International Journal of English Linguistics; Vol. 9, No. 4; 2019
ISSN 1923-869X E-ISSN 1923-8703
Published by Canadian Center of Science and Education

Influence of L1 Background on Categorical Perception of Mandarin Tones by Russian and Vietnamese Listeners

Qiongqiong Zou[1]

[1] College of Foreign Languages, Hunan University, Changsha, China

Correspondence: Qiongqiong Zou, College of Foreign Languages, Hunan University, No. 2 Lushan South Road, Yuelu District, Changsha, 410082, China. E-mail: zouqiong@hnu.edu.cn

Received: April 30, 2019 Accepted: May 28, 2019 Online Published: July 3, 2019

doi:10.5539/ijel.v9n4p275 URL: https://doi.org/10.5539/ijel.v9n4p275

Abstract

This study investigated the influence of L1 background on categorical perception of lexical tones by three language groups, namely native Mandarin, Russian and Vietnamese listeners. Tone identification and discrimination scores of two tone continua (T1-T2 and T1-T4) were measured for each participant. Results showed that the two tone language groups, i.e., Mandarin and Vietnamese listeners, perceived both tone continua categorically whereas the non-tone language group, i.e., Russian listeners, did not. More specifically, while the Russian group exhibited significantly broader identification boundaries and performed near chance level in discrimination tasks, the Mandarin and Vietnamese groups presented sharp slopes in identification curves and corresponding discrimination peaks at the cross-boundary positions. Moreover, Mandarin and Vietnamese listeners showed slightly different discrimination curves, which could be attributed to the effect of their different tone inventories. The current findings suggest that native tone language background, to some extent, can facilitate non-native tone perception.

Keywords: categorical perception, tone, language background, identification, discrimination

1. Introduction

Languages in the world can be generally divided into tone and non-tone languages, which differ in the use of pitch (fundamental frequency (F0) as its acoustic correlate) in their prosodic systems (Yip, 2002). Most tone languages, such as Mandarin Chinese, Thai and Vietnamese, employ pitch height and/or pitch contour at the word level to minimally distinguish lexical meanings between otherwise segmentally identical words (Gandour, 1978). However, non-tone languages, including English, French and Russian, use pitch variations in a broader way. Specifically, pitch can be used at the word, phrase, or sentence level (intonation) in these non-tone languages to cover various communication intentions (So & Best, 2014).

Since these two broad language types diverge in the use of pitch information, previous studies have revealed that pitch variation in a tone language, i.e., tone contrasts, are perceived differently by tone and non-tone language listeners and that listeners tend to make use of their native language prosodic systems when perceiving pitch variations in a non-native language (Huang & Johnson, 2010; So & Best, 2010; Sun & Huang, 2012; Wang, 2013; Wayland & Guion, 2003). Although a large number of studies have been conducted to investigate the influence of first language (L1) background on the perception of second language (L2) tones, their findings were still not consistent. Some reported that tone language speakers performed better than non-tone language speakers in perceiving L2 tones (Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004), but others held the reverse opinion (So, 2005; Wang, 2006). Meanwhile, there also exists another point of view that native tone and non-tone language speakers did not differ significantly in the perceptual accuracy of L2 tones (Francis, Ciocca, Ma, & Fenn, 2008; Hao, 2012).

Categorical perception (CP) serves as a more sensitive approach to addressing the issue of how lexical tones are perceived by native tone language speakers as compared to speakers of non-tone languages. As a well-known cognitive phenomenon, CP was first put forward by Liberman, Harris, Hoffman, and Griffith (1957), which posits that a continuum with several equivalently separated stimuli is perceived as a discrete and finite number of categories and between-category differences can be noticed more saliently than within-category differences. Typical categorical perception should be assessed by a sharp identification boundary between two categories and a corresponding discrimination peak at the category boundary position (Abramson, 1978; Liberman et al., 1957).

Earlier studies on categorical perception of speech sounds have mainly been directed to segmental features, i.e., consonants and vowels. While most previous research showed that stop consonants are perceived categorically (Brandt & Rosen, 1980; Carne, Widin, & Viemeister, 1977; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Liberman et al., 1957; Pisoni, 1977; Rosner, 1984; Zlatin, 1974), people tend to perceive vowels in a more continuous manner (Altmann et al., 2014; Fry, Abramson, Eimas, & Liberman, 1962). In addition to the great bulk of research on categorical perception of segmental features, a large amount of literature on CP of suprasegmental features has appeared recently, such as lexical tones in tone languages (Francis, Ciocca, & Ng, 2003; Halle, Chang, & Best, 2004; Peng et al., 2010; Shen & Froud, 2016; Wang, 1976; Xu, Gandour, & Francis, 2006). These CP studies mainly focused on the influence of language experience on the categorical perception of L2 tones. For instance, Wang (1976) investigated the perception of lexical tones by native Mandarin and English listeners. The results indicated native Mandarin listeners perceived lexical tones in a clearly categorical manner whereas English listeners did not. Similar findings that contrasted native tone language listeners with non-tone language listeners were also reported in other studies (Halle, Chang, & Best, 2004, for Mandarin vs. French; Peng et al., 2010, for Mandarin and Cantonese vs. German; Xu, Gandour, & Francis, 2006, for Mandarin vs. English).

Besides, different tone systems could further impact non-native tone perception in terms of CP. Peng et al. (2010) examined both the effect of tone language vs. non-tone language background (German vs. Chinese), and the effect of different tone systems (Mandarin vs. Cantonese), on the categorical perception of Mandarin lexical tones. Their results showed that German listeners exhibited a continuous pattern of tone perception, whereas both Mandarin and Cantonese listeners perceived lexical tones categorically, yet with different patterns shaped by their own tone inventory. Zheng et al. (2012) carried out an event-related potentials study to explore the influence of tone systems on the categorical perception of lexical tones. Their findings suggested that language background did modulate Mandarin and Cantonese listeners' behavioral and electrophysiological responses to the tone stimuli presented, and Cantonese listeners showed stronger sensitivity to pitch height and pitch slope than native Mandarin listeners in order to discriminate their denser tone inventory.

