LEARNER VS NON-LEARNER DIFFERENCE IN THE PERCEPION OF MANDARIN LEXICAL TONES: COMPARISON OF LISTENERS FROM ENGLISH AND JAPANESE FIRST LANGUAGE (L1) BACKGROUNDS

Kimiko Tsukada[1], Kaori Idemaru[2]

[1]Macquarie University, [2]University of Oregon
kimiko.tsukada@gmail.com, idemaru@uoregon.edu

ABSTRACT

Mandarin is one of the most representative tonal languages 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 first language (L1) backgrounds. We examined how individuals differing in L1 (English, Japanese) and experience with Mandarin (learners, non-learners) might respond to six pairs of Mandarin tones using a four-alternative forced-choice discrimination test. The results showed that while Japanese non-learners generally outperformed English non-learners, possibly benefitting from contrastive use of pitch accent in L1, two groups of learners did not differ in their perception of Mandarin lexical tones. This suggests that English speakers can overcome the initial disadvantage and learn lexical tones in a new language as successfully as speakers of other Asian language.

Keywords: cross-language perception, Mandarin lexical tones, English, Japanese.

1. INTRODUCTION

Cross-language processing of Mandarin lexical tones has been examined in many studies, but research involving non-native *learners* is still limited [4, 5, 19]. In particular, there seems to be a lack of studies that have directly compared learners and non-learners from multiple L1 backgrounds. To fill this gap in our current understanding, in this study, we included both learners and non-learners from two L1 backgrounds: English and Japanese. The aim was to determine if comparable L1 effect is observed for individuals with and without Mandarin learning experience. We hope to gain valuable insights into if and how different L1s might impact on the learning process of Mandarin lexical tones. The results obtained would provide useful knowledge for improving listening and communication skills in Mandarin.

The two languages of interest in this study, English and Japanese, differ in many ways. Crucially, while English has no lexical tones, Japanese uses pitch accents contrastively. For example, the word *asa* means 'morning' if it is pronounced with a high-low pitch pattern in standard Tokyo Japanese, but the meaning changes to 'linen' if it is pronounced with a low-high pitch pattern. Thus, it might be expected that native Japanese speakers are more sensitive to and skilled than native English speakers in processing Mandarin lexical tones if this skill transfers positively across languages.

However, as reviewed in [18], pitch accents in Japanese phonetically differ from lexical tones in Mandarin [14]. Specifically, Japanese pitch accent is not realized within a single syllable whereas each syllable is a tone bearer in Mandarin [12]. Further, the pitch range in Mandarin is reported to be twice as wide as that of Japanese. Thus, it is possible that Japanese speakers' prior experience with pitch accents may not be as beneficial as one might expect. In fact, there is a split in the literature on cross-language tone processing and whether L1 experience with lexical tone plays a facilitative [8] or inhibitory [12, 19] role is unresolved.

2. METHODS

2.1. Speakers and speech materials

The experimental stimuli and procedures were identical to those used in previous research [15-18]. Eight (4 males, 4 females) native Mandarin speakers with a mean age of 27.8 years (*sd* = 9.2) 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 and identified themselves as native speakers of Mandarin. They were recorded in a sound-treated studio on the 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 ___”). All materials were transcribed in Chinese characters along with pinyin (the Romanized spelling system of Chinese characters with tone symbols indicated by diacritics) 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. The tokens produced in isolation were used as the stimuli for this study. The stimuli presented to listeners were seven CV syllables (where C = /p, t, m/ and V = /i, a, u/) across all four Mandarin tones.

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”	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”

2.2. Participants

Four groups of listeners participated in this study (Table 2). The first two groups had English as their L1 and the other two groups had Japanese as their L1. Each L1 group consisted of participants with and without Mandarin learning experience. The results of Japanese participants were reported in our previous study [18].

Table 2: Participants in this study. Standard deviations are in parentheses.

L1	Mandarin experience	Gender	Age
English	Yes	13M, 10F	26.2 (12.8)
	No	8M, 18F	24.9 (9.7)
Japanese	Yes	6M, 4F	19.9 (1.2)
	No	10M, 11F	22.8 (3.8)

All participants were university students and participated in the study in their home countries or in Australia. The two learner groups included participants at different (beginner to advanced) levels of proficiency of Mandarin. Unfortunately, it is difficult to objectively control learners’ level of proficiency in foreign language speech research due to large individual variation and to different curricula at different institutions (e.g. availability of immersion and flagship programs in USA).

