Tone assimilation by Mandarin and Thai listeners with and without L2 experience

Xianghua Wua,⁎, Murray J. Munrob, Yue Wangb

a Department of East Asian Languages and Cultures, University of California, Berkeley 94720, USA
b Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

ARTICLE INFO
Article history:
Received 29 August 2013
Received in revised form 17 June 2014
Accepted 25 June 2014

Keywords:
Tone assimilation
Linguistic experience
Mandarin
Thai
Allotone
Category assimilation task

ABSTRACT
Evidence indicates that perceptual assimilation of segments is tied to L1 and L2 contrasts at a lower phonetic level for listeners without L2 experience, but at both a phonetic and a higher phonological level for those with L2 experience. It is less clear, however, that the same is true for suprasegmental features. In this examination of perceptual assimilation of lexical tones, 40 listeners, whose L1 and L2 are one of the two tone languages, Mandarin and Thai, along with another 40 native listeners of the two tone languages without L2 experience performed a mapping-rating assimilation task in which they first identified which L1 tone sounded most similar to the L2 tone they heard, and then rated the goodness of match on a 5-point scale. The inexperienced listeners assimilated L2 tones to L1 tones with the most similar acoustic properties, i.e., F0 height and contour. The experienced listeners were additionally influenced by phonological tone changes in Mandarin. In particular, falling rising tones were assimilated to the rising tone or low falling tone in Mandarin or Thai and vice versa. These findings are discussed in relation to current conceptions of perceptual assimilation.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Considerable research on the perception of segmental information has examined the factors influencing the assimilation of L2 phonetic and phonological categories to their L1 counterparts (Levy, 2009; Nishi, Strange, Akahane-Yamada, Kubo, & Trent-Brown, 2008; Schmidt, 1996; Strange, 1999; Strange, Bohn, Trent, & Nishi, 2004). In contrast with phonetic assimilation, in which L2 phonetic categories are perceived to be similar to L1 phonetic categories based on acoustic or gestural properties, phonological assimilation occurs because L1 and L2 phonetic categories have the same phonemic status. However, few studies have focused on perceptual assimilation at the suprasegmental level with respect to matters such as tone. Segmental assimilation may apply to tone, as tone functions phonemically in languages such as Mandarin and Thai; thus it is reasonable to expect that tone assimilation is affected by the acoustic properties and phonemic status of L1 and L2 tone categories. The current study explores tone assimilation by native speakers of two tone languages within the framework of recent models of cross-language speech perception.

1.1. Issues related to perceptual assimilation

Three central issues in recent models of cross-language perceptual assimilation are of particular interest in this investigation. One concerns the existence of assimilation processes at both the phonetic and phonological levels, a distinction that reflects a major difference between the Perceptual Assimilation Model (PAM, Best, 1995; Best, McRoberts, & Goodell, 2001; Best, McRoberts, & Sithole, 1988; Best & Tyler, 2007) and the Speech Learning Model (SLM, Flege, 1995, 2007). The PAM proposes that both levels of assimilation may occur, while the SLM characterizes category assimilation strictly at the phonetic level. Among other postulates, the PAM proposes six phonetic and four phonological assimilation patterns based on previous findings from segmental research. Unlike phonetic assimilation, which occurs when one sound is perceived to be the phonetic equivalent of another one, phonological assimilation is stimulated by both phonetic similarity and phonemic correspondence, in particular the latter (Best & Tyler, 2007). With
regard to the assimilation patterns involving categorized or assimilable categories, phonetic or phonological assimilation may be characterized as Two Category (two L2 categories are assimilated to two different L1 categories), Single Category (two L2 categories are assimilated to one L1 category) and Category Goodness (two L2 categories are assimilated to two L1 categories, but one fits better). A second issue of concern here is the effect of language experience on perceptual assimilation. The PAM predicts assimilation exclusively at the phonetic level for listeners without L2 experience but very likely also at the phonological level for those with L2 experience (Best & Tyler, 2007). In contrast, the SLM postulates variable phonetic assimilation for listeners with varying L2 experience. The third issue concerns the extent to which phonetic similarity can predict assimilation at the phonetic and phonological levels. According to the PAM, phonetic similarity between L1 and L2 phones may be predictive of phonetic assimilation but not necessarily phonological assimilation (Best, 1995; Best & Tyler, 2007; Bundgaard-Nielsen, Best, & Tyler, 2011; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Levy, 2009; Levy & Strange, 2008). The SLM, however, posits that despite the importance of acoustic phonetic similarity, assimilation of phonetic categories can occur not only between acoustically similar phones in L1 and L2, but also between dissimilar ones.

1.2. Phonetic assimilation

Phonetic assimilation refers to a process in which listeners attend to the acoustic or gestural similarities between incoming L2 speech sounds and existing L1 phonetic categories (Best et al., 1988; Flege, 1995, 2007). Research has shown that both inexperienced and experienced listeners demonstrate perceptual assimilation at the phonetic level in the case of vowels (e.g., Levy, 2009; Nishi et al., 2008), consonants (e.g., Best et al., 2001), and lexical tones (e.g., Leung, 2008; So, 2012; So & Best, 2010). For example, in the perception of German/French vowels, both inexperienced and experienced American English (AmE) listeners assimilated German/French vowel categories /i, ɛ, a, u/ to the AmE vowels /i, ɛ, a, u/, respectively, based on the closer acoustic space between these vowels in the two languages (Levy, 2009; Strange et al., 2004). At the suprasegmental level, studies have investigated perception of Mandarin tones by Cantonese speakers with Mandarin experience (Leung, 2008) and those without (So, 2012; So & Best, 2010). Results indicate that Mandarin tones were identified as their acoustic equivalents in Cantonese, such as the high level tones and the rising tones, confirming the effects of acoustic similarities on tone perception.

Segmental and suprasegmental studies have also indicated that the assimilated phonetic categories may differ in their degree of phonetic similarity among L1 counterparts. For example, Strange et al. (2004) noted that the German vowel /t/ was assimilated to the American English /t/ by native listeners of American English, despite the fact that the phonetic dissimilarities are greater between these vowels than those between the vowels /l/ /t/ /l/ in L1 and L2. At the suprasegmental level, Cantonese listeners have been found to identify Mandarin high falling tone as Cantonese high falling tone despite the fact that the Mandarin tone ends with a much lower fundamental frequency (F0) than the Cantonese tone, suggesting that phonetic assimilation may be induced by partially common features, in this case a falling contour (Leung, 2008; So & Best, 2010). Tone assimilation research by So (2012) also indicated that Mandarin falling rising tone was assimilated to both Cantonese low falling tone and rising tone because the former represents the initial part of falling rising tone and the latter the final part.

In some cases, it has been suggested that use of phonetic cues may affect assimilation results (Levy, 2009; Strange et al., 2004). A well-known example is Japanese listeners’ confusion between the English consonants /l/ and /l/, which differ primarily in F3, a phonetic feature not attended to by Japanese listeners who tend to use F2 to distinguish the two consonants (e.g., Iverson et al., 2003). As a result, English /l/ is primarily assimilated to Japanese flap /l/ while English /l/ is not (Hattori, 2009). The use of different perceptual cues has also been found to cause different tone assimilation patterns between tone language speakers as well as between tone and non-tone language speakers. In a multidimensional scaling study, Gandour (1983) evaluated the effects of F0 height and contour on the perceptual distance between 19 synthesized pitch contrasts created from a naturally produced tone imposed on the monosyllable [wa]. Language-specific discrimination was detected, as the listeners attached greater importance to either F0 height or contour. Native Cantonese, Mandarin and Thai listeners with tonal L1s were more sensitive to F0 contours than non-tonal English listeners, and Thai listeners were more sensitive to F0 contours than were Mandarin listeners. Furthermore, the English listeners appeared to attach more importance to the pitch at the end of a tone; thus a high level tone and a rising tone (ending with a high pitch) were perceived to be similar. Gandour’s (1983) findings indicating the selection of preferred perceptual cues by tone and non-tone language speakers have been supported by more recent studies (e.g., Francis, Ciocca, Ma, & Fenn, 2008; Guion & Pederson, 2007; Huang & Johnson, 2010).