To sum up, compared with a great deal of research on the differences in tone perception between native tone language and non-tone language listeners, there is still a paucity of empirical evidence on how L1 tones affect the categorical perception of non-native tones. What's worse, we can find that the tone languages involved in the above CP studies were either Mandarin or Cantonese, which are both variations of Chinese language. Since Mandarin is the official language in China, it seems impossible that native Cantonese speakers have never been exposed to Mandarin, or cannot speak Mandarin at all. Therefore, it is still unknown how native tone language listeners from other countries perceive Mandarin tones. To this end, the current study explores categorical tone perception by listeners from two rarely discussed languages, namely Russian (non-tone language) and Vietnamese (tone language which has different tone system from Mandarin). Using the traditional paradigm of CP, the current study aims to figure out how Russian and Vietnamese listeners perceive Mandarin tones respectively and then to reexamine the influence of language background on the categorical perception of lexical tones, thus adding convincing evidence to the aforementioned disputable topic whether tone language background could facilitate or hinder non-native tone perception.

2. Mandarin and Vietnamese Tone Systems

Mandarin, the official language in China, is probably one of the most typical tone languages in the world (Wang, 1973). It has four regular tones: Tone 1 (T1)—high level tone (55), Tone 2 (T2)—mid rising tone (35), Tone 3 (T3)—low falling rising tone (214), and Tone 4 (T4)—high falling tone (51). The numbers in parentheses suggest the relative starting and ending pitch of each tone on a 1–5 scale, with 1 referring to the lowest pitch of the speaker and 5 to the highest pitch (Chao, 1948).

Compared with Mandarin, Vietnamese has a more complex tone system, as shown in Table 1. It has six main tones: the ngang tone "level" (44), the huyền tone "deep" (32), the ngã tone "tumbling" (325), the hỏi tone "asking" (323), the sắc tone "sharp" (45), the nặng tone "heavy" (21).

Table 1. Mandarin and Vietnamese tone systems

Mandarin	Vietnamese
	The ngang tone “level” (44)
Tone 1—high level (55)	The huyền tone “deep” (32)
Tone 2—mid rising (35)	The ngã tone “tumbling” (325)
Tone 3—low falling rising (214)	The hỏi tone “asking” (323)
Tone 4—high falling (51)	The sắc tone “sharp” (45)
	The nặng tone “heavy” (21)

Although there does not exist a completely identical tone in Mandarin and Vietnamese, it is easy to find great similarities. Firstly, the ngang tone in Vietnamese and Tone 1 in Mandarin are both level tones, but the latter has a higher pitch value than the former. Secondly, the huyền tone and the nặng tone in Vietnamese, together with Tone 4 in Mandarin, are all falling tones. However, with the pitch falling from 5 to 1, Mandarin Tone 4 has the greatest degree of pitch variation among these three falling tones. Thirdly, the sắc tone in Vietnamese bears a high resemblance to Mandarin Tone 2, both of which are rising tones. Lastly, there remain the ngã tone, the hỏi tone in Vietnamese and Tone 3 in Mandarin, all of which are contour tones, and share similar pitch direction.

In the present study, the perceptual performance for two tone continua, i.e., Tone 1-Tone 2 (T1-T2) and Tone 1-Tone 4 (T1-T4), are compared across three language groups: native Mandarin listeners, native Russian listeners and native Vietnamese listeners. Based on previous findings and the tone systems of Mandarin and Vietnamese, the following hypotheses are made: 1) Mandarin and Vietnamese listeners would perceive lexical tones in a categorical manner whereas Russian listeners would perceive the two tone continua in a continuous manner. 2) Mandarin and Vietnamese listeners would exhibit different categorical patterns of tone perception due to their different tone inventories.

3. Method

3.1 Participants

Ten native Mandarin listeners (5 females and 5 males, mean age = 23.5 yr, SD = 1.4), 10 native Russian listeners (6 females and 4 males, mean age = 21.0 yr, SD = 3.5), and 10 native Vietnamese listeners (5 females and 5 males, mean age = 25.1 yr, SD = 1.8) were recruited for this study.

A background questionnaire was collected to gather information about their language experience. All native Mandarin speakers were students from Hunan University in Changsha, who were born and grew up in Northern China. All of them scored more than 89 in National Standard Mandarin Examination. Participants who spoke dialects were excluded to avoid dialectical effects. The native Russian and Vietnamese speakers were international students in Changsha. They had been learning Mandarin Chinese for about six months and had never been exposed to Mandarin or other tone languages before (except their native language for Vietnamese speakers).

None of the participants had ever received any professional music training. All of them were confirmed with no speech or hearing disorders or any other language problems. They were paid for their participation and were given informed consent in compliance with a protocol approved by Human Research Ethics Committee at Hunan University.

3.2 Materials

The target syllable /ta/ with Mandarin Tone 1, Tone 2 and Tone 4 was recorded by an adult female native speaker of Mandarin Chinese at a sampling rate of 22050 Hz, 16-bit resolution. Two tone continua (T1-T2 and T1-T4) were constructed based on the natural speech templates of these three tones. Each continuum consisted of eleven steps with equal distances re-synthesized by applying the pitch-synchronous overlap and add (PSOLA) method (Moulines & Laroche, 1995) implemented in Praat (Boersma & Weenink, 2018). Figure 1 presents a schematic diagram of the pitch contours of the eleven stimuli for the T1-T2 continuum (on the left) and the eleven stimuli for the T1-T4 continuum (on the right) (following Peng et al., 2010; Wang, 1976). In Mandarin, these three tones over the target syllable /ta/ were all real words yet with different lexical meanings: “搭 (build) ” when produced with the high level tone (T1), “答 (answer) ” when produced with the mid rising tone (T2), and “大 (big) ” when produced with the high falling tone (T4).