Four of the English-speaking learners had experience living in Mandarin-speaking locations (e.g. Beijing, Hohhot, Shanghai, Taiyuan) with an average length of residence of 2.8 years. Based on the enrolled course codes and/or self-report, this group included nine beginners, five intermediate, six upper intermediate and three advanced (flagship program) learners. Five of the Japanese-speaking learners had experience living in Mandarin-speaking locations (e.g. Beijing, Hong Kong, Shanghai, Taiwan) with an average length of residence of 5 years. Based on the instructor’s assessment and/or self-report, this group included four beginners, three intermediate, two upper-intermediate and one advanced learner.

2.3. Procedures

This study used a four-alternative forced-choice oddity task, which was employed in previous second language (L2) speech research [2, 21, 22], to assess listeners’ perception of six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4). As described in [21, 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 participants 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 [11] for the Japanese participants and English participants tested in Australia. The PRAAT program [1] was used for the English participants tested in USA. The listeners heard the stimuli at a self-selected, comfortable amplitude level over the high-quality headphones on a notebook or desktop computer. The experimental session was self-paced and the listeners could take a break after each block if they wished.

The stimuli were presented in triads and the listeners were given four (‘1’, ‘2’, ‘3’, ‘NO’) response categories. Each of the six pairs was tested by change and no-change (catch) trials. The three tokens in all trials were spoken by three different talkers of the same gender, 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 [13].

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 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 tone 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.

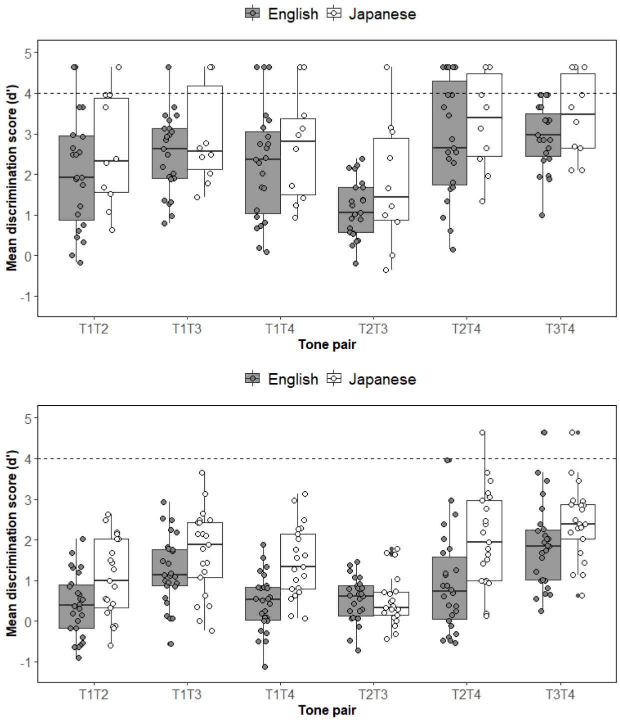
Responses to the change and no-change trials were used to calculate d-prime (d') scores by following the bias-correction procedure [9]. These scores were based on the proportion of 'hits (Hs)' and the proportion of 'false alarms (FAs)' obtained for each tone pair. The highest d' was set to 4.65.

3. RESULTS

Figure 1 shows the distributions of d' scores by two groups each of learners (top panel) and non-learners (bottom panel) as a function of tone pair. L1 effect appeared clearer for non-learners than for learners.

Averaged across the six tone pairs, the English and Japanese learner groups had mean d' scores of 2.3 and 2.8, respectively. On the other hand, the English and Japanese non-learner groups had mean d' scores of 0.9 and 1.5, respectively. The native control's mean d' score was 3.99. The extent of between-group differences varied depending on the tone pair as seen in Figure 1.

Figure 1: The distributions of d' scores for six tone pairs by two groups of learners (top) and non-learners (bottom). The bold horizontal line in each box indicates the median. The bottom and top lines of the box indicate the first and third quartiles. The small points outside the box are outliers. The dashed horizontal line indicates the mean value (3.99) for the Mandarin control group (n = 10).



A three-way repeated-measures analysis of variance (ANOVA) with L1 (L: English, Japanese) and experience (E: learner, non-learner) as between-subjects factors and tone pair (T: T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, T3-T4) as a within-subjects factor reached significance for all main and interaction effects except for a L1 x Experience interaction effect (Table 3). The significant three-way interaction was explored by separate Experience x Tone pair ANOVAs for learner and non-learner groups.

