarin tones, but height is the feature that differentiates the Low and Mid tones.

Furthermore, we see that the Japanese group underperformed all the other groups on the Low+Mid vs Low+High comparisons with statistical significance against the L1 Thai group. This suggests that the Japanese group may have difficulty in perceiving the difference between a mid and high pitch (cf. the compositional model) which is the only difference between these two disyllabic stimuli. That is, the mid and high pitches may be categorized as being the same, i.e., both "high" pitches, or as merely having pitch (vs. no pitch for the Low tone). This perspective is reinforced by the performance of the Japanese on the monosyllabic Mid vs. High tones which is their least accurate performance among all the monosyllabic comparisons. This performance was unexpected in contrast to the Low vs. Mid tone comparison which was expected as it has been noted to be difficult to perceive. This latter comparison was further reinforced by difficulty in differentiating between Low+High vs. Mid+High disyllabic comparison which requires listeners to differentiate between an initial low vs mid pitch. As such, the L1 Japanese appear to have difficulties in perceiving the height of the Mid tone when juxtaposed against either the High or Low tones. In contrast to these two comparisons, the Low+Mid vs. Mid+High comparison was relatively easy to differentiate as the tones occur in two different registers, i.e. one in the lower and one in the higher, or exhibit a larger magnitude of difference, i.e. low vs high. Thus, for Japanese listeners the mid level tone presents an issue as it appears to sit at the border between the lower and higher register, suggesting that tones are perceived as either being high or low or merely having a pitch, i.e. high pitch.

In comparison, the disyllabic pitch patterns that reflect English word stress appear not to aid the L1 speakers of English in perceiving differences between rising and falling pitch patterns. Indeed, as noted the L1 English group underperformed speakers of standard Korean which does not feature lexical pitch on many of the disyllabic patterns. This relatively lower performance by the L1 English group vis-à-vis the L1 Korean group on the disyllabic stimuli can be likely attributed to the finding that the feature of pitch accompanying the presence/absence of word stress is less salient than the other features of vowel length, vowel quality, and intensity. Thus, word stress in English appears to be equal to having no contrastive lexical pitch in the L1, e.g. standard Korean, although the superior performance by the Korean group on the disyllabic stimuli remains unaccounted for. In short, disyllabic stimuli did not validate the hypothesis that L1 English speakers would be able to more easily perceive pitch when pitch combinations mirror the salient usage in English juxtaposing a stressed (i.e. higher pitch) and an unstressed (i.e. lower pitch) syllable against one another. However, the poor performance of the L1 English group on the Low+High vs Mid+High comparison in contrast to the higher but flat performance on the other four comparisons may indicate that lexical pitch in stress is still somewhat salient in English and merely weaker relative to its salience in the other languages. Indeed, the perceptual difficulty of this one comparison may stem from both the mid pitch and low pitch being potentially mapped onto unstressed syllables in English. Additionally, most of the L1 Koreans have exposure to English as an L2 which may have boosted their perception of pitch if the correlate of pitch for stress is the most salient feature to such learners.

In summary, the varying role of lexical pitch in the native language (e.g. pitch accent, word stress) appears to shape the naïve perception of another non-native lexical pitch system (i.e. Thai). That is, the hierarchy established in the previous study (Schaefer and Darcy 2014) bears out for the perception of monosyllabic stimuli although less robustly so. However, speakers of Japanese and English do not necessarily perceive disyllabic stimuli more accurately than monosyllabic stimuli. These mixed results indicate that the overall picture of the influence of L1 lexical pitch on the naïve perception of non-native lexical pitch is not so straightforward but rather fairly nuanced. As such, performance on the perception of monosyllabic versus disyllabic stimuli offers new insights, both universal and language-specific.

NOTES

[1] I would like to express my gratitude to the following people for their guidance and support of this research: Professors Isabelle Darcy, Rex Sprouse, Laurent Dekydtspotter, Öner Özçelik, Charles Lin, Kathleen Bardovi-Harlig and the Department of Second Language Studies at Indiana University. I further thank Wannaporn Tongtako, Pawanrat Rattana, Janjarus Karachot, Natasha Branch, Stephanie Dickinson, and the Indiana Statistical Consulting Center (ISCC). I am also very grateful to the College of Arts and Sciences at Indiana University for a Louise McNutt Graduate Dissertation Year Research Fellowship and to the journal of *Language Learning* for a Dissertation Grant to support this research. I also thank the many participants. Additionally, I would like to thank one reviewer who offered immensely insightful comments, particularly concerning the Thai stimuli.