The main procedures for re-synthesizing the stimuli were as follows: (1) Scaling the target syllable with these three tones to equal duration of 500 ms and equal intensity of 70 dB. (2) Fixing the pitch contour of Tone 1 to the level frequency of 270 Hz and both the starting frequency of Tone 2 and the ending frequency of Tone 4 to 210 Hz. (3) Setting the number of pitch points to three, with the first one at the starting position, the second one at the 100 ms position, and the third one at the ending position. (The fundamental frequency of the first 100 ms was kept constant in order to make the re-synthesized stimuli more natural.) (4) Synthesizing the two tone continua by dragging the aforementioned three pitch points to their respective pitch value (following Peng et al., 2010). The starting frequency of each stimulus in the T1-T2 continuum and the ending frequency of each stimulus in the T1-T4 stimuli were calculated by the formula 210 Hz＋6 Hz × (Stimulus Number－1). All manipulations of the natural speech recordings were conducted in Praat (Boersma & Weenink, 2018).

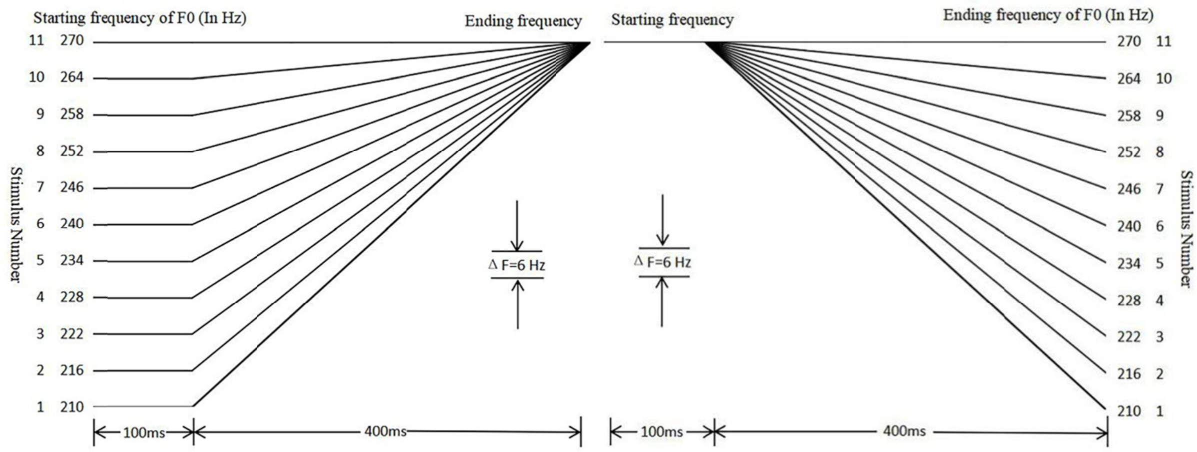

Figure 1. Schematic diagram of the pitch contours for the stimuli in the T1-T2 continuum (on the left) and in the T1-T4 continuum (on the right).

3.3 Procedures

All participants were instructed to complete two separate experimental tasks, i.e., an identification task and a discrimination task. Each participant was tested individually in a quiet room, accompanied by the experimenter. Both tasks were administered using Praat (Boersma & Weenink, 2018).

3.3.1 Identification Task

A practice block (including fourteen stimuli, Number 1, Number 2, Number 5, Number 6, Number 7, Number 10, Number 11, repeating twice randomly) (following Chen, Peng, Yan, & Wang, 2016) was provided before the test block of each continuum in order to familiarize participants with the identification procedure. For the T1-T2 continuum, the participants were instructed to click the option "Tone 1" when they heard a sound which bears the level tone and to click the option "Tone 2" when they heard a sound which bears the mid-rising tone (two-alternative forced choice, 2AFC). Similarly, for the T1-T4 continuum, the participants were instructed to click the option "Tone 1" when they heard a sound bearing the level tone and to click the option "Tone 4" when they heard a sound bearing the falling tone (two-alternative forced choice, 2AFC). Once a response was made, the next stimulus would be presented automatically following a 500 ms pause. The eleven stimuli were randomly played eight times in the test block, resulting in a total of eighty-eight test trials for each continuum. No feedback was provided for the participants during both the practice and test blocks. The order of continuum presentation was randomized among participants.

3.3.2 Discrimination Task

All participants were instructed to finish an AX two-step discrimination task. There was also a practice block which consisted of twelve stimulus pairs (5-7, 7-5, 6-8, 8-6, 1-1, 11-11, repeating twice randomly) (following Chen et al., 2016). The inter-stimulus interval (ISI) within each stimulus pair was fixed to 500 ms. After hearing each stimulus pair, the participants were required to determine whether the two stimuli bear the same tone or different tones by clicking the "Same" or "Different" option. Once a response was made, the next stimulus pair would be presented automatically following a 500 ms pause. Since AX test trials were supposed to have four possible combinations, i.e., AB, BA, AA, BB, a total of 29 stimulus pairs were presented for the test block of each continuum. Among these 29 stimulus pairs, 18 pairs were made up of two different stimuli separated by two steps (1-3, 3-1, 2-4, 4-2 ... 8-10, 10-8, 9-11, 11-9) and the other 11 pairs were made up of the 11 stimuli along the continuum each paired with itself (1-1, 2-2 ... 10-10, 11-11). These 29 pairs were randomly presented five times in the test block, leading to 145 pairs in total for each continuum. No feedback was provided for the participants during both the practice and test blocks. The order of continuum presentation was also randomized among participants.

3.4 Data Analysis

To investigate the influence of language background on categorical perception of lexical tones, three essential parameters of CP were calculated for each participant: category boundary position, boundary width, and discrimination accuracy.

For each stimulus, the identification score referred to the average percentage of T1 or T2 responses for the T1-T2 continuum and the average percentage of T1 or T4 responses for the T1-T4 continuum. Boundary position and boundary width were calculated using Probit analyses (Finney, 1971). The boundary position referred to the 50% crossover point of the identification score and the boundary width referred to the linear distance between the 25% position and 75% position (Xu, Gandour, & Francis, 2006).

