5th International Symposium on Tonal Aspects of Languages (TAL 2016)
24-27 May 2016, Buffalo, New York

Tone Pair Similarity and the Perception of Mandarin Tones by Mandarin and English Listeners

Wenyi Ling[1], Amy J. Schafer[2]

[1]Department of Second Language Studies, University of Hawaii, U.S.A.
[2]Department of Linguistics, University of Hawaii, U.S.A.
Wenyi9@hawaii.edu; aschafer@hawaii.edu

Abstract

Previous work [1]–[5] has provided evidence that the perception of lexical tones by native speakers of Mandarin can be more categorical than that of naïve foreign listeners, for the tone pairings T1-T2, T2-T4, T1-T4 and T3-T4. The present study extended this work by testing Mandarin and naïve English listeners' perception of all six possible pairwise combinations of the four Mandarin tones, using identification and AXB discrimination tasks. The results revealed significant differences in perception across tone pairs, highlighting potential challenges for language learning. They also confirmed that Mandarin listeners were more categorical in their perception than the naïve listeners, who showed higher discrimination accuracy than Mandarin listeners. The latter finding contrasts with previous results for naïve French speakers' perception of Mandarin tone [2], and is consistent with an effect of native-language supra-segmental patterns on foreign-language tone perception.
Index Terms: tone, perception, Mandarin, discrimination

1. Introduction

1.1. Tone perception by native listeners

Mandarin Chinese has four lexically distinctive tones: high level (T1), high rising (T2), low dipping (T3), and high falling (T4) [6,7]. Research going back to Wang [1], who conducted an identification and an AXB discrimination task on a continuum from T1 to T2 with Mandarin listeners, has found evidence for categorical perception (CP) by native listeners. Subsequent studies tested continua for the T1-T2 pair [2]–[5] and three other tone pairs: T2-T4 [2,3], T1-T4 [5] and T3-T4 [2], and similarly concluded that Mandarin tones are categorically perceived by native listeners, although identification curves can have shallower slopes and discrimination curves can have broader peaks compared to tests of consonants [2,8].

The present study used identification and discrimination tasks to test tone continua for all six tone pairings in Mandarin, to compare the degree of CP across tones in a single study, and provide a better picture of CP for T1-T3 and T2-T3 pairs. Both to set the stage for future work on native English second-language learners of Mandarin, and to provide a comparison group for the native Mandarin listeners, we also tested naïve English listeners – i.e. native English listeners with no prior experience with Mandarin.

1.2. Tone perception by naïve listeners

Naïve listeners have been found to exhibit less categorical and more psychophysical perception of lexical tone [2,5]. Hallé et al. [2] tested Taiwan Mandarin listeners and French listeners on three tone pairs (T1-T2, T2-T4, T3-T4) in identification and AXB discrimination tasks. Because the naïve listeners could not readily use category labels (e.g., Chinese characters) for identification, the identification task was a modified AXB task in which steps in a tone continuum (the X stimuli) were flanked by endpoint stimuli. For Mandarin listeners, Hallé et al. found steep slopes in the identification curves and corresponding peaks in the discrimination curves, indicating CP. For naïve French listeners, the identification curves showed notable slopes, but nevertheless less categorical performance than the native listeners, while the discrimination curves lacked distinctive peaks and were significantly lower in accuracy compared to native Mandarin speakers.

The current study used methods similar to Hallé et al., although with several important changes. While Hallé et al used Chinese characters as labels in the identification task for native speakers and tested naïve speakers in a modified AXB task, we used tone number labels (e.g. T1, T2) for both groups [9]. These were introduced to participants during exposure trials, described below. Use of tone labels allowed us to use the same task for each group, and avoided the attention to certain tone features that could occur with descriptive labels such as "rising" or "dipping". As for participants, our native listener population spoke mainland Mandarin, which is more widely used than Taiwan Mandarin and regarded as the standard version. Our naïve population consisted of American English listeners. The English and French suprasegmental systems differ in a number of respects, and it is known that the suprasegmental contrasts in a listener's native language can influence their tonal perception in an unfamiliar language [9,10,11,12]. Thus, our naïve English listeners might be expected to differ from Hallé et al's naïve French listeners in their responses to Mandarin tone continua, providing further insight into how previous language experience and native-language knowledge affect tone perception. As noted above, the naïve English group also provided a comparison group for in-progess work on Mandarin tone perception by native English listeners who are learning Mandarin as a second language.

