Discrimination of Cantonese lexical tones by listeners with and without learning experience of tone languages

Yee-ping WONG

College of Professional and Continuing Education, The Hong Kong Polytechnic University
kookihee@yahoo.co.jp

ABSTRACT

This study investigated the perception of the six Cantonese lexical tones by (1) native Mandarin listeners, (2) novice Japanese-speaking listeners, (3) Japanese-speaking learners of Mandarin, and (4) Japanese-speaking learners of Cantonese. Analyses of A-prime scores in an AXB discrimination task revealed that all four groups showed high difficulty in discriminating acoustically similar pairs within the level-level pairs (T1-T3, T1-T6 and T3-T6) as well as the rise-level pair (T2-T5). Although native Mandarin listeners and learners of Cantonese generally performed better, the performance varies with tone pairs. In contrast, experience learning Mandarin tones was not significantly advantageous for tonal discrimination. Moreover, the overall performance was worse when speaker's gender was female. This result could be explained by F0 overlapping of the two female speakers. Native Mandarin listeners and learners of Cantonese are more resilient to intertalker variations as they managed to outperform the novice listeners in the more challenging voice condition.

Keywords: Chinese lexical tones, L2 speech perception, Cantonese, Japanese, Mandarin

1. INTRODUCTION

Both Cantonese and Mandarin use fundamental frequency (F0) as the primary perceptual cue to distinguish lexical meanings. There are six lexical tones in Cantonese—T1 (high level, 55), T2 (high rising, 25/35), T3 (mid-level, 33), T4 (low falling, 21), T5 (low rising, 23), and T6 (low level, 22) [2]. On the other hand, there are four tones that contrast in meaning in Mandarin—T1 (high level, 55), T2 (rising, 35), T3 (dipping, 214), and T4 (falling, 51) [10].

Although results of many previous studies suggest that speakers of tone languages generally outperformed those of non-tone languages in perception of lexical tones (e.g., [9, 24]), it is also known that in lexical tone perception, listeners with different first language (L1) and linguistic experience use different perceptual cues. For example, while both native speakers of Cantonese and Mandarin attend to both F0 height and direction in distinguishing lexical tones, Cantonese speakers do not appear to make use of duration as a useful cue [5, 6, 8, 11]. Moreover, the weighting on F0 direction and height may be different according to speakers' L1. It is reported that Taiwan Mandarin speakers were more confused than their English counterparts in tone perception. This result may be attributed to the difference between the tonal systems of Cantonese and Mandarin. Native Mandarin speakers assigned more weight to F0 direction than F0 height, whereas native Cantonese speakers were sensitive to both cues [5]. Mandarin speakers also showed difficulty in perceiving the three Cantonese level tones, especially between acoustically difficult pairs such as T3 and T6 [14]. In contrast to Mandarin speakers, native English speakers tend to rely on F0 height and average F0 [3, 4, 5, 6, 7]. [2] points out that listeners of non-tone languages perceived lexical tones mainly on a psychoacoustic mode. Although speakers of non-tone language perceive tones acoustically as non-linguistic units, they could somehow show greater sensitivity to subtle F0 differences within a tonal category [14]. It is found that native speakers of English and French could categorize, albeit in a different pattern [19, 20]. Furthermore, native Japanese speakers were able to categorize the Mandarin tones [16] and Cantonese tones [24] into their native pitch-accent categories.

Apart from the influence of L1 and psychoacoustic factors, learning the target tonal language as a second language (L2) also plays an important role in tone perception (e.g., [7, 23]). For instance, late advanced native English learners of Mandarin were approaching the perception pattern of native Mandarin speakers compared to the control group by attending to both F0 contour and average F0 in [7]. Meanwhile, although the influence of L1 is well-documented, less is known about the role of L2 in the perception of lexical tones in a third language (L3). Recently, L2 learning experience of Mandarin has been shown to modulate the perception of L3 Cantonese tones for English-speaking learners [14]. In addition, [25] reported that native Japanese learners of Mandarin outperformed novice Japanese speakers in perception of L3 Cantonese as well.

The use of F0 as perceptual cue is different in Cantonese, Mandarin, and Japanese. Japanese is a non-tonal language and pitch-accent pattern is used to contrast meaning. Combinations of high (H) pitch, low (L) pitch, and an accent (*) are used to indicate the meaning contrast in lexical words [27]. Contrastive words like *hashi* (H*L) ‘chopsticks’, *hashi* (LH) ‘edge’, and *hashi* (LH*) ‘bridge’ can form minimal pairs. Japanese pitch accent is primarily realized by F0 contour [21]. Thus, Japanese speakers are expected to be skilled at processing F0 variations. However, pitch accents in Japanese is said to be phonetically different from lexical tones as they are not realized within a single syllable [18, 22].