1.3. Phonological assimilation

Phonological assimilation refers to a higher-level perceptual process in which listeners assimilate an L2 phonological category to an L1 category with similar lexical functions. In previous work, phonological assimilation has been exemplified in three main ways. One of these is mentioned by Best and Tyler (2007), who note that English learners of French assimilate French //l/, a voiced uvular fricative, to the English liquid /l/ in word recognition, even though the two sounds bear little phonetic similarity. At the lexical level, the French and English consonants appear to be interchangeable from the standpoint of English speakers.

In a second illustration of phonological assimilation, experienced listeners have been observed to relate “unassimilable” L2 phonological categories to L1 phonological categories. In Bundgaard-Nielsen et al. (2011), Japanese learners of Australian English (AE) who had more exposure to AE showed more consistent mappings between Japanese and AE vowels in terms of the types and numbers of AE categories assimilated to Japanese categories than those with less exposure to AE. For instance, the more
experienced Japanese listeners appeared to associate such vowels as /æ/, /æ/ and /æ/ more consistently with Japanese vowel categories while these vowels were less frequently identified as Japanese vowel categories by the less experienced Japanese listeners.

A third phenomenon, which is the most relevant to this study, is that, for L1 listeners, poor discrimination tends to occur between allophonic variants which share the same phonological functions (Boomershine et al., 2008; Huang, 2001). At the segmental level, for example, Boomershine et al. (2008) found that allophonic contrasts (e.g., /d/ and /r/) were less distinguishable than phonemic contrasts (e.g., /d/ and /l/) for native English speakers; and that response times were longer in the discrimination of the allophonic contrasts than the phonemic ones. At the suprasegmental level, Huang (2001) noted a significantly smaller perceptual distance between two Mandarin tones, rising tone and the falling rising tone, than between other tone pairings. In her study, Mandarin and English listeners heard pairwise Mandarin tone contrasts in a discrimination task. Response times were significantly longer when the Mandarin listeners discriminated between rising tone and falling rising tone than other tone pairings. The higher degree of misperception by the Mandarin listeners was attributed to the similar phonological use of Mandarin rising tone and falling rising tone under some conditions, in which falling rising tone is realized as rising tone when it is followed by another falling rising tone at the word level (Duanmu, 2007). In other words, Mandarin rising tone is used as an allophonic variant of falling rising tone. The allophonic tone variant is also known as allotone (Gandour, 1981; So & Best, 2010). Similar to segmental allophones, allotones are related to each other due to the same phonemic status.

Not only has perceptual confusion between allotones been observed for native speakers, but a number of studies have also indicated phonemic-level tone assimilation by Cantonese listeners due to phonological correspondence between the Mandarin high level and high falling tones (Hao, 2012; Leung, 2008; So & Best, 2010). In Cantonese, these two tones appear as allophonic variants without phonemic contrast (Hashimoto, 1972; Yu, 2007). Additionally, Tse (1978) points out that Cantonese high falling and high level tones are unconditioned free variants. Cantonese listeners with L2 Mandarin experience either assimilate the two Mandarin tones to Cantonese high level tone (Leung, 2008) or show high identification confusion between the two tones (Hao, 2012). On the other hand, some research also suggests that phonemic-level tone assimilation can be found for listeners without L2 experience. In one recent study, So (2012) asked Cantonese listeners without any previous Mandarin experience to match each of the four Mandarin tone categories to the six Cantonese categories. The results indicated phonetic assimilation for Mandarin high level, rising, and falling rising tones as mentioned above. Mandarin falling tone, which corresponds to the allotone of Cantonese high level tone, however, was assimilated to Cantonese high level tone. In another study, So and Best (2010) also found that Mandarin high falling tone was identified with relatively higher rates as Mandarin high level tone by Cantonese listeners without Mandarin experience. These findings point to phonological tone assimilation for both inexperienced and experienced listeners, suggesting that phonological assimilation may be predicted by both lower-level acoustic similarities and higher-level phonemic status.

In addition, Zee (1999) suggests that Cantonese listeners’ high sensitivity to phonological tone changes may be accounted for by their rich L1 tone system (6 Cantonese tones vs. 4 Mandarin tones). In fact, in a recent study by Zheng, Minett, Peng, and Wang (2010), Cantonese listeners exhibited greater sensitivity to the acoustic distinctions and engaged in phonological processing more strongly than their Mandarin counterparts, especially when processing speech as opposed to complex (sawtooth) tones. It was suggested that a “denser” L1 tone system in Cantonese may enhance L2 tone perception at both the phonetic and phonological levels. In a similar vein, to assess the contributions of phonetic and phonological properties to cross-language speech perception, Bohn and Best (2012) investigated the perception of English approximant contrasts by native speakers of Danish and German. Their findings revealed that inexperienced Danish listeners were better at perceiving some English approximant contrasts than native listeners probably because Danish is richer in approximants compared to English. They concluded that the broader phonological subclasses might lead to superior performance in non-native speech perception.

1.4. Summary

In summary, previous studies have revealed a process of phonetic assimilation based on acoustic characteristics and a process of phonological assimilation based on phonemic functions. At the lower acoustic level, phonetic assimilation can be predicted by physical similarities between L1 and L2 phones, although the degree of similarity necessary for such assimilation can vary. Moreover, phonetic assimilation may be determined by the use of phonetic featural cues (e.g., F0 height vs. F0 contour in tone assimilation). At the higher functional level, phonemic status, rather than phonetic similarity, leads to the association of an L2 phonological category with its counterpart in L1. Research has also indicated that language experience is related to assimilation processes at the phonetic and phonological levels. However, it is not clear how patterns of assimilation vary as a function of language experience. For example, while the PAM proposed phonological assimilation patterns only for the experienced L2 listeners (Best & Tyler, 2007), recent findings (e.g., So, 2012; So & Best, 2010) suggest that inexperienced listeners are also sensitive to the phonological correspondence between L1 and L2 phonological tone categories.

2. The current study

Although previous studies of tone perception have probed the assimilation of tone patterns according to their acoustic phonetic properties and phonemic status using discrimination (e.g., Huang, 2001) or identification (e.g., Hao, 2012; So & Best, 2010) tasks, few have employed an assimilation research paradigm to determine the extent to which an L2 tone category is selected as the
phonetic or phonological equivalent of an L1 tone category by listeners from different L1 tone language backgrounds but with L2 learning experience of the other language (So, 2012). To address this issue, the current study adopted the category assimilation method (e.g., Levy, 2009) to explore the assimilation patterns between Thai and Mandarin tone categories by inexperienced and experienced listeners. The rationale for selecting these two tone systems lay in the fact that the Thai tone system is similar to the Mandarin tone system in terms of the size of tone inventory and in the nature of acoustic correlates (Chao, 1948; Ladefoged, 2001). There are five tone categories in Thai (mid falling, low falling, high falling, high rising and falling rising) and four in Mandarin (high level, rising, falling rising and falling), both having level and contour pitch patterns. Moreover, Mandarin falling rising tone has two conditional allophonic variants: rising tone and low falling tone. As discussed earlier, falling rising tone is realized as rising tone under the condition where it precedes another falling rising tone at the word level. In most other cases, falling rising tone is realized as low falling tone when it is followed by the other three tones, i.e., high level, rising and high falling tone in a word (Shih, 1997). Thus low falling tone also functions as an allotone of falling rising tone in Mandarin (cf., Hallé, Chang, & Best, 2004).

