Interspeech 2018
2-6 September 2018, Hyderabad

Cross-language perception of Mandarin lexical tones by Mongolian-speaking bilinguals in the Inner Mongolia Autonomous Region, China

Kimiko Tsukada[1, 2], *Yurong*[3]

[1]Macquarie University, Australia
[2]University of Oregon, USA
[3]Inner Mongolia University, China
Kimiko.tsukada@gmail.com, umyurong@yahoo.co.jp

Abstract

Mandarin is a representative tonal language with four contrastive tone categories (Tone 1 (T1): high level (ā), Tone 2 (T2): high rising (á), Tone 3 (T3): dipping (ǎ), Tone 4 (T4): high falling (à)). Learning Mandarin tones is known to be difficult for speakers from diverse linguistic backgrounds. The purpose of this research was to examine how native Mongolian-speaking bilinguals perceive Mandarin lexical tones. The 24 (17 females, 7 males) participants studied Mandarin for 15 years on average in the Inner Mongolia Autonomous Region, China. A discrimination experiment was conducted to assess Mongolian bilinguals' perception of six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4). The Mongolian group was less accurate than the control group of ten native Mandarin listeners for all six pairs and the between-group difference was particularly large for T2-T3. However, large individual variation was observed and some Mongolian bilinguals perceived Mandarin tones as accurately as native Mandarin listeners, suggesting that native-like tone perception is attainable in subsequently acquired languages.

Index Terms: perception, Mandarin, Mongolian, lexical tones, bilingualism

1. Introduction

The number of individuals who learn Mandarin as a second (L2) or foreign language is growing in many parts of the world. There is also a large number of heritage learners of Mandarin in countries such as Australia and United States that have traditionally been the destinations for many migrants. There is the widely-held view that learning Mandarin tones is difficult especially for speakers from non-tonal language backgrounds [1, 2]. Accordingly, there is substantial interest among speech researchers and language teachers in gaining better understanding of the processing and/or acquisition of Mandarin lexical tones.

Within China itself, ethnic minorities including Mongols speak Mandarin in addition to their diverse first languages (L1s) although little is known about characteristics of Mandarin spoken by these people. In fact, difficulties of bilingual education in minority schools has been pointed out in previous research. According to [3], "more and more Mongolian children are opting to attend Han Chinese schools" and "in the Autonomous Regions, greater emphasis is placed upon learning Chinese, which functions as a common language between ethnic minorities, than on 'foreign' languages". From the perspectives of social advancement and economic development, "the advantages for ethnic minority populations to learn Chinese are undeniable [3]", although the conservation of ethnic identity may be challenged.

The aims of this research were to examine how native speakers of Mongolian who were born and raised in the Inner Mongolia Autonomous Region in China perceive Mandarin lexical tones and compare their pattern of tone perception to that of native Mandarin speakers. This study provides rare insights into how ethnic minorities living in China perceive Mandarin tones and has implications for bilingual education as well as cross-linguistic perception of prosodic characteristics.

Mandarin is a tonal language with four tone categories (Tone 1 (T1): high level (ā), Tone 2 (T2): high rising (á), Tone 3 (T3): dipping (ǎ), Tone 4 (T4): high falling (à)). In tonal languages, incorrect use of lexical tones leads to confusion/misunderstanding (e.g. 妈 mā 'mother' vs 马 mǎ 'horse' or 买 mǎi 'buy' vs 卖 mài 'sell'). Unlike Mandarin, Mongolian, which is spoken by more than 3 million people, is a quantity language with 12 short and 12 long vowels [4]. Thus, Mongolian has a substantially larger vowel inventory than does Mandarin [5], but it is not a tonal language. While there is agreement among phoneticians that "Mongolian stress (interpreted here as accent) has no distinctive function (no lexical or morphological function) [5, p. 63]", its exact nature (e.g. location, variable vs fixed, pitch vs stress accent) seems to be unresolved, as there is not yet sufficient research on prosodic characteristics of Mongolian [6, 7].