To obtain a more complete picture of the acquisition of lexical tones and the influence of L2, the present study further expanded the discrimination of Cantonese tone contrasts to native Japanese L2 learners of Cantonese and native Mandarin speakers. To the best of my knowledge, there is no previous studies that examined the tone perception of Japanese-speaking learners of Cantonese. This study fills this gap by comparing Mandarin and Cantonese learners’ performance in an AXB discrimination task and provides additional evidence as to the role of L1 and/or L2 in non-native tone perception.

2. METHOD

2.1. Participants

The study included four groups of listeners. Novice native Mandarin listeners (NM), who had no or minimal exposure to tonal languages other than their first language, consisted of 10 listeners (mean age: 25.5 years). Their performance provided baseline data for benchmark comparison. Novice native Japanese listeners (NJ) who had no or minimum exposure to any tone languages, consisted of 21 listeners (mean age: 30.0 years). To minimize the influence of accent pattern in dialects, only those who are from the Greater Tokyo Area were recruited. L2 Mandarin group (NJL2M) consisted of 12 learners (mean age: 25.0 years) who have learnt Mandarin as foreign or second language for an average of 2.1 years (SD=1.2). L2 Cantonese group (NJL2C) consisted of 10 learners (mean age: 47.2) who have learnt Cantonese as L2 for an average of 8.9 years (SD=8.1). NJL2M self-reported their Mandarin proficiency as beginner–upper intermediate and they had no prior experience with non-Mandarin tonal languages. NJL2C also reported the same for Cantonese. All participants reported no hearing problems or language disorders, and neither had professional musical training.

2.2. Stimuli

Two female and two male native speakers of Hong Kong Cantonese produced all stimuli. All combinations of the six tone pairs were included to allow a more thorough comparison. The two test stimuli of [jɐu] and [se] (/jau/ and /se/ in Jyutping used in Hong Kong), which were used in previous studies (e.g., [14]), were selected for the present study. The six tones with these two syllables all form real Cantonese words. The four speakers each read two sets of words (one set for /jau/ and one set for /se/) in the carrier sentence at a normal speech rate. They were asked to read several times and the tokens with duration which were closest to the average of the six tones were selected and segmented for the discrimination task. The intensity of all the stimuli was normalized at 70dB. The F0 values of the stimuli of the two male speakers were 124.6Hz and 128.2Hz, while that of the two female speakers were 184.5Hz and 132.3Hz on average. Figure 1 shows an example using F0 of [se]. Three native speakers identified the stimuli and the accuracy was approximately 96%.

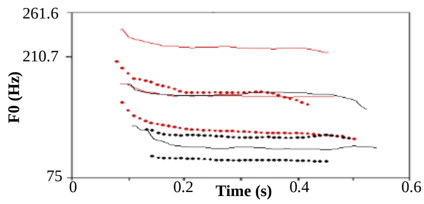

Figure 1: T1 and T6 of [se] showing F0 range of the 4 native speakers (Black=Male, Red=Female).

2.3. Procedure

The AXB discrimination task was carried out using Praat. There were 15 A–B pairs for the 6 tones and all four possible combinations (AAB, ABB, BAA and BBA) were included. Participants received 480 trials (2 syllables x 15 pairs x 4 combinations x 2 voices x 2 repetitions) presented in a mixed-talker design, in random order, and blocked by 60 trials. Within each trial, the first and third sound were uttered by one speaker and the second sound was uttered by another speaker of the same gender. 10 extra trials were included for practice. The interstimulus interval was set to 1s and the intertrial interval was 3s. For each trial, participants chose either the label ‘1’ (X=A) or the number ‘3’ (X=B) on the keyboard. The process was self-paced and lasted for around 1 hour, with short breaks scheduled between blocks.

2.4. Data analysis

Responses of each tone pairs were used to calculate A-prime (A’) scores [16], an index of discrimination accuracy. A’ was calculated based on the proportion of ‘hits’(H) and ‘false alarms’(F) for each pair. If H was equal to or exceeded F, A’ was calculated as (1). Otherwise, A’ was calculated as (2).