[2] Tone marks for Mandarin reflect pinyin usage and not IPA notation while tone marks for Thai adopt IPA notation with the exception of the mid tone which is left unmarked (vs nā: in IPA).

REFERENCES

ABRAMSON, ARTHUR. S. 1975. The tones of Central Thai: Some perceptual experiments. Studies in Thai Linguistics in Honor of William J. Gedney, ed. by Jimmy G. Harris and James R. Chamberlain, 1-16. Bangkok: Central Institute of English Language.

_____ . 1976. Thai tones as a reference system. Tai Linguistics in Honor of Fang-Kuei Li, ed. by Thomas W. Gething, Jimmy G.

Harris and Pranee. Kullavanijaya, 1-12. Bangkok: Chulalongkorn University Press.

_____ . 1978. Static and dynamic acoustic cues in distinctive tones. Language and Speech 21: 319-325.

BECKMAN, MARY, JULIA HIRSCHBERG and STEFANIE SHATTUCK-HUFNAGEL. 2005. The original ToBI system and the evolution of the ToBI framework. Prosodic Typology: The Phonology of Intonation and Phrasing, ed. by Sun-Ah Jun, 9-54. Oxford: Oxford University Press.

BENNETT, J. FRASER. 1994. Iambicity in Thai. Studies in the Linguistic Sciences 24 (1/2): 39-57.

BEST, CATHERINE T. 1995. A direct realist view of cross-language speech perception. Speech Perception and Linguistic Experience: Issues in Cross-language Research, ed. by Winifred Strange, 171-204. Baltimore, MD: York Press.

BOERSMA, PAUL and DAVID WEENINK. 2015. Praat: Doing phonetics by computer [Computer program]. Version 5.4.08, http://www.praat.org/

BRAUN, BETTINA, TOBIAS GALTS and BARIŞ KABAK. 2014. Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. Second Language Research 30 (3): 323-350.

BURNHAM, DENIS, ELIZABETH FRANCIS, DI WEBSTER, SUDAPORN LUKSANEEYANAWIN, FRANCISCO LACERDA and CHAYADA ATTAPAIBOON. 1996. Facilitation or attenuation in the development of speech mode processing? Tone perception over linguistic contexts. Proceedings of the Sixth Australian International Conference on Speech Science and Technology, ed. by Paul McCormack and Alison Russell, 587-592. Canberra: Australian Speech Science and Technology Association.

BURNHAM, DENIS, KATHRYN KIRKWOOD, SUDAPORN LUKSANEEYANAWIN and SUPATRA PANSOTTEE. 1992. Perception of Central Thai tones and segments by Thai and Australian adults. Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics, 546-560. Bangkok: Chulalongkorn University Press.

CHAO, YUEN REN. 1948. Mandarin Primer. Cambridge, MA: Harvard University Press.

CHEN, YIYA and XU, YI. 1980. Production of weak elements in speech-evidence from F0 patterns of neutral tone in standard Chinese. Phonetica 63 (1): 47-75.

CUTLER, ANNE. 2015. Lexical stress in English pronunciation. The Handbook of English Pronunciation, ed. by Marnie Reed and John M. Levis, 106-124. Oxford: Wiley Blackwell.

ELIASSON, STIG. 1997. Tone in second language acquisition. Language History and Linguistic Modeling: Vol. 2, ed. by Raymond Hickey and Stanislaw Puppel, 1273–1289. Berlin: Mouton de Gruyter.

FISHER, RONALD. A. 1925. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

FORSTER, KENNETH I. and JONATHAN C. FORSTER. 2003. DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments, and Computers 35 (1): 116-124.