Furthermore, as discussed previously, research findings have not been consistent with respect to phonetic and phonological assimilation patterns as a function of L2 experience. Thus the current study investigates tone assimilation at the phonetic and phonological levels by native speakers of Mandarin and Thai who differed not only in L1, but also in L2 tone experience. Compared to the inexperienced listeners who had no Mandarin experience, the experienced listeners had 0.5–2 years of exposure to Mandarin (for the experienced Thai listeners) or Thai (for the experienced Mandarin listeners).

The experiments employ a cross-language perceptual assimilation task, also known as a category assimilation task, which has been used widely to probe segmental assimilation (Levy, 2009; Nishi et al., 2008; Schmidt, 1996; Strange, 2007; Strange et al., 2004) as well as suprasegmental assimilation (So, 2012; So & Best, 2011). The task consists of a mapping portion and a rating portion, which are performed sequentially for each token. In the mapping portion, listeners choose a native tone category which is perceived most similar to the incoming non-native tone; then, in the rating portion, they rate the similarity between the native and non-native tones for goodness-of-fit on a Likert scale. As outlined below, two research questions are addressed.

2.1. Does language experience affect phonetic and phonological tone assimilation?

It is expected that both L1 and L2 experience will have effects on tone assimilation. According to previous findings (e.g., Gandour, 1983; Zheng et al., 2010), listeners from a relatively larger L1 tone inventory appear to be more sensitive to the acoustic distinctions between tone categories. Therefore, compared to their Mandarin counterparts, Thai listeners should be more sensitive to the acoustic similarities between Thai and Mandarin tone categories, in particular similarities in F0 contour (Gandour, 1983). In terms of the effects of L2 experience, inexperienced and experienced listeners are expected to assimilate L2 tones to L1 tone categories which are acoustically similar in F0 height and contour. In addition, due to their exposure to the L2 tone system, experienced listeners are also expected to discover the phonemic equivalence of some L2 tone categories in spite of the nature of the acoustic correspondence. Given the findings from So and Best (2010) and So (2012) showing the inexperienced listeners could detect the phonological correspondence between L1 and L2 tone categories, it is hypothesized that inexperienced Mandarin listeners will also associate L2 rising tone and low falling tone to L1 falling rising tone because of their familiarity with the allophonic tone variants in L1 (cf., Cantonese listeners).

2.2. What factors are predictive of tone assimilation at the phonetic and phonological levels?

This question concerns the factors tied to tone assimilation at the two different levels. Given the similarities between the Mandarin and Thai tones in terms of F0 height and contour, it seems reasonable to expect that speakers of either language will perceive the tones of the other language as "categorizable" or "assimilable" in terms of the Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007). In other words, both Mandarin and Thai listeners are expected to assimilate L2 tones to L1 tone categories according to the similarities in F0 height and contour (Gandour, 1983). Furthermore, on the basis of previous findings (e.g., Huang, 2001; Lueng, 2008; So & Best, 2010), tone assimilation can be expected to occur at the phonological level. In particular, phonological tone assimilation is likely to arise from Mandarin allophonic tone changes such that a falling rising tone will be assimilated to its allotonic variants, a rising tone or a low falling tone, and vice versa. Thus these three tones are very likely to be perceived as phonologically similar tone categories, despite the fact that they differ in F0 height and contour. For the rest of the tones, the acoustic similarity between L1 and L2 tones is expected to determine the assimilation patterns.

3. Acoustic similarity between Mandarin and Thai tones

3.1. Comparison of Mandarin and Thai tones

Table 1 presents the four Mandarin and five Thai tones with their pitch changes on Chao's (1948) 5-point scale, as well as the conventional descriptions of the pitch changes in terms of F0 height and contour with 1 indicating the lowest F0 level and 5 the highest (Ladefoged, 2001). In this table, low falling rising (214) and high falling (51) tones exist in both languages, while the Mandarin high rising tone (35) is very similar to the Thai high rising tone (45). Additionally, the Thai low falling tone (21) is realized as a partial component of the falling rising tone. In contrast, the Mandarin high level tone (55) and the Thai mid falling tone (32) appear to lack
acoustic equivalents in the other language. Nevertheless, the high F0 of the Mandarin level tone may be associated with the Thai high rising tone (45) and the falling contour of the Thai mid falling tone (32) may be related to the Mandarin falling tone (51).

In order to examine the assimilation patterns of Thai and Mandarin tones, it is crucial to investigate the actual acoustic properties of the two tonal systems. Conventional descriptions indicating categorical distinctions of the tones may not reflect the acoustic properties of tone categories naturally produced by different speakers in different circumstances (Gandour & Potisuk, 1991). Acoustically, a tone category can be produced with varying pitch trajectories in citation form. For instance, Abramson (1978) and Gandour (1978) described Thai mid tone as a mid (level) tone while Ladefoged (2001) referenced it as a mid falling tone. In Mandarin, rising tone and falling rising tone can be produced from highly distinguishable to very confusable depending on the realization of their acoustical parameters (Shen & Lin, 1991). Thus, in the current research, prior to the tone assimilation study, F0 correlates of Thai and Mandarin tones were measured acoustically. Following previous studies (e.g., Bent, 2005; Ciocca, Whitehill, & Ng, 2002; Wang, Jongman, & Sereno, 2003), the current research used the primary acoustic correlates of lexical tone, i.e., F0 height and contour, to compare and categorize Mandarin and Thai tones in that F0 height and contour have been verified as the most important perceptual cues contributing solely or jointly to tone identification (Thai, Abramson, 1997; Mandarin, Guion & Pederson, 2007; Lin, 1988). Other acoustic features of tone, such as tonal duration or voice quality may affect tone perception to some extent, but are not considered in the current study due to their relatively small effect on tone perception (e.g., Taiwanese, Lin & Repp, 1989).

3.2. Method

3.2.1. Talkers and speech material

A gender-balanced set of recordings of Thai and Mandarin tones was collected from two speakers each of Mandarin and of Thai for acoustic analysis. The Mandarin speakers were born and raised in Beijing and were in their early 20s at the time of recording. The male speaker had 1.5 years of residence in North America and the female speaker, 2.5 years. The Thai speakers, who had a mean age of 22 years, were born and raised in Bangkok. The male speaker had resided in North America for 4 years and the female speaker 2 years. The four speakers all reported normal speaking and hearing ability. All were undergraduate students of Simon Fraser University in Canada, and none had experience with other tone languages.

The stimulus set consisted of 12 syllables, each produced with the four Mandarin and five Thai tones, as well as hums which contain only tonal (but no segmental) information (Lee, Vakoch, & Wurm, 1996). As indicated in Table 2, 10 of the targets were real words in Mandarin or Thai varying in syllabic structure [CV(V), CVC, CCV(V)], and two were nonsense syllables in both languages. The composition of this stimulus set was designed to ensure the diversity of tokens in tone productions, covering a wide range of segmental, syllabic, and lexical semantic contexts in which tone may occur. Therefore, data from all the word and non-word conditions, including the hummed tones, were pooled in the current analyses.

3.2.2. Procedure

3.2.2.1. Recording. The stimuli were recorded in a sound-treated room in the Language and Brain Lab at Simon Fraser University, using a Presonus Digital Audio 24 B27/96K firewire recording interface and a Shure KSM 109 microphone, with a sampling rate
of 44.1 kHz and 16-bit resolution. The Mandarin words were read in the carrier sentence “我读一个______” (I read a _____), with the target word appearing in the final underscored position (Duanmu, 2007). The Thai words were produced within the frame “รู้การตั้ง______” by the male speaker and in the frame “รู้การตั้ง______” by the female speaker. Both sentences mean “I want to read_____,” but with gender-appropriate personal pronouns (ฉัน for male and ฉัน for female). Target items were randomly presented three times each to be read from a computer screen. The Mandarin words appeared in pinyin and Chinese characters, while the Thai words were shown in Thai orthography. In order to produce hummed tones, Mandarin tone marks superimposed on /m/ and Thai tone marks superimposed on ุ (/m/ in Thai orthography) were presented to the Mandarin and Thai speakers, respectively. Before the formal recording, the speakers practiced the full set of items several times to ensure fluent and natural productions.