(1) $A' = 0.5 + ((H-F)*(1+H-F)) / (4H*(1-F))$
(2) $A' = 0.5 - ((F-H)*(1+F-H)) / (4F*(1-H))$

Mixed-effects model ANOVA was then performed on the A' scores data with Tone Pair and Speaker Gender as within-subjects factors, and Group as between-subjects factor. In the model, participant and syllable were crossed random variables. Post-hoc pairwise comparisons (with Bonferroni adjustment) were conducted to further explore the effects and interactions using the emmeans package of R.

3. RESULTS

The results of mixed-effects model ANOVA using Satterthwaite approximation revealed that the main effects of all three factors were significant: tone pair ($F(14,1421) = 161.6$, $p < .001$); gender ($F(1, 1421) = 632.1$, $p < .001$); group ($F(3, 49) = 3.6$, $p = .02$). The two-way interactions of the three factors (pair x group: $F(42, 1421) = 1.5$, $p = .02$; gender x group: $F(3, 1421) = 52.4$, $p < .001$; pair x gender: $F(14, 1421) = 52.4$, $p < .001$) are also all significant. Three-way interaction was non-significant. Paired t-tests were then carried out for multiple comparisons.

First, the overall scores across the 15 tone pairs were significantly lower for the difficult tone pairs which have similar F0 direction. The A' score was the lowest for T3-T6, followed by T1-T3, T1-T6, and T2-T5, and then T3-T4 and T4-T6. The scores of other pairs were significantly higher. Second, the A' score was higher when the stimuli were presented in female voice than in male voice. Third, the averaged A' scores across the 15 tone pairs of NJL2C group was significantly higher than the NJ group (0.894 vs. 0.835). The NM group, with mean score 0.889, also outperformed NJ group with marginal significance ($t(49) = 2.5$, $p = .09$).

The concerning interaction term of the current study—group x tone pair, is illustrated in Figure 2. Results of post-hoc analyses showed that NJL2C outperformed NJ for T1-T3 ($t(259) = 3.8$, $p = .001$) and T1-T6 ($t(259) = 3.0$, $p = .02$). NJL2C also outperformed NM for T2-T5 ($t(259) = 2.7$, $p = .05$). The A' score of NJL2C was also marginally significantly higher than NJL2M for T3-T6 ($t(259) = 2.5$, $p = .08$). On the other hand, NM outperformed NJ for T1-T6 ($t(259) = 2.6$, $p = .06$), T3-T4 ($t(259) = 2.4$, $p = .09$), T3-T5 ($t(259) = 2.5$, $p = .09$), T4-T6 ($t(259) = 2.6$, $p = .05$), and T5-T6 ($t(259) = 2.4$, $p = .09$). Further, the score of NM was also significantly higher than NJL2M for T3-T6 ($t(259) = 3.1$, $p = .01$).

Table 1 displays the mean scores of the four groups in voice of different gender. The performance was always better for all four groups when the speaker's voice was male (all $p < .001$). Moreover, only when the stimuli were presented in a female voice, significant between-group difference was observed (NJL2C > NJ, $t(59) = 3.6$, $p = .004$; NM > NJ, $t(59) = 4.0$, $p = .001$).

The mean discrimination scores by tone pair between the two types of speaker gender are given in Table 2. The score differences between the speaker gender were significant for all but level-rising tone pairs T1-T2, T2-T3, T2-T6, T3-T5, and T5-T6. Additionally, when the speaker is female, the scores of pairs T1-T6, T1-T3, and T3-T6 were substantially lower, followed by T2-T5, then T4-T6 and T3-T4. By contrast, when the speaker is male, the scores were significantly lower for T3-T6, followed by T2-T5, then T1-T3, and next T1-T6, T5-T6, and T4-T6.

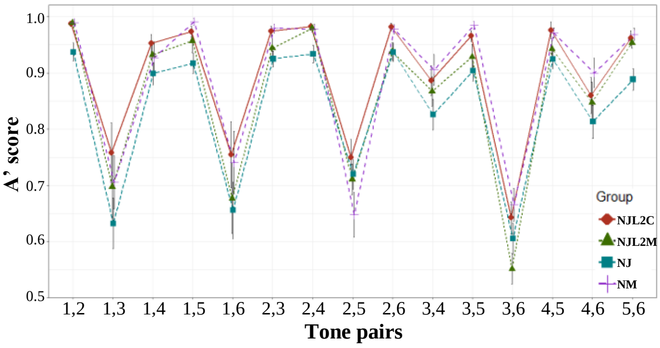

Figure 2: Average A' scores for each tone pair by the four groups (error bars = ±1 standard error).