FRANCIS, ALEXANDER L., VALTER CIOCCA, LIAN MA and KIMBERLY M. FENN. 2008. Perceptual learning of Cantonese lexical tones by tonal and non-tonal language speakers. Journal of Phonetics 36 (2): 268-294.

FRY, D. B. 1958. Experiments in the perception of stress. Language and Speech 1: 126-52.

GANDOUR, JACKSON 1983. Tone perception in Far Eastern languages. Journal of Phonetics 11: 49-175.

GANDOUR, JACKSON and RICHARD HARSHMAN. 1978. Crosslanguage differences in tone perception: a multidimensional scaling investigation. Language and Speech 21: 1-33.

GUION, SUSAN G. and ERIC PEDERSON. 2007. Investigating the role of attention in phonetic learning. Language Experience in Second Language Speech Learning, ed. by Ocke-Schwen Bohn and Murray J. Munro, 57-77. Amsterdam: John Benjamins.

HALLÉ, PIERRE A., YUEH-CHIN CHANG and CATHERINE T. BEST. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs French listeners. Journal of Phonetics 32: 395-421.

HIRST, DANIEL and ALBERT DI CRISTO. 1998. A survey of intonation systems. Intonation Systems, ed. by Daniel Hirst and Albert Di Cristo, 1-45. Cambridge, UK: Cambridge University Press.

HYMAN, LARRY M. 2009. How (not) to do phonological typology: The case of pitch-accent. Language Sciences 31 (2-3): 213-238.

KIM-RENAUD, YOUNG-KEY. 2009. Korean: An Essential Grammar. New York: Routledge.

KOCHANSKI, GREG, ESTHER GRABE, JOHN COLEMAN and BURTON ROSNER. 2005. Loudness predicts prominence: Fundamental frequency lends little. The Journal of the Acoustical Society of America 118 (2): 1038-1054.

KONDO, MARIKO. 2007. Acoustic realization of lexical accent and its effects on phrase intonation in English speakers' Japanese. Proceedings of the 16th International Congress of Phonetic Sciences 2007 [ICPhS XVI], ed. by Jürgen Trouvain and William J. Barry, 1649-1652. Saarbrücken, Germany: Dudweiler Pirrot.

KUBOZONO, HARUO. 1999. Mora and syllable. The Handbook of Japanese Linguistics, ed. by Natsuko Tsujimura, 31-61. Malden, MA: Blackwell Publishers

LEATHER, JONATHAN. 1983. Speaker normalization in perception of lexical tone. Journal of Phonetics 11: 373-382.

_____ . 1987. F0 pattern inference in the perceptual acquisition of second language tone. Sound Patterns in Second Language Acquisition, ed. by Allan James and Jonathan Leather, 59-81. Dordrecht: Foris Publications.

_____. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second-Language Speech, ed. by Jonathan Leather and Allan James, 72-89. Amsterdam: University of Amsterdam.

LEE, CHAO-YANG. 2007. Does horse activate mother? Processing lexical tone in form priming. Language and Speech 50 (1): 101-123.

_____. 2009. Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study. Journal of the Acoustical Society of America 125 (2): 1125-1137.

LEE, IKSOP and S. ROBERT RAMSEY. 2000. The Korean Language. Albany, NY: State University of New York Press.

LEE, WAI-SUM and ZEE, ERIC. 2008. Prosodic characteristics of the neutral tone in Beijing Mandarin. Journal of Chinese Linguistics 36: 1-29.

LI, CHARLES N. and SANDRA A. THOMPSON. 1977. The acquisition of tone in Mandarin-speaking children, Journal of Child Language 4 (2): 185-199.

MCALLISTER, ROBERT, JAMES E. FLEGE and THORSTEN PISKE. 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, SCOTT. 1996. Tonal distinction errors by beginning Chinese language students: A comparative study of American English and Japanese native speakers. Chinese Pedagogy: An Emerging Field, ed. by Scott McGinnis, 81-91. Columbus, OH: Ohio State University Foreign Language Publications.

MIXDORFF, HANSJÖRG, SUDAPORN LUKSANEEYANAWIN, HIROYA FUJISAKI and PATAVEE CHARNAVIT. 2002. Perception of tone and vowel quality in Thai. Proceedings of the 7th International Conference on Spoken Language Processing, ICSLP2002 – INTERSPEECH 2002, ed. by John H. L. Hansen and Bryan L. Pellom, 753-756. Denver, CO: International Speech and Communication Association.