2. Methods

2.1. Participants

Thirty native Mandarin listeners from the East China Normal University (China) community and thirty English listeners with no experience of Mandarin or any other tonal language from the University of Hawaii (U.S.) community participated in the study. Each completed both experimental tasks. An additional 4 listeners were tested but eliminated from analysis for excessive errors (3) or misunderstanding the

10.21437/TAL.2016-19

instructions (1). The age range (18-50 yrs) was similar for the two groups. No participants reported any professional music experience or any hearing or vision problems. Participants were rewarded for their time with a small amount of course credit or $10.

2.2. Stimuli

The two syllables /pa/ and /pi/ were used for the experimental materials, and /kwo/ was used for practice stimuli. Both experimental syllables combine with the four Mandarin tones to create common words. Because word frequency could potentially influence the results [13,14], but there is disagreement about the total number of homophones for each tonal syllable in Mandarin [13] the log frequency of the most common homophone for each tonal syllable was obtained, following [2]. A Chi-square analysis showed no significant difference across the tone types ($X^2 = 0.05$, $p = 1$). After data collection, we discovered technical errors with some of the /pa/ stimuli. In order to report a balanced data set, only the results from /pi/ are presented here.

To avoid effects of co-articulation and tone sandhi, the sound stimuli were constructed from isolated Mandarin words spoken by a female native speaker. Two native speakers blind to the research question selected tokens for the experiment based on intensity and sound quality. Eight selected sounds (2 syllables * 4 tones) were used to construct the six tone pairs. Six 9-step tone continua were generated by equalizing duration within the continuum and linearly changing the pitch and intensity between the endpoints (see Figure 1) with the PSOLA method [15] in Praat [16]. Six 4-step tone continua for /kwo/ were synthesized in the same manner for practice stimuli. Two native Mandarin speakers confirmed the naturalness of all stimuli.

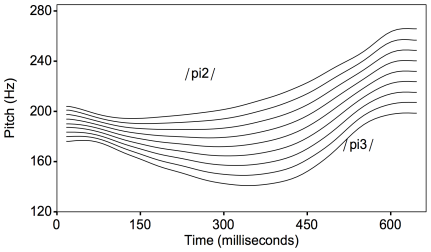

Figure 1. *Examples of synthesized continua of T2-T3.*

2.3. General procedure and tasks

2.3.1. General procedure

First, all participants completed an online questionnaire about their personal language background and professional music experience. Half of the participants in each group performed the identification task first; the other half started with the discrimination task. Participants took a break between the tasks. After finishing both tasks, participants were asked about their difficulty. Most Mandarin listeners (29/30) thought the identification task was easier, while all English listeners thought discrimination was easier.

2.3.2. Identification task

Stimuli were blocked by tone pair to reduce the English listeners' memory load and avoid alternative answers (e.g. the middle step of the T2-T4 pair has a flat pitch, which is similar to T1). Block order was counterbalanced across participants. In each block, all nine steps for the tone pairs for each syllable were presented three times in random order. This produced a total of 324 testing trials (2 syllables * 6 tone pairs * 9 steps * 3 presentations). Because the naïve listeners were unfamiliar with Mandarin tones and could not readily use category labels (e.g., Chinese characters), step 0 and step 8 tokens for the relevant tone pairs were presented at the beginning of each block, along with tone number labels (e.g. T1, T2). For each tone pair, a block consisted of two exposure trials with the tone-labelled practice syllable, four practice trials, four exposure trials with tone-labelled critical syllables, and then the critical trials. Participants were required to classify each sound token as quickly as possible by pressing the appropriate tone-number key on the keyboard (e.g, "1" or "2" in block T1-T2). Participants initiated the next trial or block by pressing the space bar, at their own pace. The task took about 15 minutes to finish.