Table 3: Results of L1 x Experience x Tone ANOVA.

Factor	*df*	*F*	*p*
L	1, 76	7.4	< .01
E	1, 76	42.3	< .001
T	5, 380	88.2	< .001
L x E	1, 76	0.1	*ns*
L x T	5, 380	3.3	< .01
E x T	5, 380	7.4	< .001
L x E x T	10, 380	3.3	< .01

Tables 4 and 5 show the results of two-way ANOVAs which assessed the effects of L1 and tone pair for learner and non-learner groups, respectively.

Only the tone effect reached significance for the learner groups (Table 4), who had the greatest difficulty with T2-T3. T2-T3 is known to be highly confusing for listeners from diverse L1 backgrounds [e.g. 3-7, 10, 16-18, 20, 23, 24].

Table 4: Results of L1 x Tone pair ANOVA: Learners.

Factor	*df*	*F*	*p*
L	1, 31	1.4	*ns*
T	5, 155	33.3	< .001
L x T	5, 155	0.1	*ns*

Table 5: Results of L1 x Tone pair ANOVA: Non-learners.

Factor	*df*	*F*	*p*
L	1, 45	7.6	< .01
T	5, 225	64.4	< .001
L x T	5, 225	8.0	< .001

For the non-learner groups, the main effects of L1 and tone and a L1 x Tone pair interaction reached significance (Table 5). While the Japanese non-learners were least accurate for T2-T3, the English non-learners discriminated T1-T2, T1-T4 and T2-T4 poorly in addition to T2-T3. Both groups discriminated T3-T4 most accurately, which is consistent with our findings involving listeners from other L1 backgrounds [15, 17].

Table 6: Results of *t*-tests assessing the between-group differences: Non-learners.

Tone pair	*df*	*t*	*p*	Between-group comparisons
T1-T2	38.3	-2.5	< .05	E < J
T1-T3	38.4	-1.6	*ns*	--
T1-T4	37.6	-4.3	< .001	E < J
T2-T3	40.5	0.2	*ns*	--
T2-T4	42.8	-3.3	< .01	E < J
T3-T4	44.6	-2.0	*ns*	--

Table 7: Results of one-way ANOVAs assessing the effect of Tone pair and multiple comparison tests (Bonferroni-adjusted): Non-learners.

L1	*df*	*F*	*p*	Between-pair comparisons
English	5, 150	11.0	< .001	T1-T2, T1-T4 < T1-T3, T3-T4; T2-T3 < T3-T4; T2-T4 < T3-T4
Japanese	5, 120	11.2	< .001	T2-T3 < T1-T4, T1-T3, T2-T4, T3-T4; T1-T2 < T2-T4, T3-T4; T1-T4 < T3-T4

Tables 6 shows the results of Welch's *t*-tests (not assuming equal variances) which assessed the difference between the English and Japanese non-learners. For three out of the six tone pairs, the Japanese group was significantly more accurate than the English group even though both groups were naïve to Mandarin. Table 7 shows the results of one-way ANOVAs which assessed the effect of Tone pair and post-hoc tests (Bonferroni-adjusted) for non-learners. The English and Japanese non-learners' discrimination accuracy varied widely depending on the tone pair (Figure 1).

4. DISCUSSION

This study examined the perception of Mandarin lexical tones by four groups of listeners differing in their L1 (English, Japanese) and experience with Mandarin (learner, non-learner). We were particularly interested in determining how these two factors might interact. In other words, would the two learner groups resemble each other to a greater extent than the two non-learner groups in discriminating Mandarin tone pairs despite the L1 difference?

We found that the L1 effect was limited to the non-learner groups. Familiarity with L1 pitch accent may have initially aided the Japanese non-learners, but the Japanese learners did not outperform the English learners. This discrepancy may be related to variability in pitch accent patterns across different Japanese dialects and also to a limited functional load of pitch accents in comparison to Mandarin lexical tones. Admittedly, the group size was much smaller for the Japanese learners than for the other groups and the two learner groups were not tightly matched in their level of Mandarin proficiency. These limitations need to be addressed in our future work to verify the results obtained in the present study.