All signal manipulations were performed using Praat (Version 5.1.05, Boersma & Weenink, 2009). The target words and the hums were first excised from the digital recordings and saved as individual sound files. Items were then screened by native raters, two speakers each of Mandarin and of Thai, neither of whom had participated in this study as a speaker. They listened through AKG headphones at a comfortable level and evaluated the goodness of the tone productions by first identifying the tones presented through E-prime software 1.0 (Psychology Software Tools, Inc., Sharpsburg, PA) and then selecting the best tone from the three repetitions. The mean correct identification rate was 100% across the two raters for each language, showing high accuracy of tone productions. The best-rated exemplar of each target tone superimposed on each syllable and hum was selected for the acoustic analysis.

3.3. Results

The results of the acoustic analyses are displayed in Fig. 1a and b, which show tone trajectories in Mandarin and Thai respectively and in Table 3, which provides the mean values of the acoustic parameters described above for all nine tones. The small number of speakers for each tone system precluded the use of statistical analyses for evaluating the difference in each parameter between Mandarin and Thai tones. As a result, these tones were categorized into four types on the basis of the mean value of the parameters presented three times each to be read from a computer screen. The Mandarin words appeared in pinyin and Chinese characters, while the Thai words were shown in Thai orthography. In order to produce hummed tones, Mandarin tone marks superimposed on /m/ and Thai tone marks superimposed on ุ (/m/ in Thai orthography) were presented to the Mandarin and Thai speakers, respectively. Before the formal recording, the speakers practiced the full set of items several times to ensure fluent and natural productions.

All signal manipulations were performed using Praat (Version 5.1.05, Boersma & Weenink, 2009). The target words and the hums were first excised from the digital recordings and saved as individual sound files. Items were then screened by native raters, two speakers each of Mandarin and of Thai, neither of whom had participated in this study as a speaker. They listened through AKG headphones at a comfortable level and evaluated the goodness of the tone productions by first identifying the tones presented through E-prime software 1.0 (Psychology Software Tools, Inc., Sharpsburg, PA) and then selecting the best tone from the three repetitions. The mean correct identification rate was 100% across the two raters for each language, showing high accuracy of tone productions. The best-rated exemplar of each target tone superimposed on each syllable and hum was selected for the acoustic analysis.

3.2.2.2. Pitch analysis. Each tone production was evaluated via a set of acoustic parameters to establish F0 height and contour. These comprised F0 at five successive temporal points (0%, 25%, 50%, 75% and 100% of a tone), maximum F0, minimum F0, average F0, pitch range of falling and rising contours, and timing of the turning point for falling rising tones. Timing of the turning point refers to the duration from onset of the tone to the point of change in F0 direction (Moore & Jongman, 1997; Shen & Lin, 1991), a parameter which has been suggested to cue Mandarin rising tone and falling rising tone (Bent, 2005; Khow & Ciocca, 2007; Wang et al., 2003). The timing of each turning point was converted to a % value, by computing the ratio of the duration between the starting point and turning point to that of the whole tone contour. Falling and rising ranges were calculated to capture the pitch change from onset to offset for the falling and rising tones, and from the onset to the turning point as well as from the turning point to the offset for the dipping tones.

As discussed earlier, the most important acoustic correlates of lexical tones are F0 height and contour (Abramson, 1975; Gandour, 1983); thus acoustical analyses in the current study focused on the F0 variations of Mandarin and Thai tones. To facilitate comparisons of the acoustic parameters among the target tones across speakers and syllable structures, the F0 contours were first equalized for duration using Praat (see also Wang et al., 2003). The duration normalization would consequently not allow the examination of durational effects on either acoustic analysis of tone categories or listeners’ perception of these tones. However, previous research has indicated that the influence of duration on tone perception is secondary compared to F0 height and contour (e.g., Lin, 1988; Lin & Repp, 1989). Controlling for duration thus allows association of observed perceptual patterns with the F0 dimension, the primary cue to tone perception. In order to preserve all F0 information from the signal, the longest F0 contour of each tone category was first identified (Wang et al., 2003). All other tokens from the same tone category were then lengthened to this value using the enhanced pitch-synchronous overlap-and-add (PSOLA) technique (Boersma & Weenink, 2009), which manipulates the duration of a sound signal without altering pitch.

Pitch tracks for the equalized stimuli were then extracted automatically using the autocorrelation method, with ranges set to 70–400 Hz for the female speakers and 50–300 Hz for the male speakers (Boersma & Weenink, 2009). F0 measurements (in Hz) were made at 10 ms intervals to generate, for example, 51 F0 values for a 500-ms tone. Each F0 value was then converted from Hz to a logarithm-based T value using the following formula, to normalize for pitch range differences among speakers (Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Peabody & Seneff, 2009; Rose, 1987; Wang et al., 2003):

\[
T = \frac{\lg X - \lg L}{\lg H - \lg L} \times 5
\]

where \(X\) is the pitch (in Hz) at any given point, \(L\) is the lowest pitch and \(H\) the highest pitch produced by a given talker. \(T\) values ranged from 0 to 5, corresponding to Chao’s (1948) pitch scale for Mandarin tones. According to this scale, 0 refers to the lowest pitch (when \(\lg X = \lg L\)) and 5 the highest pitch (when \(\lg X = \lg H\)) in the production of a given talker.

**3.3. Results**

The results of the acoustic analyses are displayed in Fig. 1a and b, which show tone trajectories in Mandarin and Thai respectively and in Table 3, which provides the mean values of the acoustic parameters described above for all nine tones. The small number of speakers for each tone system precluded the use of statistical analyses for evaluating the difference in each parameter between Mandarin and Thai tones. As a result, these tones were categorized into four types on the basis of the mean value of the parameters mentioned above. Since the categorization was based on acoustic analysis for Mandarin and Thai tones collected for the current study, it is not entirely consistent with the categories in Table 1. The four tone types were (1) level tones – Mandarin high level (M-HL) and Thai mid level (T-ML); (2) falling tones – Mandarin high falling (M-HF), and Thai high falling (T-HF) and low falling (T-LF); (3) rising tones – both Mandarin (M-R) and Thai (T-R); and (4) falling rising tones – both Mandarin (M-FR) and Thai (T-FR). The majority of the tones displayed F0 heights and contours similar to those indicated in Table 1, except for the Thai mid tone. In Table 1, Thai mid tone is described as a mid falling tone (32) with a falling range similar to the Thai low falling tone (21), while in the current study, the Thai
mid level tone carried a much flatter falling range (0.3) compared to Thai low falling tone (2.6). Therefore, the former was grouped with level tones and the latter with falling tones. Nevertheless, the Thai mid tone having the features of a level tone is consistent with Abramson (1978) and Gandour (1978).

The tone trajectories show that M-HL and T-ML have a relatively level pattern with negligible pitch variability, i.e., little difference between maximum $F_0$ and minimum $F_0$ (M-HL: 0.2 and T-ML: 0.4). Nevertheless, M-HL and T-ML differ in $F_0$ values. The five designated points, the maximum $F_0$, the minimum $F_0$, and the average $F_0$ are all relatively close to $T$ values of 4 for M-HL and 3 for T-ML. For this reason, M-HL can be considered a high level tone, and T-ML a mid level tone.

In contrast, M-HF, T-LF and T-HF can all be categorized as falling tones because of a clear pitch drop between the onset and offset for each tone (M-HF: 2.7; T-HF: 2.2; T-LF: 2.6). Moreover, as the starting $F_0$ values of M-HF and T-HF are higher than those of T-LF (M-HF: 4.2; T-HF: 3.6; T-LF: 3.2), T-LF can be regarded as a low falling tone.