We are interested in determining if and to what extent Mongolian bilinguals approximate to or diverge from the native Mandarin speakers in their cross-language tone perception. Due to limited knowledge of prosodic characteristics of Mongolian as briefly mentioned above, it is difficult to predict the Mongolian bilinguals' perceptual assimilation patterns and discrimination accuracy of Mandarin tones by applying current theories of cross-language perception such as Perceptual Assimilation Model for Supersegmentals [8, 9]. Nevertheless, it might be expected that having no existing tone categories which can be assimilated to Mandarin tones may make perceptual learning difficult. Alternatively, having no tone categories to "overwrite" may facilitate learning especially at young age, as this may set non-native learners free from L1-specific phonetic and/or phonological constraints. Despite lack of lexical tones in their L1 Mongolian, Mandarin language education starting at primary level and beyond may enable the Mongolian bilinguals to successfully acquire Mandarin lexical tones. The Mongolian bilinguals who participated in this study all started learning Mandarin at the age of six.

10.21437/Interspeech.2018-48

2. Methods

A perception experiment was conducted to assess Mongolian bilinguals' discrimination accuracy of six Mandarin tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4). The stimuli were produced by multiple native speakers of Mandarin as described below.

2.1. Speakers and speech materials

Eight (4 males, 4 females, *mean* = 27.8 years, *sd* = 9.2) native Mandarin speakers were recruited from the undergraduate student population at a university in Sydney. Their mean length of residence in Sydney was 1.6 years. While some of them spoke regional dialects in addition to Mandarin, they all received primary and secondary education in standard Mandarin (Putonghua) prior to arriving in Australia as young adults and identified themselves as native speakers of Mandarin. The participants were recorded in a sound-treated studio on university campus under the supervision of a Mandarin-English bilingual experimenter and received monetary reward for their participation.

A total of 76 monosyllabic words including the 28 test words (Table 1) were presented on the computer screen one word at a time in random order and produced twice in isolation and once in a short carrier sentence (我读______这个字 wǒ dú ___ zhè ge zì "I read the word ___").

Table 1: *Test words used in this study (V = Vowel).*

V	Tone 1	Tone 2	Tone 3	Tone 4
/i/	眯 mī "blind"	迷 mí "lost, confused"	米 mǐ "rice"	密 mì "secret"
	逼 bī "narrow"	鼻 bí "nose"	笔 bǐ "to write"	必 bì "must"
	低 dī "low"	敌 dí "to fight"	底 dǐ "foundation"	弟 dì "younger brother"
/a/	妈 mā "mother"	麻 má "hemp"	马 mǎ "horse"	骂 mà "to scold"
	八 bā "eight"	拔 bá "to extract"	把 bǎ "to hold"	爸 bà "father"
	答 dā "answer"	达 dá "to extend"	打 dǎ "to beat"	大 dà "big"
/u/	都 dū "capital city"	读 dú "to read"	赌 dǔ "to gamble"	度 dù "occasion"

Tokens produced in isolation were used as stimuli in this study. The stimuli presented were seven CV syllables (where C=/m, p, t/ and V=/i, a, u/) across all four Mandarin tones[1]. All materials were transcribed in Chinese characters with pinyin (the Romanized spelling system of Chinese characters with tones indicated by diacritics) on top to minimize any ambiguity of pronunciation. The pace of presentation was controlled by the experimenter. The recorded speech materials were digitized at 44.1 kHz and the target words were segmented and stored in separate files to be presented as stimuli.

2.2. Participants

The 24[2] (17 females, 7 males, *mean* = 24.4 years, *sd* = 3.9) Mongolian-speaking participants were born and brought up in the Inner Mongolia Autonomous Region, China. All of them identified themselves as native speakers of Mongolian and participated in the study at university campus in Hohhot. They started learning Mandarin at the age of six in primary school. In Inner Mongolia, students learn Mandarin weekly for approximately 8-10 hours, 6-8 hours, 4-6 hours at primary, junior high and senior high schools, respectively. As for teachers' language background, the majority are native speakers of Mongolian at primary and junior high schools and native speakers of Mandarin at senior high schools. The Mongolian bilinguals were undergraduate or postgraduate students at Inner Mongolia University in Hohhot and studied Mandarin for 15 years on average at the time of testing. In general, the Mongolian bilinguals use more Mongolian than Mandarin in their daily life at home and university, but use Mandarin at places such as university canteens or when they go shopping.

A group of ten (8 females, 2 males, *mean* = 25.4 years, *sd* = 4.3) college-educated native Mandarin speakers participated as controls in our previous research [10, 11, 12]. None of them participated in the recording sessions. The Mandarin listeners participated in the study in Australia (*n* = 5), Japan (*n* = 1) or Singapore (*n* = 4) according to their place of residence and availability. The Mandarin listeners lived in these countries temporarily (for 0.5 months to 5 years) and identified themselves as native speakers of Mandarin (Putonghua).