5. CONCLUSIONS

While the Japanese non-learners tended to be more accurate than the English non-learners in their Mandarin tone discrimination, the two learner groups did not differ from each other. This suggests that Japanese speakers may benefit from using pitch contrastively in their L1. However, given that the two learner groups did not differ from each other, English speakers may be as capable as, if not more than, Japanese speakers in learning lexical tones in a new language in the long run.

6. ACKNOWLEDGEMENTS

This research was supported by the 2018 Endeavour Research Fellowship to the first author. We thank Hayli Brown and Laura Walter for research assistance and participants for making the study possible. Finally, we thank three anonymous reviewers for their time and input.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2016. Praat: Doing phonetics by computer [version 6.0.19]. http://www.praat.org.

[2] Flege, J. E., MacKay, I. R. A., Meador, D. 1999. Native Italian speakers' perception and production of English vowels. *J. Acoust. Soc. Am.* 106, 2973–2987.

[3] Hallé, P., Chang, Y.-C., Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *J. Phon* 32, 395–421.

[4] Hao, Y.-C. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *J. Phonetics* 40, 269–279, 2012.

[5] Hao, Y.-C. 2018. Second language perception of Mandarin vowels and tones. *Language and Speech* 61, 135–152.

[6] Huang, T., Johnson, K. 2010. Language specificity in speech perception: Perception of Mandarin tones by native and non-native listeners. *Phonetica* 67, 243–267.

[7] Kiriloff, C. 1969. On the auditory perception of tones in Mandarin. *Phonetica* 20, 63–67.

[8] Lee, Y.-S., Vakoch, D. A., Wurm, L. H. 1996. Tone perception in Cantonese and Mandarin: A crosslinguistic comparison. *J. Psycholinguistic Research* 25, 527–542.

[9] MacMillan, N. A., Creelman, C. D. 1991. Detection Theory: A User's Guide. Cambridge, United Kingdom: Cambridge University Press.

[10] Shen, X. S., Lin, M. 1991. A perceptual study of Mandarin tones 2 and 3. *Language and Speech* 34, 145–156.

[11] Smith, S. 1997. *User Manual for UAB Software.* University of Alabama at Birmingham (UAB), Department of Rehabilitation Sciences, Birmingham, Alabama, USA.

[12] 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, 273–293.

[13] Strange, W., Shafer, V. L. 2008. Speech perception in second language learners: The re-education of selective perception. In: Hansen Edwards, J. G. & Zampini, M. L. (eds.), *Phonology and Second Language Acquisition.* Amsterdam, the Netherlands: John Benjamins, 153–191.

[14] Sunaoka, K., Sun, Q., Hiki, S. 2009. Corpus-based Chinese language education: Development and assessment of Chinese tone discrimination CAI self-learning system based on learner corpus. *J. Liberal Arts (The School of Political Science and Economics, Waseda University)* 126, 79–98.

[15] Tsukada, K. Accepted for publication. Are Asian language speakers similar or different? The perception of Mandarin lexical tones by naïve listeners from tonal language backgrounds: A preliminary comparison of Thai and Vietnamese listeners. *Australian J. Linguistics.*

[16] Tsukada, K., Han, J.-I. In press. The perception of Mandarin lexical tones by native Korean speakers differing in their experience with Mandarin. *Second Language Research.*

[17] Tsukada, K., Kondo, M. In press. The perception of Mandarin lexical tones by native speakers of Burmese. *Language and Speech.*

[18] Tsukada, K., Kondo, M., Sunaoka, K. 2016. The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience. *J. Second Language Pronunciation* 2, 225-252.

[19] Wang, X. 2013. Perception of Mandarin tones: The effect of L1 background and training. *Modern Language J.* 97, 144–160.

[20] Wang, Y., Spence, M. M., Jongman, A., Sereno, J. A. 1999. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am.* 106, 3649–3658.

[21] Wayland, R., Guion, S. 2003. Perceptual discrimination of Thai tones by naïve and experienced learners of Thai. *Applied Psycholinguistics* 24, 113–129.

[22] Wayland, R. P., Guion, S. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54, 681–712.

[23] Wong, P. 2013. Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *J. Acoust. Soc. Am.* 133, 3649–3658.

[24] Wong, P., Schwartz, R. G., Jenkins, J. J. 2005. Perception and production of lexical tones by 3-year old, Mandarin-speaking children. *J. Speech, Language and Hearing Research* 48, 1065–1079.