M-FR and T-FR carry a complex falling rising (dipping) contour. The falling and rising ranges are distinctive, with relatively high $T$ values for both tones (M-FR: falling, 2; rising, 2.4, and T-FR: falling, 2; rising, 2.8). M-FR and T-FR are also very close in terms of the timing of the turning point, despite T-FR showing a higher average $F_0$.

The rising tones M-R and T-R both start with a short and slightly falling slope before moving upwards. Compared to the falling rising tones, the amount of falling is negligible. Moreover, because the rising ranges are greater than the falling ranges (M-R: falling 0.3; rising: 1 and T-R: falling 0.2; rising, 0.6), these two tones are best classified as rising tones.

---

Table 3
Mean acoustic measurements for Mandarin and Thai tones. $F_0$ is shown in $T$ values.

<table>
<thead>
<tr>
<th>Tone type</th>
<th>Tone</th>
<th>Pitch at</th>
<th>Max. $F_0$</th>
<th>Min. $F_0$</th>
<th>Avg. $F_0$</th>
<th>Turning point</th>
<th>Falling range</th>
<th>Rising range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0%</td>
<td>25%</td>
<td>50%</td>
<td>75%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>M-HL</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.8</td>
<td>3.9</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>T-ML</td>
<td>3.4</td>
<td>3.2</td>
<td>3.1</td>
<td>3.1</td>
<td>3.1</td>
<td>3.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Falling</td>
<td>M-HF</td>
<td>4.2</td>
<td>4.1</td>
<td>3.6</td>
<td>2.6</td>
<td>1.9</td>
<td>4.3</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>T-HF</td>
<td>3.6</td>
<td>3.7</td>
<td>3.6</td>
<td>2.3</td>
<td>1.6</td>
<td>3.8</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>T-LF</td>
<td>3.2</td>
<td>2.6</td>
<td>1.7</td>
<td>1.7</td>
<td>1.2</td>
<td>3.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Falling Rising</td>
<td>M-FR</td>
<td>2.6</td>
<td>1.3</td>
<td>0.7</td>
<td>2.3</td>
<td>2.6</td>
<td>3.1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>T-FR</td>
<td>2.8</td>
<td>1.6</td>
<td>1.2</td>
<td>2.9</td>
<td>3.7</td>
<td>3.8</td>
<td>1</td>
</tr>
<tr>
<td>Rising</td>
<td>M-R</td>
<td>3</td>
<td>2.8</td>
<td>3</td>
<td>3.3</td>
<td>3.5</td>
<td>3.7</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>T-R</td>
<td>3.4</td>
<td>3.2</td>
<td>3.3</td>
<td>3.5</td>
<td>3.7</td>
<td>3.8</td>
<td>3</td>
</tr>
</tbody>
</table>

Fig. 1. Tone trajectories (expressed in T units) produced by two Mandarin (the top panel a) and two Thai native speakers (the bottom panel b).
In summary, the results of the acoustic analysis illustrate acoustic correspondences between the Mandarin and Thai tone inventories. The results are consistent with the numerical descriptions by Ladefoged (2001), except for T-ML which he describes as a mid falling tone. These results will be used to interpret the assimilation data described below.

4. Assimilation study

4.1. Method

4.1.1. Listeners

The listeners were 72 native speakers of either Mandarin \((N=36)\) or Thai \((N=36)\) grouped according to L2 tone language experience (Table 4). The inexperienced Mandarin (IM) listeners had no experience with other tone languages. All were born and raised in the Beijing area and were undergraduate students at Beijing universities at the time of testing. The experienced Mandarin (EM) group had been Thai majors for two years at the Guangxi University for Nationalities in China and for three months at the University of the Thai Chamber of Commerce in Thailand. The EM listeners started learning Thai during their first year in the Guangxi university. The 20 inexperienced Thai (IT) listeners had no experience with other tone languages. All were undergraduate students at the University of the Thai Chamber of Commerce in Bangkok. The 16 experienced Thai (ET) listeners had 0.5–2 years of Mandarin experience as Chinese as a second language students at Beijing College of Chinese Language and Culture. The ET listeners had no exposure to Mandarin before they started the beginner’s courses in Chinese at the Beijing college.

All listeners reported normal hearing and speech ability, and all were right-handed and had very limited or no musical training. The experienced listeners reported more daily use of L2 than L1 during the stay in the target language speaking countries. On average, they studied the target language for 5–6 hours per week, with the target language being the teaching language.

4.1.2. Stimuli

The stimuli consisted of the four syllables listed in Table 2, including two Mandarin-only syllables (/tuol/, /fej/) and two Thai-only syllables (/kha(a)/, /sjau/) as well as a hummed syllable, each with the four Mandarin and the five Thai tones. These tokens had been recorded and modified for the acoustic analysis described above; thus the tonal duration was normalized as described earlier. This stimulus set with tones appearing in diverse contexts was selected to minimize bias in tone perception by the non-native listeners due to segmental, syllabic or lexical semantic influence from their L1s.

4.1.3. Procedure

Listening sessions were conducted in quiet rooms in Bangkok for the IT and EM listeners and in Beijing for the IM and ET listeners. Stimuli were presented using Paradigm (2009, Beta Version) software via a laptop and AKG headsets. The perceptual assimilation task entailed both categorical response (i.e., mapping) and scalar rating (Levy, 2009; Nishi et al., 2008; Strange, 2007; Strange et al., 2004). Instructions in the listeners’ native languages were displayed on the computer screen. After hearing each non-native stimulus, the listeners chose the native tone category most similar to the presented tone by pressing a labeled key from one of 4 choices for the Mandarin listeners and one of 5 for the Thai listeners. Immediately afterward, they rated the degree of similarity between the native and non-native tones on a 5-point Likert scale, with 1 representing “least similar” and 5 “most similar.” During a familiarization session, 10 non-native words were presented for mapping and rating. In the subsequent experimental session, 40 Mandarin words (5 syllables × 4 tones × 2 speakers) and 50 Thai words (5 syllables × 5 tones × 2 speakers) were presented twice in random order to the Thai and Mandarin listeners respectively. Both the mapping and rating tasks were self-paced, but the listeners were instructed to respond as quickly as possible. Each session lasted about 20 minutes, including a self-paced break between the blocks.

4.1.4. Data analysis

Data collected from the tone assimilation tasks were analyzed to obtain two measures: fit index and mapping-based degree of response diversity.

Fit index is a metric which combines response rates and goodness ratings. In previous studies, fit index has been used to determine the modal response category (or the closest category), which is the one with the maximum index value (Guion et al., 2000; Hallé, Best, & Levitt, 1999; Iverson & Evans, 2007; Lengeris, 2009). Therefore, a larger fit index indicates a shorter perceptual
distance between L1 and L2 categories. Fit index is calculated by multiplying the mean percentage of responses and mean goodness rating in order to express the assimilation fit of an L2 category to L1 categories.

The degree of response diversity was calculated to assess the consistency in responses for each L2 tone category being mapped to every L1 tone category. Frequencies of responses in each L1 tone category were tallied, and the degree of response diversity \((K')\) was calculated using the following formula (Koopman, Personal communication; Simpson, 1949):

\[
K' = \frac{1}{\sum_{i=1}^{R} P_i^2}
\]

where \(R\) is the total number of L1 tone categories and \(P_i\) is the percentage of responses in which an L2 category \((i)\) was assimilated to a particular L1 tone category. Minimum diversity \((K' = 1)\) indicates that an L2 tone category was consistently mapped to a single L1 tone category. At maximum diversity, \(K'\) is equal to the number of tone categories – 5 for the Thai listeners and 4 for the Mandarin listeners – and indicates that an L2 stimulus was equally mapped to all L1 tone categories. Also, according to Krumhansl (1978), if one object within a stimulus set is similar to many objects, it should be less similar to any particular object. Therefore, if \(K'\) is large, an L2 tone category is less similar to the modal response category; if \(K'\) is close to its minimum, an L2 category is very similar to the modal response category.