2.3. Procedures

This study applied the method for testing L2 segmental perception to lexical tone perception. A categorial discrimination test (CDT) with a four-alternative forced-choice oddity task employed in previous L2 speech research [13, 14, 15, 16] was used to assess Mongolian speakers' perception of six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4). As described in [15, p. 118], this is 'a version of the ABX discrimination task' and 'is designed to minimize response bias (guessing)'. A high level of performance in this task would require not only the use of purely auditory information but also the establishment of phonetic categories for one or both sounds in a given sound (tone) pair. The listeners were tested individually in a session lasting approximately 45 to 60 minutes. The presentation of the stimuli and the collection of perception data were controlled by the UAB (University of Alabama at Birmingham) software [17] for the Mandarin group and the PRAAT program [18] for the Mongolian group. The listeners heard the stimuli at a self-selected, comfortable amplitude level over the high-quality headphones on a notebook computer. The experimental session was self-paced and the listeners could take a break after each block if they wished.

[1] While we have not yet examined the influence of individual segments, contextual vowels and/or consonants may also affect listeners' responses. We thank an anonymous reviewer for reminding us of this possibility.

[2] One Mongolian listener's data were excluded from the analysis, as her scores were lower than 0.5 for 5 of the 6 tone pairs, which suggests that she may not have fully understood the instructions for the task.

The stimuli which consisted of monosyllabic CV words differing in lexical tones were presented in triads and the listeners were given four (‘1’, ‘2’, ‘3’, ‘NO’) response categories. Each of the six pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4) was tested by change and no-change (catch) trials. The three tokens in all trials were spoken by three different talkers, and so were always physically different even in no-change trials, as this was considered a better measure of listeners’ perceptual capabilities in real world situations [19].

The listeners were asked to choose an odd “word” that was different from the other two, if there was any. The change trials contained an odd item. For example, a change trial testing the T1-T2 pair might consist of ‘mā$_2$’-‘mā$_1$’-‘má$_3$’ (where the subscripts indicate different talkers). The correct response for change trials was the button (‘1’, ‘2’, or ‘3’) indicating the position of the odd item, which occurred with equal frequency in all three possible serial positions. The serial position of the odd item in a triadic change trial was not fixed, which increased task uncertainty. The change trials tested the participants’ ability to respond appropriately to relevant phonetic differences between tokens and distinguish tones drawn from two different categories.

The correct response to no-change trials, which contained three different instances of a single category (e.g. /tǐ/$_3$ /tǐ/$_1$ /tǐ/$_2$ or /pà/$_1$ /pà/$_3$ /pà/$_2$), was a fourth button marked ‘NO’. The no-change trials tested the participants’ ability to ignore audible but phonetically irrelevant within-category variation (in e.g. voice quality). The participants were required to respond to each trial, and were told to guess if uncertain. A trial could be replayed as many times as the listener wished, but responses could not be changed once given. The inter-stimulus interval in all trials was 0.5 s.

A total of 360 trials were presented in three blocks of 120 trials. A different randomization was used for each block. The first eight trials in each block were for practice and were not analyzed. The resulting 336 (3 blocks x 112) trials consisted of 252 change trials testing six pairs (42 trials each for T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4) and 84 no-change trials (21 trials each for T1, T2, T3, T4). In selecting the stimuli, care was taken so that tokens by each of the eight speakers would be distributed as evenly as possible.

Responses to the change and no-change trials were used to calculate A-prime (A') scores [20], an index of discrimination accuracy. These scores were based on the proportion of ‘hits (Hs)’ obtained for each tone pair and the proportion of ‘false alarms (FAs)’. If the proportion of Hs equaled the proportion of FAs, then A' was set to 0.5. If H exceeded FA, then $A' = 0.5+((H–FA)*(1+H–FA))/((4*H)*(1–FA))$. However, if FA exceeded H, then $A' = 0.5–((FA–H)*(1+FA–H))/((4*FA)*(1–H))$. An A' score of 1 indicated perfect sensitivity, whereas an A' score of 0.5 or lower indicated a lack of sensitivity.

3. Results

Averaged across six tone pairs, the mean discrimination scores were 0.98 and 0.90 for the Mandarin and Mongolian groups, respectively. The mean discrimination scores for each tone pair for each group are given in Table 2.