MORÉN, BRUCE and ELIZABETH C. ZSIGA. 2006. The lexical and post-lexical phonology of Thai tones. Natural Language and Linguistic Theory 24 (1): 113-178.

OTAKE, TAKASHI and ANNE CUTLER. 1999. Perception of suprasegmental structure in a nonnative dialect. Journal of Phonetics 27 (3): 229-253.

RAAIJMAKERS, JEROEN G. W., JOSEPH M. C. SCHRIJNEMAKERS and FRANS GREMMEN. 1999. How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. Journal of Memory and Language 41 (3): 416-426.

SCHAEFER, VANCE. 2015. Influence of the first and second language on the perception of Thai tones. Indiana University doctoral dissertation.

SCHAEFER, VANCE and ISABELLE DARCY. 2014. Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. Laboratory Phonology 5 (4): 489-522.

SO, CONNIE K. 2006. Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. Proceedings of the 11th Australian International Conference on Speech Science and Technology, ed. by Paul Warren and Catherine I. Watson, 438-443. Auckland, New Zealand: University of Auckland.

_____ . 2010. Categorizing Mandarin tones in Japanese pitch-accent categories: The role of phonetic properties. Interspeech 2010 Second Language Studies: Acquisition, Learning, Education and Technology (pp. O1-3). Waseda University, Tokyo, Japan.

SO, CONNIE K. and CATHERINE T. BEST. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. Language and Speech 53: 273-293.

SUGIYAMA, YUKIKO. 2012. The Production and Perception of Japanese Pitch Accent. Cambridge, UK: Cambridge Scholars Publishing.

STAGRAY, JAMES R. and DAVID DOWNS. 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. Journal of Chinese Linguistics 21 (1): 143-163.

TUC, HO-DAC. 2003. Vietnamese-English Bilingualism: Patterns of Code-switching. New York: RoutledgeCurzon.

TUMTAVITIKUL, APILUCK. 1995. Tonal movements in Thai. Proceedings of the 13th International Congress of Phonetic Sciences Volume 1, ed. by Kjell Elenius and Peter Branderud, 118-121. Stockholm: Royal Institute of Technology and Stockholm University.

VAN LANCKER, DIANA and VICTORIA A. FROMKIN. 1973. Hemispheric specialization for pitch and 'tone': Evidence from Thai. Journal of Phonetics 1: 101-109.

WANG, YUE, DAWN M. BEHNE, ALLARD JONGMAN and JOAN A. SERENO. 2004. The role of linguistic experience in the hemispheric processing of lexical tone. Applied Linguistics 25: 449-466.

WANG, YUE, ALLARD JONGMAN and JOAN A. SERENO. 2001. Dichotic perception of Mandarin tones by Chinese and American listeners. Brain and Language 78: 332–348.

_____ . 2006. L2 acquisition and processing of Mandarin Chinese tones. The Handbook of East Asian Psycholinguistics, ed. by Ping Li, Elizabeth Bates, Li Hai Tan and Ovid J. L. Tseng, 250-256. Cambridge, UK: Cambridge University Press.

WANG, YUE, MICHELLE M. SPENCE, ALLARD JONGMAN and JOAN A. SERENO. 1999. Training American listeners to perceive Mandarin tones. Journal of the Acoustical Society of America 106 (6): 3649-3658.

WAYLAND, RATREE P. and SUSAN G. GUION. 2003. Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. Applied Psycholinguistics 24: 113-129.

_____ . 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. Language Learning 54: 681-712.

YIP, MOIRA. 2002. Tone. Cambridge, UK: Cambridge University Press.

ZSIGA, ELIZABETH and RATTIMA NITISAROJ. 2007. Tone features, tone perception, and peak alignment in Thai. Language and Speech 50 (3): 343-383.

Vance Schaefer
Dept. of Modern Languages, Bondurant Hall C-115
University of Mississippi
University, MS 38677
[schaefer@olemiss.edu]