4.2. Results

4.2.1. Assimilation of Mandarin tones to Thai tones

4.2.1.1. Fit index. Mean percentage of response and mean goodness rating were first calculated for the assimilation of each input Mandarin tone to the five Thai tone categories by the inexperienced and experienced Thai listeners (see Table A1 in Appendix A for a list of mean response rates and goodness ratings for the Thai listeners). Fit index was then computed by multiplying the two measures for each mapping. The fit index values for the assimilation of Thai tones to Mandarin tone categories are listed in Table 5.

With respect to the fit index, both IT and ET listeners mapped M-HL mainly onto T-ML (IT: 2.36; ET: 4.49), and M-HF mainly onto T-HF (IT: 2.18; ET: 2.89). In contrast, a between-group difference was observed for M-R and M-FR in terms of assimilation fit. While the IT group mapped M-R mainly to T-R (2.21) and M-FR to T-FR (2.02), the ET listeners mapped M-R primarily to T-FR (2.03) and M-FR to T-LF (1.59). To assess group differences between the maximum fit indexes for the consistent mappings, one-factor ANOVAs using group as the independent variable further indicated a higher assimilation fit for the ET listeners than the IT listeners when mapping M-HL to T-ML \(F(1, 34) = 37.75, p < .0001\), and M-HF to T-HF \(F(1, 34) = 8.83, p = .005\).

The assimilation results indicated perceptual differences between the inexperienced and experienced groups. Unlike the inexperienced listeners, who tended to map L2 tones to L1 tone categories with the same \(F_0\) height and contour, the experienced group evidently did not base their mappings solely on the acoustic characteristics of the stimuli.

4.2.1.2. Degree of diversity. As noted above, the Mandarin tones varied in the consistency with which they were mapped to Thai tone categories. To quantify this variability, the degree of mapping diversity \((K')\) was computed for each Mandarin tone and is presented in Fig. 2 for the IT and ET listeners.

Table 5
Assimilation fit of Mandarin tones to Thai tone categories for the inexperienced Thai (IT) and experienced Thai (ET) listeners. Bold indicates fit index values for the modal responses.

<table>
<thead>
<tr>
<th>Perceived as</th>
<th>Presented tones</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M-HL</td>
<td>M-R</td>
<td>M-FR</td>
<td>M-HF</td>
</tr>
<tr>
<td>IT</td>
<td>2.36</td>
<td>0.04</td>
<td>0.30</td>
<td>0.04</td>
</tr>
<tr>
<td>ET</td>
<td>4.04</td>
<td>0.07</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>T-ML</td>
<td>0.08</td>
<td>0.08</td>
<td>0.91</td>
<td>0.62</td>
</tr>
<tr>
<td>T-LF</td>
<td>0.09</td>
<td>0.23</td>
<td>0.25</td>
<td>0.56</td>
</tr>
<tr>
<td>T-HF</td>
<td>0.89</td>
<td>0.24</td>
<td>0.56</td>
<td>2.18</td>
</tr>
<tr>
<td>T-R</td>
<td>0.03</td>
<td>0.69</td>
<td>2.02</td>
<td>0.02</td>
</tr>
<tr>
<td>T-FR</td>
<td>0.02</td>
<td>0.97</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Mapping diversity \((K')\) for the four Mandarin tones as perceived by the IT and ET listeners in terms of Thai categories.
A two-factor mixed-design ANOVA was then computed using $K'$ as the dependent variable, experience group (IT and ET) as a between-groups factor, and input tone (M-HL, M-R, M-FR and M-HF) as a repeated measure. Main effects of group, $F(1, 34)=7.2, p<.011$ and tone, $F(3, 102)=11.6, p<.0001$, as well as the interaction of group and tone, $F(3, 102)=8.2, p<.0001$, were significant. To evaluate the interaction effect, follow-up one-factor ANOVAs were performed using tone as the independent variable for each group. A significant effect of tone was detected for the ET group, $F(3, 45)=25.1, p<.0001$, but not for the IT group $F(3, 57)=1.1, p=.35$. Bonferroni-adjusted post hoc analyses for the ET group indicated significant differences in $K'$ between M-HL and M-R ($p<.0001$), M-HL and M-FR ($p<.0001$), M-HL and M-HF ($p<.0001$), as well as between M-FR and M-HF ($p=.06$), showing a lower diversity of mapping for M-HL compared to the other three tones and for M-HF compared to M-FR. Additional one-factor ANOVAs were performed to compare the two groups on M-HL, M-R, M-FR and M-HF. Significant between-group differences were observed for M-HL, $F(1, 35)=27.9, p<.0001$ and M-HF, $F(1, 35)=8.5, p=.006$, though not for M-R, $F(1, 35)=.27, p=.61$ or M-FR, $F(1, 35)=.03, p=.87$, indicating lower diversity of mapping of M-HL and M-HF onto Thai tone categories by the ET listeners. Therefore, as was the case for modal responses, $K'$ revealed differences in mapping consistency between the inexperienced and experienced groups in terms of the mapping of L2 level and high falling tones onto their L1 counterparts. However, the ET listeners’ perception of Mandarin and Thai level tones was more similar than the IT listeners’, as the latter group associated the Mandarin level tone with a larger range of Thai tone categories (ET: 1.1 vs. IT: 2.2).

### 4.2.2. Assimilation of Thai tones to Mandarin tones

#### 4.2.2.1. Fit index

Fit index derived from percentages of mapping response and goodness ratings was analyzed to pinpoint any group effects on mapping patterns and to determine the closest L1 tone categories for the input L2 tone. Table A2 in Appendix A lists Fit index values for all mapping patterns of the five Thai tones assimilated to the Mandarin tone categories.

As shown in Table 6, both the IM and EM listeners assimilated T-ML to M-HL (IM: 3.96, EM: 4.06), and T-HF to M-HF (IM: 3.7, EM: 3.98) with high frequency. In addition, the IM and EM listeners mapped T-R to M-R with higher response rates (IM: 2.68, EM: 1.37) than to other Mandarin tones. Inconsistency between the two groups was also observed for T-LF and T-FR. Whereas the IM listeners mapped T-LF mainly to M-HF (3.02) and T-FR to M-FR (2.87), the EM listeners mapped T-LF (2.93) and T-FR (3.52) with higher response rates. However, inconsistency between the two groups was also observed for T-LF and T-FR.

![Fig. 3. Mean degree of diversity for Thai tones being mapped onto Mandarin tones by the IM and EM listeners.](image-url)

**Table 6**

<table>
<thead>
<tr>
<th>Perceived as</th>
<th>Presented tones</th>
<th>T-LF</th>
<th>T-HF</th>
<th>T-R</th>
<th>T-FR</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-ML</td>
<td>IM</td>
<td>3.96</td>
<td>4.06</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>4.06</td>
<td>4.06</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>T-LF</td>
<td>IM</td>
<td>0.24</td>
<td>0.36</td>
<td>0.29</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>T-HF</td>
<td>IM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>T-R</td>
<td>IM</td>
<td>3.02</td>
<td>3.70</td>
<td>2.87</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>0.22</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>T-FR</td>
<td>IM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>EM</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
in terms of modal responses. Furthermore, L2 experience was found to affect the mappings of the Thai tones to the Mandarin tones, in particular the contour tones, as opposed to the level tones.

4.2.2.2. Degree of diversity. The $K'$ scores were obtained for the mappings of each Thai input tone, with the results displayed in Fig. 3. A mixed-design ANOVA was computed, using $K'$ as the dependent variable, experience group (IM and EM) as the between-group factor, and input tone (T-ML, T-LF, T-HF, T-R and T-FR) as the repeated measure. A significant main effect of tone was observed, $[F(4, 136)=16.5, p<.0001]$, whereas the effect of group, $[F(1, 34)=.49, p=.49]$ and the interaction between group and tone, $[F(4, 136)=1.6, p=.17]$, were not significant. Further Bonferroni-adjusted post hoc analyses indicated significantly less diversity for T-ML than T-LF ($p=.001$), T-R ($p<.0001$) and T-FR ($p<.0001$), and also significantly less diversity for T-HF than T-R ($p=.001$) and T-FR ($p<.0001$). Thus the diversity of mappings of each Thai tone showed no significant influence of L2 experience. However, Thai level tone and high falling tones were mapped more consistently to Mandarin tones than other Thai tone categories.