The Mongolian group was less accurate than the control group of ten Mandarin speakers for all six pairs. T2-T3 was the hardest pair for the Mongolian bilinguals (0.72). T2-T3 confusion has been frequently reported in the literature for listeners from diverse L1 backgrounds [1, 2, 8, 9, 10, 11, 12, 21, 22, 23] and even for advanced learners [12, 22]. This may be due to acoustic-phonetic similarity of T2 and T3, as both tones have an initial dip in pitch followed by a rising pitch contour when spoken in isolation. The native Mandarin listeners' discrimination accuracy was consistently high for all pairs and they showed little inter-speaker variability.

Table 2: *Mean discrimination (A') scores by two (Mandarin, Mongolian) groups of listeners. Standard deviations are in parentheses.*

Tone pairs	Mandarin (n = 10)	Mongolian (n = 24)
T1-T2	0.99 (0.009)	0.91 (0.081)
T1-T3	0.98 (0.009)	0.93 (0.060)
T1-T4	0.98 (0.010)	0.94 (0.063)
T2-T3	0.97 (0.014)	0.72 (0.225)
T2-T4	0.99 (0.011)	0.95 (0.039)
T3-T4	0.98 (0.013)	0.95 (0.039)

Figure 1 shows the distribution of A' scores by two groups of listeners as a function of tone pairs. The between-group difference in the mean A' scores was largest for T2-T3 (0.25) and smallest for T2-T4 and T3-T4 (0.032). However, large individual variation was observed and some Mongolian listeners perceived Mandarin tones as accurately as native Mandarin listeners. The number of Mongolian listeners whose discrimination scores fell within the range set by the Mandarin group was 6 (25%) for T1-T2, 9 (38%) for T1-T3, 12 (50%) for T1-T4, 0 (0%) for T2-T3, 14 (58%) for T2-T4 and 14 (58%) for T3-T4, respectively. As for T2-T3, the lowest score by the Mandarin group was 0.95. There were four Mongolian bilinguals whose score for this pair was 0.94 (16%).

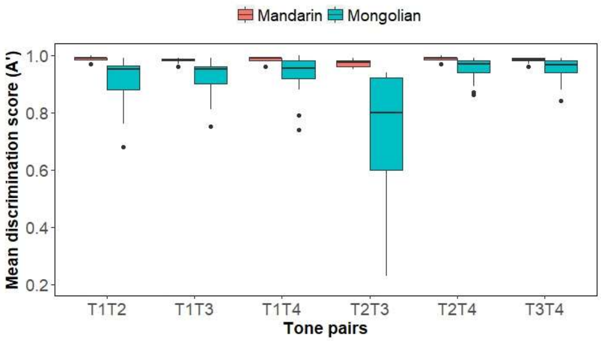

Figure 1: *The distribution of discrimination (A') scores by two (Mandarin, Mongolian) groups of listeners. The horizontal line in the box indicates the median. The bottom and top lines of the box indicate the first and third quartiles. The points outside the box are outliers.*

A two-way repeated-measures analysis of variance (ANOVA) with Group (Mandarin, Mongolian) as a between-subjects factor and Tone Pair (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4) as a within-subjects factor was run to explore differences between the two groups of listeners. The main effects of Group (G) and Tone Pair (T) reached significance and so did the two-way interaction [G: $F(1, 32) = 12.9$, $p < .01$, T: $F(5, 160) = 27.8$, $p < .001$, G x T: $F(5, 160) = 9.7$, $p < .001$]. Post-hoc t-tests (Bonferroni-adjusted) showed that the Mandarin group was significantly more accurate than the

Mongolian group in discriminating all six tone pairs [$F(1, 32) = 5.0\text{-}12.5$, $p < .05$].

Welch's F-tests with Tone Pair as the independent variable reached significance only for the Mongolian group [$F(5, 63.1) = 6.2$, $p < .001$]. Table 3 shows the results of Welch's F-tests and Dunnett's Modified Tukey-Kramer pairwise multiple comparison post hoc tests. For the Mongolian group, the effect of Tone Pair reached significance for five comparisons (T2-T3 vs T1-T2, T2-T3 vs T1-T3, T2-T3 vs T1-T4, T2-T3 vs T2-T4, T2-T3 vs T3-T4). In other words, the Mongolian bilinguals discriminated T2-T3 less accurately than all other tone pairs ($p < .05$) as is clearly seen in Figure 1.