To measure the discrimination scores, the 145 discrimination trials were divided into nine stimulus pairs. As mentioned above, each stimulus pair had four possible combinations, namely AB, BA, AA, BB, among which, AB and BA represented different pairs, whereas AA and BB represented same pairs (Xu, Gandour, & Francis, 2006). Overlapping AA or BB pairs were included in adjacent stimulus pairs. For example, the 6-6 pair was included in both 4-6 and 6-8 stimulus pairs. The discrimination accuracy (P) for each stimulus pair was obtained by the following formula (Xu, Gandour, & Francis, 2006):

$$P = P(\text{'S'}/S) \times P(S) + P(\text{'D'}/D) \times P(D) \tag{1}$$

P('S'/S) represented the percentages of "same" ('S') responses to all the same (S) trials and P('D'/D) represented the percentage of "different" ('D') responses to all the different (D) trials. That is to say, P('S'/S) and P('D'/D) denoted the percentages of correct responses. P(S) and P(D) stood for the percentages of "same" and "different" trials in each stimulus pair, which were both 50% in the current study (Peng et al., 2010).

4. Results

4.1 T1-T2 Continuum

Identification and discrimination curves for the T1-T2 continuum among different language groups are depicted in Figure 2 and the estimated boundary position and boundary width are shown in Table 2. As demonstrated in Figure 2, the two tone language groups, namely Mandarin and Vietnamese listeners, perceived the T1-T2 continuum categorically, i.e. steep slopes in identifications curves and corresponding discrimination peaks. However, the identification curves for Russian listeners just exhibited gentle slopes and their discrimination peak did not correspond with the category boundary, suggesting that the Russian group perceived Mandarin Tone 1 and Tone 2 in a continuous manner.

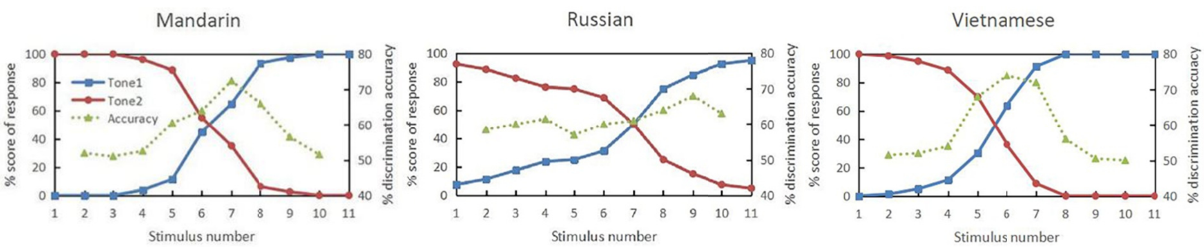

Figure 2. The identification curves (solid lines, with square dots denoting T1 response and round dots denoting T2 response) and discrimination curves (dashed lines with triangle dots).

The left y-axis represents the percentage score of T1 or T2 response, while the right y-axis represents the overall discrimination accuracy.

Table 2. Derived categorical boundary position and width for the T1-T2 continuum

Group	Position		Width	
	Mean	*SD*	*Mean*	*SD*
Mandarin	6.23	0.65	1.24	0.59
Russian	6.10	0.69	4.35	3.19
Vietnamese	5.44	0.74	1.34	0.44

One-way ANOVA was conducted to examine the influence of language background on boundary position, boundary width, and discrimination accuracy. Tukey's HSD post hoc analysis was used to make pairwise comparisons when necessary. All effects were reported at a significant level of $p < 0.05$.

A significant effect of *group* was found both for the perceptual boundary position [F (2, 27) = 3.716, p = 0.038, η^2 = 0.216] and boundary width [F (2, 27) = 8.764, p = 0.001, η^2 = 0.394]. Tukey's HSD post hoc pairwise comparisons of the three groups revealed that the perceptual boundary position for the Mandarin group occurred at a significantly larger stimulus number than that for the Vietnamese group (p = 0.044), but that the boundary positions perceived by Mandarin and Russian listeners were not significantly different (p = 0.915). As for boundary width, Tukey's HSD post hoc pairwise comparisons indicated that both Mandarin and Vietnamese listeners had significantly narrower boundary widths than Russian listeners (ps < 0.01), but no significant difference in boundary width was found between the Mandarin and Vietnamese listeners (p = 0.993). It indicated that Vietnamese learners, with a tone background, had sharper identification boundaries between Mandarin Tone 1 and Tone 2 than Russian learners with a non-tone background.

With respect to the discrimination task, the overall accuracies of the nine stimulus pairs among the three language groups are shown in Figure 3. First, the discrimination accuracy for the Mandarin group reached its maximum at the stimulus pair 6-8, which straddled its own boundary position (6.23). One-way ANOVA confirmed a significant effect of *group* for the discrimination accuracy at this position [F (2, 27) = 5.275, p = 0.012, η^2 = 0.281]. Tukey's HSD post hoc analysis revealed that the accuracies for both the Mandarin group (72.5%) and the Vietnamese group (72%) were significantly higher than that for the Russian group (61%) (ps < 0.05) at this position, but there was no significant difference between the Mandarin and Vietnamese groups (p = 0.991). Second, the discrimination peak for the Vietnamese group was 74% at the stimulus pair 5-7, also straddling its boundary position (5.44). One-way ANOVA also demonstrated a significant difference in the discrimination accuracy at this position across the three language groups [F (2, 27) = 4.634, p = 0.019, η^2 = 0.256]. Tukey's HSD post hoc comparisons showed that the accuracy for the Vietnamese listeners (74%) was significantly higher than that for the Russian listeners (60%) (p = 0.017). However, no significant difference was observed between the Mandarin and Vietnamese listeners or between the Mandarin and Russian listeners (ps > 0.05). Third, the discrimination peak for the Russian group was 68% at the stimulus pair 8-10, missing its boundary position (6.10). Results of one-way ANOVA indicated a significant effect of *group* at this position [F (2, 27) = 19.680, p < 0.001, η^2 = 0.593]. Tukey's HSD post hoc pairwise comparisons showed that the discrimination accuracy for the Russian group (68%) was significantly greater than those for the Mandarin (56.5%) and the Vietnamese (50.5%) groups (ps < 0.01). But Mandarin and Vietnamese listeners did not exhibit significantly different discrimination accuracies at this position (p = 0.105). All these results showed that the discrimination peaks of native tone language listeners corresponded with their respective boundary positions, suggesting linguistic boundaries. However, the discrimination peak of the Russian listeners was located at the level end, reflecting the psychophysical boundaries.