4.2.3. Summary

These results show that both inexperienced and experienced listeners tended to assimilate L2 tones to L1 tone categories with the most similar F0 features. As shown in Table 7, the inexperienced Thai (IT) and listeners experienced Thai (ET) assimilated M-HL to T-ML and M-HF to T-HF, and inexperienced Mandarin (IM) and experienced Mandarin (EM) listeners assimilated T-ML to M-HL, T-HF to M-HF and T-R to M-R. However, noteworthy differences in mapping patterns were also observed between the inexperienced and experienced listeners. For the Thai listeners, M-R was assimilated to T-R by the IT listeners, but to T-FR by the ET listeners, and M-FR was assimilated to T-FR by the IT listeners, but to T-LF by the ET listeners; while for the Mandarin listeners, T-LF was assimilated to M-HF by the IM listeners, but to M-FR by the EM listeners, and T-FR was assimilated to M-FR by the IM listeners and to M-R by the EM listeners. The group differences were confirmed by statistical evidence on the basis of percentage of response and similarity rating for the consistent mapping patterns.

Additionally, the results indicated differences in degree of mapping diversity between the IT and ET listeners. The ET listeners showed less diversity when mapping M-HL to T-ML and M-HF to T-HF than the IT listeners. In terms of individual tone patterns, level tones (M-HL and T-ML) and high falling tones (M-HF and T-HF) showed less diversity relative to other types of tones, in particular for the ET, IM and EM listeners.

5. Discussion

This study investigated tone assimilation by speakers of languages representing two different tonal systems (Mandarin and Thai) whose tone categories employ comparable F0 height and contour, as well as complex phonological variations (e.g., allotone). Previous research has suggested that speech perception can be affected not only by the acoustic properties of phonetic categories, but also by their linguistic functions in L1 and L2. In the current study, inexperienced and experienced listeners from both L1 backgrounds showed perceptual sensitivity to acoustic phonetic and phonological similarities between L1 and L2 tones. In particular, the inexperienced listeners assimilated L2 tones primarily to their L1 acoustic counterparts, while the experienced ones were sensitive to both acoustic and phonological influences. In the following sections, the results will be discussed with respect to the two research questions raised earlier.

5.1. How does language experience affect phonetic and phonological tone assimilation?

The current results indicate that both L1 and L2 experience had effects on tone assimilation for the listeners in this study. Despite considerable similarities between Mandarin and Thai tone systems, the slightly smaller range of choices (four vs. five) for the Mandarin listeners may have led to a lower degree of mapping diversity for that group compared to the Thai listeners. In particular,
the Mandarin listeners appeared to perceive L2 tones to be more similar to L1 tone categories than the Thai listeners, especially those from the inexperienced group. For instance, both IT and IM listeners assimilated L2 level and high falling tones to their L1 acoustic phonetic counterparts. The IT listeners, however, assimilated L2 tones to a more diverse range of L1 tone categories (e.g., T-R) than the IM listeners. Previous studies (e.g., Bohn & Best, 2012; Zheng et al., 2010) have proposed that a larger range of L1 phonetic categories may enhance listeners’ sensitivity to phonetic distinctions. For instance, Cantonese listeners showed greater sensitivity to both phonetic and phonological differences between tone categories compared to Mandarin listeners (Zheng et al., 2010). In the present study, Thai has a larger number of tone categories than Mandarin. The current results indicate that the IT listeners tended to assimilate L2 tones to a wider range of L1 tone categories, suggesting that more attention was paid to the dissimilarities between L1 and L2 tones compared to the Mandarin listeners. This finding supports previous research showing a difference in perceptual sensitivity due to the size of phonetic categories in L1 compared to L2.

Noteworthy differences between the experienced and inexperienced listeners were observed in both Mandarin and Thai groups. As discussed previously, these differences are associated with listeners’ sensitivity to tone categorization at the phonetic and phonological levels. While the inexperienced listeners showed sensitivity only to phonetic similarities, the experienced ones were sensitive to both phonetic and phonemic similarities between L1 and L2 tones. According to Flege (2007), L1 and L2 categories exist in a common space in the development of the L2 phonological system. As a result, it is very likely that not only the EM but also the ET listeners had underlying knowledge of the phonemic equivalence between falling rising tone and rising or low falling tone in Mandarin. Therefore, falling rising tones were primarily assimilated to their phonemic equivalents, i.e., rising and low falling tones. In this study, the difference between inexperienced and experienced groups was reflected in both modal response and rating data. Nevertheless, the degree of diversity showed comparable levels for inexperienced and experienced groups when assimilating allotonic variants (i.e., falling rising, rising and low falling tones) to their L1 counterparts. This can be accounted for by the perceptual difficulties arising from acoustic similarities between the three tones (see also discussion in Section 5.2). As suggested by Flege (1988, 1990), L2 phonetic categories that are similar but not identical to L1 phones are most problematic to perceive correctly. The perceptual difficulty may lead L2 tones to be mapped diversely to L1 tone categories instead of to one specific tone category.

5.2. What factors are predictive of tone assimilation at the phonetic and the phonological levels?

The outcomes of this study align with previous findings that assimilation processes can be realized at both the phonetic and the phonological levels (Best & Strange, 1992). Phonetic perception relates to listeners’ sensitivity to acoustic characteristics or articulatory gestures, while phonological perception depends on the recognition of the linguistic functions which phonetic categories serve in speech communication. Listeners from the four groups showed great sensitivity to the acoustic similarities between L1 and L2 tone categories. First, L1 and L2 tones with comparable F0 height and contour were perceived as most similar, such as high falling tones and rising tones. Second, even with only partial features in common, such as F0 contour, L2 tones tended to be assimilated to similar L1 contour tones, thus M-HL was assimilated to T-ML an T-LF to M-HF. It should be noted that the assimilation of L2 tones to L1 tone categories with partial acoustic features might be accounted for by the fact that listeners were obliged to select an L1 tone to map the input L2 tone in a forced-choice identification task.

However, it was also probable that listeners made their choices from multiple L1 tone categories with similar partial features such as F0 at the extremities of a tone contour, as discussed by Guion and Pederson (2007). In particular, M-HL and T-R both start with a high F0, and M-HL and T-HF both start with a high F0. However, both Mandarin and Thai listeners paid little attention to the F0 level, suggesting greater sensitivity to F0 contour than F0 height, consistent with previous findings (cf., Gandour, 1983). On the one hand, this finding appears to support Flege’s (1988, 1990) proposal that listeners are more confident when assimilating L2 categories which are identical to their L1 counterparts than those which partially share acoustic properties (“similar phones”). On the other hand, it may suggest that listeners are likely to associate a particular L2 phone with an L1 phone similar (or “dissimilar”) to it simply because there is no better match. In terms of the PAM, assimilation favored the closer match for an L1 tone with two possible phonetic counterparts in L2, an instance of Category Goodness assimilation.

In addition to the phonetic assimilation process described above, the experienced listeners appeared to assimilate L2 tones to L1 tone categories on the basis of phonological factors arising from the allophonic tone changes in Mandarin (Gandour, 1981; Huang, 2001). As reviewed earlier, M-FR is realized as M-R and a low falling tone under certain conditioned contexts. Thus M-R and low falling tone are allophonic variants of M-FR. In this study, as expected, the experienced listeners assimilated L2 rising or low falling tone to L1 rising falling tone and vice versa. For instance, the EM listeners assimilated T-LF to M-FR and T-FR to M-R, while the ET listeners assimilated M-R to T-FR and M-FR to T-LF. These outcomes echo previous findings of tone assimilation by native and experienced listeners (Huang, 2001; Leung, 2008). In other words, phonological assimilation may emerge from L2 exposure, which enhances the acquisition of allophonic variation.