2.3.3. Discrimination task

Discrimination was tested with an AXB task using steps two intervals apart (e.g. step 1 vs. 3). On each trial, participants were asked to press a key to indicate whether token X matched A or B. Stimuli were blocked by tone pair to create six testing blocks. There was a total of 336 trials (2 syllables * 6 tone pairs * 7 step pairs * 4 AXB combinations). Each block began with exposure trials (identical to those in the identification task) and 6 AXB practice trials, followed by the critical stimuli presented in random order. The inter-stimuli interval (ISI) was set to 500 ms, following common practice [4,8] and data suggesting that this value produces a high correlation between discrimination and indentification performance [8,17]. However, in order to keep the experiment to a reasonable duration, it was shorter than Hallé et al.'s ISI of 1 sec. Participants were required to hear the complete AXB set in a trial before responding. This task was also self-paced and took about 35 minutes.

3. Results

3.1. Identification

For the identification task, the responses for Sound A versus B were fit by mixed-effects logistic regression models for each tone pair within each group [4,18], including step as a continuous fixed effect and participants' intercepts and slopes for step as random effects. This produced group-wise models of the identification curve for each tone pair for each group and model-based intercepts and slopes (in log odds) for each tone pair for each participant. Table 2 shows the mean by-participant intercepts and slopes by group and tone pair. The position in the step continuum of a potential categorical boundary (CB) was calculated for each participant as the point where proportions of responses to the two choices were predicted by the model to be equal [4]. Figure 2 shows the identification curves for each tone pair by group. Mandarin listeners' identification curves were sigmoid in shape, with steeper slopes than those of naïve English listeners. English listeners showed semi-horizontal identification curves for T2-T3, T2-T4, and T3-T4 pairs, indicating categorization difficulty, as well as near-chance performance on the T4 end of the T1-T4 continuum, suggesting difficulty in identifying any token within a few steps from a T4 endpoint.

Table 2. *Identification: Mean model-based intercepts and slopes (in log odds), and mean predicted category boundary locations (CBs) in the step continuum for Mandarin and English listeners.*

Tone pair	Intercept	Slope	CB
Mandarin listeners (N = 30)			
T1-T2	9.00	-2.54	3.75
T1-T3	7.12	-2.70	2.78
T1-T4	3.59	-1.75	2.19
T2-T3	8.68	-1.85	4.63
T2-T4	8.72	-2.44	3.69
T3-T4	9.82	-2.77	3.58
English listeners (N = 30)			
T1-T2	2.99	-0.75	4.62
T1-T3	4.13	-1.23	2.49
T1-T4	2.11	-0.43	6.05
T2-T3	1.72	-0.38	7.22
T2-T4	1.00	-0.24	7.40
T3-T4	1.65	-0.35	4.92

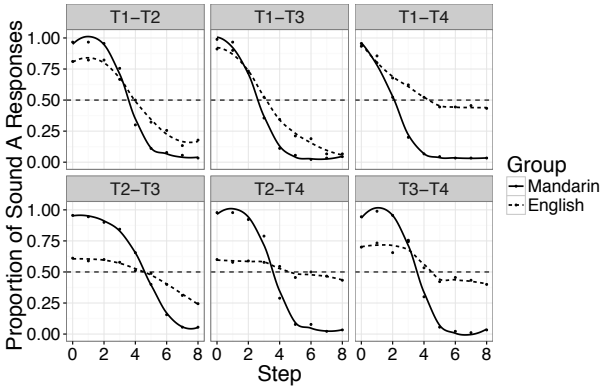

Figure 2. *Identification curves pooled across participants for each tone pair. The solid lines represent the proportion of Sound A responses at each step (0-8) for Mandarin native listeners; the dashed lines for English listeners.*

In order to analyze the degree of CP in the identification results across listener groups and tone pairs, we fit a new mixed-effects linear regression model using as a dependent measure the slope values previously calculated for each participant within each tone pair [4]. Group, tone pair, and their interaction were included as fixed effects (centered), and participants' intercepts and tone pair slopes were added as random effects. The results showed significant effects of group ($b = -1.78$, $t = -12.58$, $p < 0.01$) and tone pair ($b = 0.33$, $t = 2.22$, $p < 0.05$), and their interaction ($b = -0.79$, $t = -2.66$, $p < 0.01$), which indicates native listeners had steeper slopes than naïve English listeners, that slope varied across tone pairs, and that the changes in slope across tone pairs was significantly different for the two groups. Overall, the identification data revealed that Mandarin listeners perceived tones more categorically than naïve English listeners and, within naïve listeners, pairs T1-T2 and T1-T3 showed the most categorical behavior.