Since one of the most crucial characteristics of CP is better discrimination performance across category boundaries than within the same category, these nine stimulus pairs were further divided into between-category pairs and within-category pairs. As shown in Figure 4, the mean within-category discrimination accuracy for the Mandarin, Russian, and Vietnamese listeners was 55.7%, 61.7%, 55.2% respectively, and their respective average between-category discrimination accuracy was 68.3%, 60.5%, and 71%. The results of paired samples t test revealed that both the Mandarin group [t (9) = 3.933, p = 0.003] and the Vietnamese group [t (9) = 7.186, p < 0.001] had significantly higher between-category discrimination accuracy than within-category discrimination accuracy. No significant difference was found for the Russian group [t (9) = -0.794, p = 0.448]. Such findings showed that between-category differences could be perceived as more salient than within-category differences by Mandarin and Vietnamese listeners but not by Russian listeners, implying that native tone language listeners could perceive Mandarin Tone 1 and Tone 2 more categorically than native non-tone language listeners.

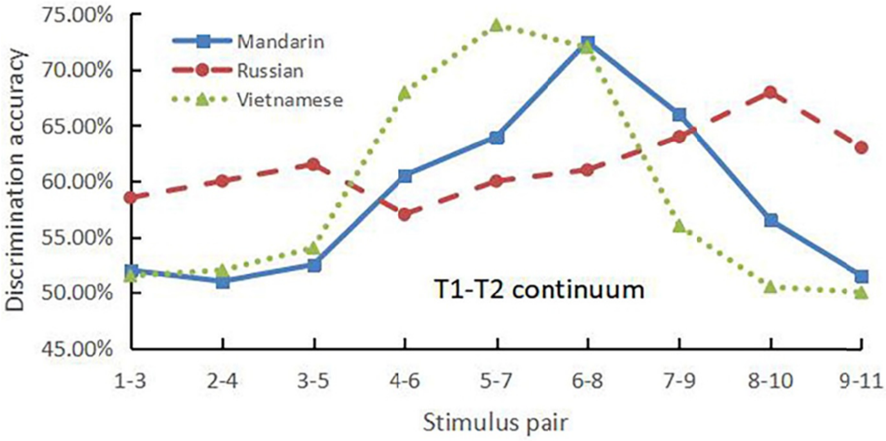

Figure 3. The discrimination accuracy of nine stimulus pairs across the three language groups in the T1-T2 continuum.

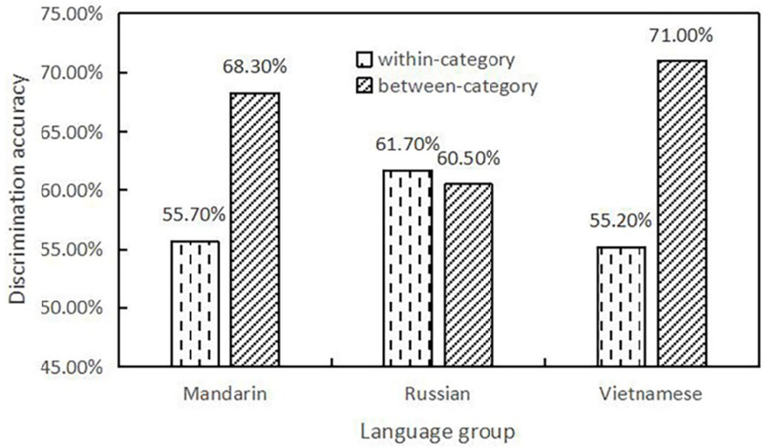

Figure 4. Within-category and between-category discrimination accuracy across the three language groups in the T1-T2 continuum.

4.2 T1-T4 Continuum

A similar statistical analysis was applied to the T1-T4 continuum. Figure 5 depicts the identification and discrimination curves for each language group and Table 3 shows their respective boundary position and width.

One-way ANOVA yielded a significant effect of *group* for both the boundary position [F (2, 27) = 7.165, p = 0.003, $\eta^2 = 0.347$] and boundary width [F (2, 27) = 5.344, p = 0.011, $\eta^2 = 0.284$]. Tukey's HSD post hoc analysis revealed that native Mandarin and Russian listeners were significantly different in both perceptual boundary position and boundary width, with the non-tone language group having significantly smaller boundary positions (p = 0.002) and broader boundary widths (p = 0.010). However, the differences between the Mandarin and Vietnamese groups were not statistically significant. Such results revealed that for the T1-T4 continuum, tone language listeners also exhibited steeper slopes in identification curves than non-tone language listeners.

In terms of the discrimination task, as Figure 6 displays, the discrimination accuracies for the Mandarin and Vietnamese listeners reached their maxima at the stimulus pair 6-8 (71.5%) and 7-9 (68%) respectively, which straddled their own boundary position (Mandarin: 7.53; Vietnamese: 7.13). One-way ANOVA yielded a marginally significant effect of *group* at stimulus pair 6-8 [F (2, 27) = 3.251, p = 0.054, $\eta^2 = 0.194$]. Tukey's HSD post hoc comparisons indicated that the accuracy for the Mandarin group (71.5%) was marginally significantly greater than that for the Russian group (63%) (p = 0.050). However, no significant difference in discrimination accuracy was observed across these three language groups at the stimulus pair 7-9 [F (2, 27) < 0.001, p = 1.000, $\eta^2 < 0.001$]. In addition, the discrimination peak for the Russian group was 71% at the stimulus pair 8-10, missing its boundary position (6.66). Similarly, no significant difference at this position was found across the three language groups [F (2, 27) = 2.523, p = 0.099, η^2 = 0.157]. Hence, native Mandarin and Vietnamese listeners exhibited linguistic boundaries in discrimination curves for the T1-T4 continuum whereas the Russian listeners exhibited only psychophysical boundaries.