However, the finding that the IM listeners appeared to perceive Thai tones less frequently as phonological equivalents despite their familiarity with allophonic variants in L1, contradicts So and Best (2010), who reported that inexperienced listeners’ perception of L2 tones was also associated with phonemically equal but acoustically distinct tone productions in L1. The discrepancy may arise from a higher degree of acoustic similarity between Mandarin high level and high falling tones, specifically the initial F0 height of both tones (e.g., Guion & Pederson, 2007). As observed by Gandour (1983), Cantonese listeners were more sensitive to F0 height compared to native speakers of other tone languages (e.g., Thai and Mandarin). Therefore, the inexperienced Cantonese listeners’ sensitivity to the phonological correspondence between high level and high falling tones can be accounted for by their acoustic
Similarities. Furthermore, the discrepancy may be explained by the fact that the Cantonese tone system encompasses more tone categories and more complex phonological tone changes relative to many other tone systems, such as Thai and Mandarin. As suggested by Zhang et al. (2010), a ‘denser’ L1 tone system can improve phonetic and phonological processing of tone categories in other languages, a view also supported by Bohn and Best (2012), who propose that a larger L1 phonetic system may help listeners distinguish unfamiliar L2 phonological categories. Finally, the discrepancy may be accounted for by the difference in phonological status of the allotones in the two languages. In Cantonese, high falling and high level tones are unconditioned free variants, and high falling tone is not an independent tone category (Tse, 1978). In Mandarin, however, falling rising tone is realized as rising tone or low falling tone under certain conditions. Additionally, rising tone exists in Mandarin tone system as an independent category. Therefore, the Mandarin allotones are phonologically more distinctive compared to the Cantonese ones. As a result, the Cantonese allotones tend to be more difficult to discriminate, in particular when the input tone is a high falling tone.

The current results also point to the interplay of the lower-level acoustic and higher-level phonological influences on tone assimilation for the experienced listeners. The experienced listeners had 0.5–2 years of L2 experience. According to Derwing and Munro (2013), L2 speech skills typically improve in the first 1–2 years of L2 experience. The assimilation findings are testament to the fact that the experienced listeners have acquired L2 phonological tone categories and associated them with L1 phonological tone categories. However, the current study also found that ET listeners assimilated M-FR not only to its allotonic variant T-LF, but also to its acoustic counterpart T-FR. Moreover, the EM listeners assimilated T-R primarily to its acoustic equivalent M-R rather than its allotonic variant, M-R. These findings support Best and Tyler’s (2007) view that perceptual assimilation occurs at phonetic and phonological levels for experienced listeners. However, they also suggest that phonetic similarities continue to exert influence on speech assimilation for intermediate L2 learners (cf. naïve listeners, Bohn & Best, 2012; Hallé et al., 1999). With relatively long-term L2 exposure, listeners may form a common space for L1 and L2 phonological systems (Flege, 1995). It is likely that newly formed L2 phonological categories may be associated with their L1 phonological and phonetic equivalents simultaneously or with more importance being attached to one type; thus discrepancies can be observed when phonetic and phonological categories are different, such as T-FR and T-LF in the current study.

Despite the findings for assimilation patterns, tone perception at the acoustic phonetic and phonological levels seemed to be less distinctive in terms of degree of diversity. First, no difference in diversity was detected between the IT and ET listeners for assimilations of M-R to T-R and T-FR, or M-FR to T-LF and T-FR. Also, no difference was observed for T-FR mappings to M-R and M-FR for either IM or EM listeners. Therefore, degree of diversity measures indicated more diverse mappings for the tones entailing allotonic tone changes (e.g., falling rising, rising and low falling tones) compared to those without (i.e., level and high falling tones). It appears, then, that the interactions of acoustic and phonological influences increase the perceptual difficulties in distinguishing L2 tone categories for the experienced listeners. For the inexperienced listeners, the findings that they gave high similarity ratings to tone and its allotonic variants may be explained by the partial acoustic similarities between these tones, while group differences in assimilation patterns suggest that tone assimilation was associated with both phonetic and phonemic correspondence between L1 and L2 tones.

6. Concluding remarks

The current study provides evidence supporting the view that recent L2 speech perception models can be extended to the suprasegmental level. The results indicate a significant and systematic influence of language experience on tone assimilation, revealing that perceptual sensitivity to the distinctions between tone categories depends on previous exposure to a tone system. Allophonic tone variation is found to be tied to tone assimilation at the phonological level. However, as this phenomenon does not exist in Thai, in the current study it was possible to examine tone assimilation only in relation to the Mandarin tone system. In future work, researchers should consider investigating tone assimilation in language pairs employing different tone systems that both carry allophonic tone contrasts, such as Mandarin and Cantonese (Yu, 2007; Zee, 1999). Additionally, phonological tone assimilation may also be explored by replacing L1 tone categories with L2 tone categories on a number of real words in L1 (Best & Tyler, 2007). Given only a difference in tone, if the replacement (e.g., M-HL replaced by T-ML for a Mandarin word) does not affect word meanings, such a finding may suggest that the two tone categories share the same phonological function.

Future research may also investigate assimilation of other prosodic features, such as Japanese pitch accent and English word stress. Like lexical tones, these features are realized at the word level with distinctive phonetic properties (Wu, Tu, & Wang, 2012); thus it is worth comparing the effects of phonetic and phonological factors in the assimilation of different prosodic patterns. Moreover, since suprasegmental assimilation is linked with listeners’ L2 experience, further work may incorporate L2 experience at different stages, allowing for a longitudinal observation of the establishment of an L2 phonological system or the formation of L2 phonological categories.

Acknowledgments

This project was supported by a doctoral fellowship to the first author from Social Sciences and Humanities Research Council of Canada (752-2009-1922). Some of the results were presented at the 17th International Congress of Phonetic Sciences (ICPhS XVII) and the 159th meeting of the Acoustical Society of America. We thank the Beijing College of Chinese Language and Culture in China and the University of the Thai Chamber of Commerce in Thailand for their help with participant recruitment.
Appendix A

See Table A1 and A2.

Table A1
Mapping matrix for the four Mandarin tones perceived by Inexperienced Thai (IT) and Experienced Thai (ET) listeners, with mean percent responses (%) and mean similarity ratings (1-5, in parentheses) for the five Thai tone alternatives.

<table>
<thead>
<tr>
<th>Perceived as</th>
<th>Presented tones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IT</td>
</tr>
<tr>
<td>M-HL</td>
<td>T-ML</td>
</tr>
<tr>
<td>M-R</td>
<td>T-LF</td>
</tr>
<tr>
<td>M-FR</td>
<td>T-HF</td>
</tr>
<tr>
<td>M-HF</td>
<td>T-R</td>
</tr>
<tr>
<td></td>
<td>T-FR</td>
</tr>
</tbody>
</table>

Table A2
Assimilation matrix for the five Thai tones perceived by Inexperienced Mandarin (IM) and Experienced Mandarin (EM) listeners, with mean percent responses (%) and mean similarity ratings (1-5, in parentheses) for the four Mandarin tone alternatives.

<table>
<thead>
<tr>
<th>Perceived as</th>
<th>Presented tones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IM</td>
</tr>
<tr>
<td>M-HL</td>
<td>T-ML</td>
</tr>
<tr>
<td>M-R</td>
<td>T-LF</td>
</tr>
<tr>
<td>M-FR</td>
<td>T-HF</td>
</tr>
<tr>
<td>M-HF</td>
<td>T-R</td>
</tr>
</tbody>
</table>

References