3.2. Discrimination

Discriminary accuracy was analyzed with mixed effects logistic regression models using discrimination accuracy as the dependent variable, and group, tone pair, step pair and their interactions as fixed effects. Participant intercepts and slopes for tone pair and step pair were included as random effects. All of the fixed effects were centered, so that the model output would represent main effects of each factor and their interactions. Because discrimination accuracy should relate to the steepness of the curve found with identification, tone pairs were arranged by average steepness of identification slope from steepest to shallowest across the two groups. Then tone pair and step pair were treated as continuous numerical factors. To assess the possibility of non-linear accurary curves across step pairs, an additional model was fit that added a quadratic term for step pair. Model comparisons using ANOVA found no improvement with the quadratic term, and so we report results from the simpler linear model.

The model results (see Table 3) showed a significant group effect, significant tone pair and step pair effects, and a significant interaction of tone pair and step pair. None of the remaining interactions were significant. Figure 3 displays the mean proportion of accurate responses for the discrimination task, averaged across step pairs, by tone pair and group. Accuracy was higher overall for English (92%) than Mandarin listeners (83%). This difference was due to Mandarin listeners' lower accuracy to within-category step pairs, which can be seen in Figure 4, where the data are further divided by step pair.

Table 3. Parameter estimates of discrimination accuracy

Parameter estimates			
	Est.	z	p <
Intercept	2.25	21.27	.001
Group	0.92	4.41	.001
Tone pair	-0.53	-4.59	.001
Step pair	-0.41	-3.38	.001
Grp*TP	-0.21	-1.06	.29
Grp*Stp	0.31	1.49	.14
TP*Stp	1.04	3.55	.001
Grp*TP*Stp	-0.55	-0.96	.34

M1: glmer(Accuracy~Grp*TP*Stp+(1+TP+Stp|Participant))

Since discrimination peaks indicate CP, the group effect in the overall model suggests a potential difference between the two groups in the degree of categorical perception. To examine this further, we ran separate mixed-effects models on each group. The effect of tone pair was significant for each group (native: $b = -0.46$, $z = -3.37$, $p < 0.01$; naïve: $b = -0.56$, $z = -2.81$, $p < 0.01$), illustrated by the variable discrimination accuracy across tone pairs for each group shown in Figure 3. For both groups, T1-T4 and T2-T3 produced the lowest overall accuracy (native: 75% & 78%; naïve: 88% & 87%), while T1-T3 showed the highest (native: 89%; naïve: 96%). Mandarin listeners showed a significant effect of step pair ($b = -0.52$, $z = -3.67$, $p < 0.01$), but English listeners did not ($b = -0.35$, $z = -1.68$, $p = 0.09$) consistent with the relatively flat discrimination curves of naïve English listeners shown in Figure 4. The interaction of tone pair and step pair was significant with native listeners ($b = 1.31$, $z = 3.76$, $p < 0.01$), but not with naïve listeners ($b = 0.71$, $z = 1.49$, $p = 0.14$). Naïve English listeners generally showed high discrimination accuracy across step pairs for each tone pair, while the accuracy of native listeners varied more across tone and step pair combinations.

Finally, we can compare the CBs generated by logistic regression on the identification data (Table 2) with the discrimination curves across step pairs (Figure 4). For native Mandarin listeners, all of the tone pairs showed some degree of correspondence between the predicted CB and step pairs with high discrimination accuracy. However, for naïve English listeners, no obvious peaks were observed at

plausible category boundaries. Overall, the discrimination data showed evidence of categorical perception by Mandarin listeners, but not by naïve English listeners.

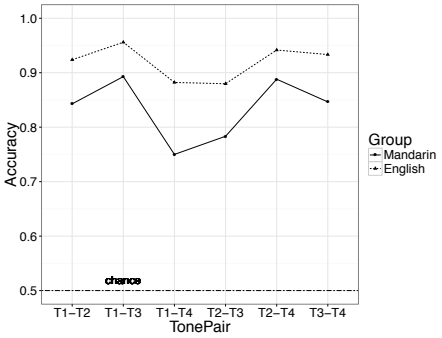

Figure 3. *Discrimination scores pooled across participants by group and tone pair (y-axis: accuracy for the 0.5-1 range; x-axis: tone pair). The solid line represents the identification performance of Mandarin listeners; the dashed line of English listeners.*