A further comparison was also conducted for between-category pairs and within-category pairs. As shown in Figure 7, the mean within-category discrimination accuracy for the Mandarin, Russian, and Vietnamese listeners was 54.4%, 62%, 55.3% respectively, and their respective average between-category discrimination accuracy was 69.8%, 58%, and 66.8%. Similar to the T1-T2 continuum, the results of paired samples t test revealed that both the Mandarin group [t (9) = 9.507, p < 0.001] and the Vietnamese group [t (9) = 5.157, p = 0.001] had significantly higher between-category discrimination accuracies than within-category discrimination accuracies in the T1-T4 continuum. Still no such significant difference was observed for the Russian group [t (9) = -1.746, p = 0.174], suggesting that Russian listeners dealt with between-category pairs and within-category pairs in the same way.

In conclusion, from the identification curves shown in Figure 2 and Figure 5, we could see clearly that changes from one tone to another were more abrupt for the two tone language groups. As for the discrimination curves, between-category discrimination accuracies were significantly higher than within-category discrimination accuracies for native Mandarin and Vietnamese listeners, indicating that native tone language listeners showed stronger sensitivity to between-category differences than to within-category differences. However, compared with these two tone language groups, native Russian listeners perceived both T1-T2 and T1-T4 continua more continuously, with gentle slopes in identification curves and no salient peaks in discrimination curves.

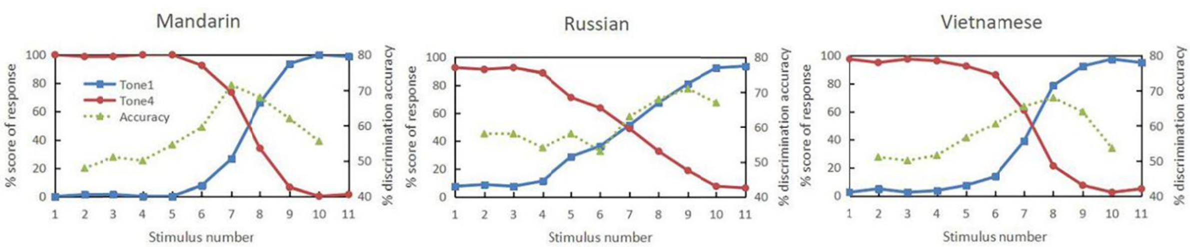

Figure 5. The identification curves (solid lines, with square dots denoting T1 response and round dots denoting T4 response) and discrimination curves (dashed lines with triangle dots).

The left y-axis represents the percentage score of T1 or T4 response, while the right y-axis represents the overall discrimination accuracy.

Table 3. Derived categorical boundary position and width for the T1-T4 continuum

Group	Position		Width	
	Mean	***SD***	***Mean***	***SD***
Mandarin	7.53	0.33	1.64	0.56
Russian	6.66	0.49	3.98	2.35
Vietnamese	7.13	0.67	2.26	1.85

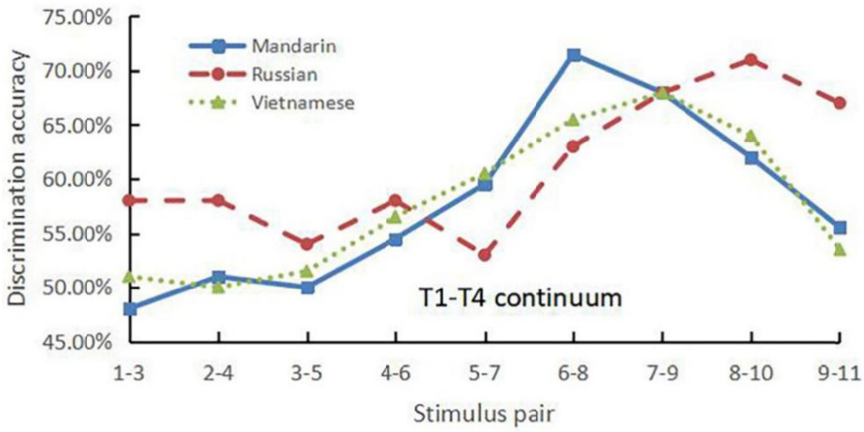

Figure 6. The discrimination accuracy of nine stimulus pairs across the three language groups in the T1-T4 continuum.

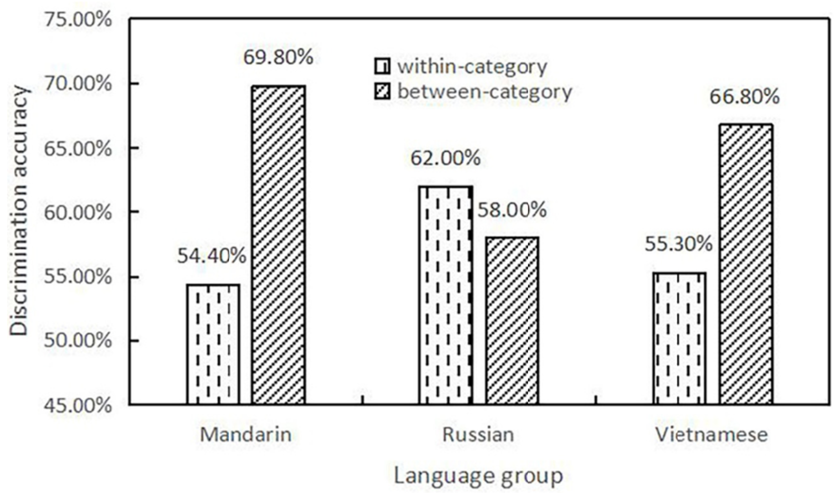

Figure 7. Within-category and between-category discrimination accuracy across the three language groups in the T1-T4 continuum.