Table 3: *Results of Welch's F-test assessing the effects of Tone Pair for each group and Dunnett's Modified Tukey-Kramer pairwise multiple comparison tests (significance level at .05).*

Group	***df***	***F*-value**	***p*-value**	**Between-pair comparisons**
Mandarin	5, 25.1	2.5	5, 25.1	--
Mongolian	5, 63.1	6.2	5, 63.1	T2-T3 < T1-T2, T1-T3, T1-T4, T3-T4, T2-T4

4. Discussion

This study examined the discrimination accuracy of Mandarin lexical tones by two groups of listeners: 24 native speakers of Mongolian born and raised in the Inner Mongolia Autonomous Region in China and ten native speakers of Mandarin as controls. We found that the Mongolian bilinguals, as a group, were less accurate than the native Mandarin listeners in discriminating six Mandarin tone pairs, but that some bilinguals showed native-like tone perception. What differentiates bilinguals who are native-like and who are not in their Mandarin tone perception? What prevents some bilinguals from attaining native-like Mandarin perception? All Mongolian bilinguals had similar age and educational backgrounds (i.e. college level or higher) and it is unlikely that they differ fundamentally in their tone processing capabilities. However, they may differ substantially in their use of Mandarin in their daily life. In the case of ethnic minorities, the role of cultural identity may be influential in their Mandarin proficiency. In this respect, the Mongolian bilinguals may have similarities to heritage language speakers who are also highly variable in their linguistic competence [23].

The Mongolian bilinguals' discrimination accuracy for T2-T3 was significantly lower than that for the other five tone pairs. T2 and T3 have been frequently reported to be phonetically similar and difficult for listeners from diverse L1 backgrounds to differentiate [1, 2, 8, 9, 10, 11, 12, 21, 22, 23]. It would be useful to conduct a categorization task[3] to characterize how Mongolian bilinguals *identify* Mandarin tones, in particular, T2 and T3. Given that some studies [1, 2, 21] suggest T2 may be more problematic than T3, there may be a bias or difference in the direction of misperception of these two tones.

This study has implications for the effect of early bilingualism on cross-language tone perception and it also bears on language policy in multi-ethnic nations. If and to what extent the native Mongolian bilinguals resemble other ethnic minorities such as Tibetans and Uyghurs in their Mandarin tone processing skills remains unknown. Advancing knowledge on this topic is important for language policy, L2 acquisition and L1 maintenance in multi-ethnic nations. Both from the perspectives of language maintenance and cross-language speech processing, in future, it would be interesting to examine the Mongolian bilinguals' L1 Mongolian speech processing abilities to determine if and how their L1 and L2 perception skills interact. Comparing their L1 Mongolian speech characteristics to that of native Mongolian speakers without knowledge of Mandarin (i.e. Mongolian speakers born and raised in Mongolia) would facilitate the assessment of the effects of bilingualism on L1 maintenance on the one hand. Comparing their L2 Mandarin speech characteristics to that of late L1 Mongolian learners of Mandarin residing in Mongolia would shed light on the effects of bilingualism on L2 speech learning. The current study is limited to the Mongolian bilinguals' tone *perception* abilities and it would be important to examine their tone *production* abilities to assess the effects of bilingualism on tone processing as a whole.

The present study contributes to our understanding of how Mongolian bilinguals born and brought up in the Inner Mongolia Autonomous Region in China process lexical tones in L2 Mandarin. This research has theoretical implications for ultimate attainment of L2 speech learning by Mongolian-Mandarin bilinguals. It also has pedagogical implications for bilingual education policy, as very little data is currently available for the acquisition of Mandarin by ethnic minorities in China.

5. Conclusions

The Mongolian-speaking group was significantly less accurate than the control group of ten Mandarin speakers for all six tone pairs with T2-T3 showing the largest between-group difference. However, large individual variation was observed and some Mongolian bilinguals perceived Mandarin tones as accurately as native Mandarin speakers. Thus, native-like tone perception is attainable for bilinguals whose L1 is non-tonal.

6. Acknowledgements

This work was supported by the 11th Hakuho Foundation Japanese Research Fellowship (2016-2017) and the 2018 Endeavour Research Fellowship. We thank our host organizations (Waseda University for the first author and National Institute for Japanese Language and Linguistics for the second author) for their support and four anonymous reviewers for their time and input.