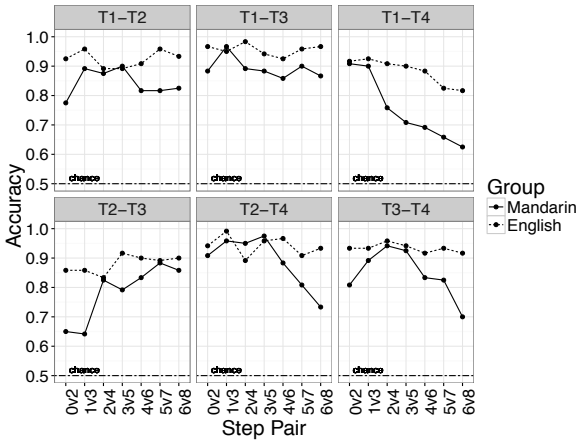

Figure 4. *Discrimination curves pooled across participants for each tone pair (y-axis: accuracy for the 0.5-1 range; x-axis: step pair). The solid lines represent the discrimination performance of Mandarin native listeners; the dashed lines of English listeners*

4. General discussion

The results of our identification and discrimination tasks indicated that native Mandarin listeners perceived the tone pairs categorically, while naïve English listeners perceived Mandarin tones non-categorically. This pattern replicated the native-naïve distinction of [2] with native listeners of another variety of Mandarin (mainland versus Taiwan) and naïve listeners from a different native language background (English versus French).

Mandarin listeners' identification curves for each tone pair showed the sigmoid shapes and steep slopes typical of CP. Their discrimination accuracy varied across step pairs, as would be expected for CP. And, by comparing the locations of CBs predicted by the identification results with their discrimination curves, we also found performance broadly consistent with CP in this group, reinforcing the results of previous studies [1]–[5].

The naïve English listeners showed sigmoid shapes for T1-T2 and T1-T3, but non-categorical identification for the remaining pairs, i.e., for any tone pair involving T4 and for the T2-T3 pair. T4 was previously found to be difficult for naïve English listeners to categorize in research that used unaltered Mandarin tones, in a task in which all four tone categories were available as response options [9]. In addition, [11] found relatively low accuracy for T1-T4 categorization, but not that of T2-T4 or T3-T4, in a task using natural tones within carrier sentences (non-citation forms). Common across this research is relatively good categorization by naïve English listeners for T1-T2 and T1-T3, and weaker categorization for T1-T4 and T2-T3.

The tone pairs that had not previously been tested with naïve listeners and Mandarin tone continua were T1-T3 and T2-T3. T2 and T3 had very similar fall-rise contours in the citation forms used for the present stimuli (see Figure 1), and, as mentioned, were difficult for naïve English listeners to discriminate, echoing other research [e.g. 9,11,12,19,20]. T1 and T3 are quite distinct acoustically: the two tones differ in contour (level versus falling-rising) and in height (high versus low). Both groups of listeners showed steep slopes for the T1-T3 pair in identification, exhibiting a typical CP identification curve, and very high discrimination accuracy across all step pairs. This suggests that both groups of listeners were able to draw on salient acoustic differences for this tone pair (in addition to knowledge of tonal categories, for the native speakers). Prior research has shown that tone height is a salient cue for native English as well as native Mandarin listeners [12]. This pair had the greatest difference in average F0 of all tone pairs, accounting for its near-ceiling discrimination performance in each group. Since acoustic memory traces decay after about 250-500 ms [21], tests with a longer ISI or a more difficult task might reveal greater CP for T1-T3 within the native Mandarin group.

Our results differed from Hallé et al. [2] in that discrimination accuracy of our native group (83%) was lower than that of our naïve English listener group (92%), versus Hallé et al.'s (88%) for native Mandarin and (74%) for naïve French listeners. There are several differences between the two studies that may have led to this reversal. First, the ISI in Hallé et al. was longer (1 sec vs. 500 ms), which could have led to greater memory trace decay in their study, especially for naïve listeners who did not have long-term representations or categories to draw on. The accuracy of the native groups was comparable between the two studies, while the accuracy of the two naïve groups was dissimilar, supporting this explanation. Second, suprasegmental differences between English and French could have aided the naïve English listeners in their Mandarin tone discrimination [e.g., 5,9,10,11,22,23]. Previous studies have demonstrated an influence of native-language suprasegmental experience on non-native listeners' perception of lexical tones [e.g., 5,9,10,12,22,23], and [11] found an advantage for naïve French listeners over English ones in their test of natural Mandarin tones in carrier sentences. Future work that tests naïve English and French listeners on the same continua in a single study could clarify the relative effects of ISI (and other task differences) versus native language background.