5. Discussion

The current study investigated the influence of L1 background on categorical perception of lexical tones by three language groups: native Mandarin listeners, native Russian listeners and native Vietnamese listeners. Results showed that the two tone language groups perceived lexical tones categorically, whereas the non-tone language group perceived both tone continua in a continuous manner. Such findings are in line with previous studies (Halle, Chang, & Best, 2004; Peng et al., 2010; Wang, 1976; Xu, Gandour, & Francis, 2006). Moreover, possibly due to the different tone systems of Mandarin and Vietnamese, these two tone language groups exhibited slightly different identification and discrimination curves. These findings can provide new evidence concerning cross-language differences in categorical tone perception.

In the current study, Russian listeners did not exhibit significantly different boundary positions from native Mandarin listeners for the T1-T2 continuum, which confirms the results reported in previous studies (Halle, Chang, & Best, 2004; Peng et al., 2010; Xu, Gandour, & Francis, 2006). However, their boundary positions for the T1-T4 continuum occurred at a significantly smaller stimulus number than native Mandarin listeners. As shown in Figure 1, the stimuli with smaller numbers have steeper slopes. Hence, the smaller boundary position meant that Russian listeners required a steeper slope to perceive the stimulus as a contour tone. The reason why boundary positions of Mandarin and Russian listeners were the same for the T1-T2 continuum but different for the T1-T4 continuum is possibly due to the impact of intonation in Russian. Similar to Mandarin Tone 2, interrogative sentences in Russian usually have rising pitch contours (Crosby, 2013). Therefore, Russian listeners could identify a rising pitch fairly well, resulting in the same boundary positions for the T1-T2 continuum. Nevertheless, contrary to the sharp pitch falling of Mandarin Tone 4, declarative sentences in Russian often have level or slightly falling pitch contours (Crosby, 2013). Thus, they showed little sensitivity to pitch falling unless there was a sharp drop, leading to smaller boundary positions than native Mandarin listeners for the T1-T4 continuum.

In the identification curves of Vietnamese listeners, they showed significantly smaller boundary positions than the Mandarin group for the T1-T2 continuum, which coincides with the result reported in Wen and Wang (2019), where native Korean listeners showed smaller boundary positions than native Mandarin listeners. They argued that since Korean is a non-tone language, L1 experience might lower their sensitivity to the variations of pitch contours. Therefore, Korean listeners needed a sharper slope to identify the stimulus as a rising tone. However, in the present study, Vietnamese is a tone language. Different boundary positions between the Mandarin and Vietnamese listeners may be attributed to their different tone systems. As shown in Table 1, the ngang tone in Vietnamese and Tone 1 in Mandarin are both level tones, but Mandarin T1 has a higher pitch value than its Vietnamese counterpart. It means that Vietnamese listeners tend to perceive a stimulus with a lower pitch as the level tone, whereas native Mandarin listeners need a higher pitch to do so. It is generally known that due to the restriction of the pronunciation mechanism, it is difficult to pronounce a level tone without any F0 fluctuation. Instead, there probably exists a slight pitch slope for a level tone in natural speech. Thus, from a low F0 onset to a high F0 onset, the Vietnamese group presented the boundary positions earlier than the Mandarin group.

As for boundary width, the Russian group showed broader boundary widths than the Mandarin group both for the T1-T2 continuum and for the T1-T4 continuum, which suggested that Russian listeners did not exhibit sharp boundaries between different tone categories. This finding is highly consistent with several previous studies (Francis, Ciocca, & Ng, 2003; Halle, Chang, & Best, 2004; Peng et al., 2010; Wang, 1976; Xu, Gandour, & Francis, 2006). However, no significant difference in boundary width was found for Mandarin and Vietnamese listeners either in the T1-T2 continuum or in the T1-T4 continuum, which confirms the results reported in Peng et al. (2010), where Mandarin and Cantonese listeners exhibited comparable boundary widths. Deutsch, Henthorn, Marvin, and Xu (2006) found that tone language experience could improve listeners' sensitivity to absolute pitch. By comparing the perception of both Mandarin and Cantonese tones, Lee, Vakoch, and Wurm (1996) found that native Cantonese listeners could discriminate Mandarin tones fairly well, but native Mandarin listeners performed not that well in discriminating Cantonese tones. They ascribed the results to the denser tone system of Cantonese. Correspondingly, Vietnamese has six tones, most of which are contour tones (see Table 1). Moreover, it can be inferred from their pitch values that the pitch variations in Vietnamese are relatively small, which indicates that Vietnamese listeners are required to have strong sensitivity to the variation of pitch height and pitch slope in order to discriminate their complex tone system. Therefore, as non-native speakers, the Vietnamese group can still identify Mandarin tones fairly well.