7. References

[1] C. Kiriloff, "On the auditory perception of tones in Mandarin," *Phonetica*, vol. 20, no. 2-4, pp. 63-67, 1969.

[2] Y. Wang, M. M. Spence, A. Jongman, and J. A. Sereno, "Training American listeners to perceive Mandarin tones," *Journal of the*

[3] We thank one of the anonymous reviewers for raising this possibility.

Acoustical Society of America, vol. 106, no. 6, pp. 3649-3658, 1999.
[3] Y. Gao, "Development of English language education in ethnic minority schools in Inner Mongolia Autonomous Region," *Intercultural Communication Studies*, vol. 20, no. 2, pp. 148-159, 2011.
[4] A. Iivonen and H. Harnud, "Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt," *Journal of the International Phonetic Association*, vol. 35, no. 1, pp. 59-71, 2005.
[5] W.-S. Lee and E. Zee, "Standard Chinese (Beijing)," *Journal of the International Phonetic Association*, vol. 33, no. 1, pp. 109-112, 2003.
[6] A. Karlsson, "The intonational phonology of Mongolian", In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing*, pp. 187-215. Oxford University Press, 2014.
[7] Yurong, *Corpus-based Comparative Study of Prosodic Features in Japanese and Mongolian*, Hakuho Foundation Japanese Research Fellowship Final Report, 2017.
[8] C. So and C. T. Best, "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," *Language and Speech*, vol. 53, no. 2, pp. 273-293, 2010.
[9] C. So and C. T. Best, "Phonetic influences on English and French listeners' assimilation of Mandarin tone to native prosodic categories," *Studies in Second Language Acquisition*, vol. 36, no. 2, pp. 195-221, 2014.
[10] K. Tsukada, H. L. Xu and N. Xu Rattanasone, "The perception of Mandarin lexical tones by listeners from different linguistic backgrounds," *Chinese as a Second Language Research*, vol. 4, no. 2, pp. 141-161, 2015.
[11] K. Tsukada, M. Kondo and K. Sunaoka, "The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience," *Journal of Second Language Pronunciation*, vol. 2, no. 2, pp. 225-252, 2016.
[12] K. Tsukada and J.-I. Han, "The perception of Mandarin lexical tones by native Korean speakers differing in their experience with Mandarin," *Second Language Research*, DOI: 10.1177/0267658318775155, In press.
[13] J. E. Flege, I. R. A. MacKay and D. Meador, "Native Italian speakers' perception and production of English vowels," *Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2973-2987, 1999.
[14] J. E. Flege, "Methods for assessing the perception of vowels in a second language," In E. Fava and A. Mioni (Eds.), *Issues in Clinical Linguistics*, pp. 19-44, Padova: UniPress, 2003.
[15] R. Wayland and S. Guion, "Perceptual discrimination of Thai tones by naïve and experienced learners of Thai," *Applied Psycholinguistics*, vol. 24, no. 1, pp. 113-129, 2003.
[16] R. P. Wayland and S. Guion, "Training English and Chinese listeners to perceive Thai tones: A preliminary report," *Language Learning*, vol. 54, no. 4, pp. 681-712, 2004.
[17] S. Smith, *User Manual for UAB Software.* University of Alabama at Birmingham (UAB), Department of Rehabilitation Sciences, 1997.
[18] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer* [version 6.0.19], retrieved from http://www.praat.org (Last viewed June 13, 2016), 2016.
[19] W. Strange and V. L. Shafer, "Speech perception in second language learners: The re-education of selective perception," In J. G. Hansen Edwards and M. L. Zampini (Eds.), *Phonology and Second Language Acquisition*, pp. 153-191, Amsterdam: John Benjamins, 2008.
[20] J. G. Snodgrass, G. Levy-Berger, and M. Haydon, *Human Experimental Psychology.* New York: Oxford University Press, 1985.
[21] Y.-C. Hao, "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," *Journal of Phonetics*, vol. 40, no. 2, pp. 269-279, 2012.
[22] Y.-C. Hao, "Second language perception of Mandarin vowels and tones," *Language and Speech*, vol. 61, no. 1, pp. 135-152, 2018.
[23] C. B. Chang and Y. Yao, "Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin," *Heritage Language Journal*, vol. 13, no. 2, pp. 134-160, 2016.