In summary, we found that native listeners perceived Mandarin tones more categorically than naïve listeners, that both groups varied in CP patterns across tone pairs, and that naïve English listeners displayed very high discrimination ability. These results suggest promising avenues for investigation of the developmental path for tone use in adult second language learners of Mandarin (L2ers). Continuing research in our laboratory is extending these tasks to native English adult learners of Mandarin.

5. References

[1] W. S.-Y. Wang, "Language Change," Ann. N. Y. Acad. Sci., vol. 280, no. 1 Origins and E, pp. 61–72, 1976.

[2] P. A. Hallé, Y.-C. Chang, and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," J. Phon., vol. 32, no. 3, pp. 395–421, 2004.

[3] J. Xi, L. Zhang, H. Shu, Y. Zhang, and P. Li, "Categorical perception of lexical tones in Chinese revealed by mismatch negativity," Neuroscience, vol. 170, no. 1, pp. 223–231, 2010.

[4] Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am., vol. 120, no. 2, pp. 1063-1074, 2006.

[5] G. Peng, H.-Y. Zheng, T. Gong, R.-X. Yang, J.-P. Kong, and W. S.-Y. Wang, "The influence of language experience on categorical perception of pitch contours," J. Phon., vol. 38, no. 4, pp. 616–624, 2010.

[6] Y. R. Chao, Mandarin primer: An intensive course in spoken Chinese. Harvard University Press, 1948.

[7] M. Yip, Tone. Cambridge University Press, 2002.

[8] A. J. van Hessen and M. E. Schouten, "Modeling phoneme perception. II: A model of stop consonant discrimination," J. Acoust. Soc. Am., vol. 92, no. 4, pp. 1856–1868, 1992.

[9] C. K. So and C. T. Best, "Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences," Lang. Speech, vol. 53, no. 2, pp. 273–293, May 2010.

[10] C. K. So and C. T. Best, "Do English speakers assimilate Mandarin tones to English prosodic categories?," in Interspeech 2008: Proceedings of the 9th Annual Conference of the International Speech Communication Association, pp. 22–26, 2008.

[11] C. K. So and C. T. Best, "Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories," Stud. Second Lang. Acquis., vol. 36, no. 2, pp. 195–221, 2014.

[12] J. Gandour, "Tone perception in far eastern-languages," J. Phon., vol. 11, no. 2, pp. 149–175, 1983.

[13] K. Zidian, "Kangxi Dictionary," by Zhang Yushu. Shanghai Shanghai Shudian Chubanshe, 1985.

[14] Y. Huang, "Frequency of Mandarin words based on Google@," [Data file]. Retrieved from http://yong321.freeshell.org/misc/ChineseCharFrequencyG.txt, 2005.

[15] E. Moulines and J. Laroche, "Non-parametric techniques for pitch-scale and time-scale modification of speech," Speech Commun., vol. 16, no. 94, pp. 175–205, 1995.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 5.3.84, retrieved 27 August 2014 from http://www.praat.org/, 2014.

[17] J. F. Werker and J. S. Logan, "Cross language evidence for three factors in speech perception," Percept. Psychophys., vol. 37, no. 1, pp. 35–44, 1985.

[18] T. F. Jaeger, "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," J. Mem. Lang., vol. 59, no. 4, pp. 434–446, 2008.

[19] T. Huang, "Tone perception by speakers of Mandarin Chinese and American English," Ohio State Univ. Work. Pap. Linguist., vol. 55, 2001.

[20] T. Huang, "Language-specificity in auditory perception of Chinese tones," Dr. Diss. Ohio State Univ., vol. 1, 2004.

[21] D. B. Pisoni, "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys., vol. 13, no. 2, pp. 253–260, 1973.

[22] B. Yang, "A model of Mandarin tone categories--a study of perception and production." Dr. Diss. The University of Iowa, 2010.

[23] Y. Wang, A. Jongman, and J. A. Sereno, "L2 acquisition and processing of Mandarin tone," Handb. Chinese Psycholinguist., pp. 250–257, 2006.