With respect to discrimination curves, the discrimination peaks of the Russian listeners were located at the stimulus pair 8-10 for both T1-T2 continuum and T1-T4 continuum, which implied that they attached great importance to absolute pitch height when discriminating Mandarin tone contrasts. Thereby, the Russian group exhibited only psychophysical boundaries in discrimination tasks, which supports the previous findings (Halle, Chang, & Best, 2004; Peng et al., 2010; Wang, 1976; Xu, Gandour, & Francis, 2006). For native Vietnamese listeners, their discrimination peak for the T1-T2 continuum was located at the stimulus pair 5-7, which occurred a little bit earlier than that of native Mandarin listeners (at the stimulus pair 6-8). As discussed above, the Vietnamese group had smaller boundary positions for the T1-T2 continuum than the Mandarin group and they would perceive a stimulus as Tone 1 earlier than Mandarin listeners. Correspondingly, it will shape Vietnamese listeners' sensitivity in distinguishing Tone 1 and Tone 2, resulting in an earlier discrimination peak. For the T1-T4 continuum, the discrimination peak of Vietnamese listeners was located at the stimulus pair 7-9, which occurred a little bit later than that of Mandarin listeners (at the stimulus pair 6-8). Such findings can be possibly ascribed to the influence of the two Vietnamese falling tones, i.e., the huyền tone and the nặng tone. Unlike the sharp pitch falling of Mandarin Tone 4, these two Vietnamese tones only fall slightly in terms of their pitch contours. Therefore, just a small degree of pitch falling can help Vietnamese listeners to perceive a stimulus as the falling tone. These two falling tones will shape Vietnamese listeners' sensitivity in distinguishing Tone 1 and Tone 4, leading to the discrimination peak occurring at a larger stimulus pair. In a word, in discrimination tasks, both Mandarin listeners and Vietnamese listeners exhibited linguistic boundaries for both continua even though these linguistic boundaries were further shaped by their respective tone inventory, which was also in accord with previous relevant literature (Peng et al., 2010).

6. Conclusion

The present study has investigated the influence of L1 background on Mandarin tone perception by using the traditional framework of CP. Mandarin and Vietnamese listeners, with exposure to native tone language, could perceive both T1-T2 and T1-T4 continua categorically, whereas Russian listeners, as non-tone language listeners, perceived lexical tones continuously. In identification tasks, compared with those for Russian listeners, the identification curves for Mandarin and Vietnamese listeners in both T1-T2 and T1-T4 continua had steeper slopes at the cross-boundary positions with narrower boundary widths. Moreover, the Mandarin and Russian groups showed significantly different boundary positions for the T1-T4 continuum, and the Mandarin and Vietnamese groups showed significantly different boundary positions for the T1-T2 continuum. In discrimination tasks, Russian listeners exhibited only psychophysical boundaries, whereas native Mandarin and Vietnamese listeners exhibited slightly different linguistic boundaries. All these differences between native Mandarin group and the other two non-native language groups could be attributed to their L1 influence. These findings converge to suggest that native tone language background, to some extent, can facilitate non-native tone perception.

Acknowledgments

The author gratefully thanks Junzhou Ma and Yuxiao Yang for their constructive help in improving the paper.

References

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*(4), 319–325. https://doi.org/10.1177/002383097802100406

Altmann, C. F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., & Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, *64*, 13–23. https://doi.org/10.1016/j.neuropsychologia.2014.09.006

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Computer program, version 6.0.43). Retrieved September 8, 2018, from http://www.praat.org/

Brandt, J., & Rosen, J. J. (1980). Auditory phonemic perception in dyslexia: categorical identification and discrimination of stop consonants. *Brain and Language*, *9*(2), 324–337. https://doi.org/10.1016/0093-934x(80)90152-2

Carne, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*(4), 961–970. https://doi.org/10.1121/1.381590

Chao, Y.-R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674732889

Chen, F., Peng, G., Yan, N., & Wang, L. (2016). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, *44*(6), 1413–1434. https://doi.org/10.1017/S0305000916000581

Crosby, C. F. (2013). *L1 Influence on L2 Intonation in Russian Speakers of English*. MA Thesis, Portland State University, Portland, USA.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. S. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, *119*(2), 719–722. https://doi.org/10.1121/1.2151799

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge, UK: Cambridge University Press.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Francis, A. L., Ciocca, V., & Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029–1044. https://doi.org/10.3758/BF03194832

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41–76). New York, NY: Academic Press. https://doi.org/10.1016/B978-0-12-267350-4.50007-8

Halle, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. https://doi.org/10.1016/j.wocn.2011.11.001

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and non-native listeners. *Phonetica*, *67*(4), 243–267. https://doi.org/10.1159/000327392

Lee, Y. S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542. https://doi.org/10.1007/BF01758181

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication*, *16*(2), 175–205. https://doi.org/10.1016/0167-6393(94)00054-E

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, *61*(5), 1352–1361. https://doi.org/10.1121/1.381409

Rosner, B. S. (1984). Perception of voice-onset-time continua: A signal detection analysis. *Journal of the Acoustical Society of America*, *75*(4), 1231–1242. https://doi.org/10.1121/1.390775

Shen, G. N., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *Journal of the Acoustical Society of America*, *140*(6), 4396–4403. https://doi.org/10.1121/1.4971765

So, C. K. (2005). The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *Journal of the Acoustical Society of America*, *117*(4), 2427–2427. https://doi.org/10.1121/1.4786607

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

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. https://doi.org/10.1017/S0272263114000047

Sun, K. C., & Huang, Ts. (2012). A cross-linguistic study of Taiwanese tone perception by Taiwanese and English listeners. *Journal of East Asian Linguistics*, *21*(3), 305–327. https://doi.org/10.1007/s10831-012-9092-9

Wang, W. S.-Y. (1973). The Chinese language. *Scientific American*, *228*(2), 50–63. https://doi.org/10.1038/scientificamerican0273-50

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, *28*, 61–72. https://doi.org/10.1111/j.1749-6632.1976.tb25472.x

Wang, X. (2006). *Perception of L2 tones: L1 lexical tone experience may not help*. In Proceedings of Speech Prosody, Dresden, Germany.

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, *97*(1), 144–160. https://doi.org/10.1111/j.1540-4781.2013.01386.x

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics*, *24*(1), 113–129. https://doi.org/10.1017/s0142716403000067

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

Wen, B. Y., & Wang, Y. L. (2019). The perception of Mandarin tones by Korean native speakers. *Language*

Teaching and Research, *195*, 31–41 (In Chinese).

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Zheng, H.-Y., Minett, J. W., Peng, G., & Wang, W. S.-Y. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Process*, *27*(2), 184–209. https://doi.org/10.1080/01690965.2010.520493

Zlatin, M. A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, *56*(3), 981–994. https://doi.org/10.1121/1.1903359

Copyrights