Group	Female	Male
NJ (n=21)	0.754 (0.236)	0.915 (0.123)
NM (n=10)	0.843 (0.197)	0.935 (0.123)
NJL2M (n=12)	0.795 (0.223)	0.927 (0.126)
NJL2C (n=10)	0.835 (0.194)	0.953 (0.088)

Table 1: Mean A' scores of the four groups by speaker's gender. Standard deviations are in parentheses.

Tone pairs	Female	Male
T1-T2	0.953 (0.093)	0.984 (0.036)
T1-T3	0.477 (0.208)	0.892 (0.116)
T1-T4	0.864 (0.126)	0.980 (0.041)
T1-T5	0.921 (0.110)	0.981 (0.071)
T1-T6	0.444 (0.210)	0.946 (0.068)
T2-T3	0.936 (0.095)	0.962 (0.061)
T2-T4	0.939 (0.083)	0.985 (0.047)
T2-T5	0.637 (0.157)	0.783 (0.127)
T2-T6	0.946 (0.082)	0.961 (0.081)
T3-T4	0.769 (0.151)	0.954 (0.065)
T3-T5	0.918 (0.117)	0.955 (0.072)
T3-T6	0.559 (0.157)	0.664 (0.136)
T4-T5	0.911 (0.101)	0.984 (0.034)
T4-T6	0.740 (0.164)	0.952 (0.089)
T5-T6	0.916 (0.097)	0.949 (0.090)

Table 2: Mean A' scores for the 15 Cantonese tones by gender of the speaker. Standard deviations are in parentheses.

4. DISCUSSION

In general, tonal discrimination was poorest between the level-level tone pairs (i.e. T3-T6, T1-T3, T1-T6) and the rising-rising tone pair T2-T5 across the four groups. The performance was also poor for the lower register tone pairs, T3-T4 and T4-T6. These pairs were reported to be more difficult contrast pairs (e.g., [12, 15]). Previous studies have pointed out that the acoustic properties of the tonal stimuli robustly affect listeners' perception of tones regardless of their language backgrounds. Tone pairs having the same F0 direction within each pair are deemed to be acoustically hard to distinguish. Moreover, although substantial differences were found due to the voice of speaker, the discrimination of level-rising tone pairs were minimally affected. This suggests that phonetic similarities between tone pairs might be the most influential factor in the perception of non-native tones in the present study. Some L2 speech perception models (e.g., [1]) propose that phonetic similarities determine the assimilation or categorization of L2 sounds into L1 sounds. Tones that are more dissimilar acoustically will be easier to discern whereas tones with more similar features are more likely to cause confusion and learning difficulties.

Overall, the findings of this study show that the three native Japanese speaker groups exhibit different perception patterns in relation to their linguistic experience of tone languages. NJL2C group and NM group performed considerably better than the other two groups. Furthermore, NJL2C outperformed NJ group for the more difficult pairs, T1-T3, T1-T6, T2-T5, and T3-T6. This suggests the positive learning effect of Cantonese. In contrast, while NM group also performed better for six relatively difficult tone pairs (i.e. T1-T6, T3-T4, T3-T5, T3-T6, T4-T6, and T5-T6), it is surprising that NJL2M did not have advantage over the NJ group. This result is in contrast with findings of many previous studies [e.g., 14, 26]. Specifically, in [26], the data of 21 novice Japanese listeners and 10 Japanese-speaking learners of Mandarin were also included in the present study. The learner group showed certain advantages over the novice NJ group. Additionally, [14] reported that native English learners of Mandarin benefited from both their L1 and L2 (Mandarin) in perceiving L3 (Cantonese) tones and they outperformed the native English controls only when the F0 direction of the tone pairs were different (T1-T2 and T2-T6). By contrast, the results of this study revealed that NJL2M group did not perform better with pairs that differ in F0 direction nor pairs that differ in F0 height.

There are no middle level nor real level low-low pitch accent pattern in Japanese. A sequence of high tones is also not allowed in word-initial position in Tokyo Japanese. Additionally, there are no rising pitch accent patterns which differ in the magnitude of F0 change. It is assumed that NJ's selections depended on phonetic similarities. Meanwhile, NM are sensitive to F0 direction. It seems that their L1 (Mandarin) profoundly affects the discrimination of pairs with different F0 direction, especially when F0 are crowded in the lower register region, i.e. T3-T4, T3-T5, T4-T6 and T5-T6. NM outperformed the NJ group in distinguishing tone pairs with close F0. It is possible that the complex Cantonese tone system may have posed difficulties for Mandarin learners since their phonetic system have been reorganizes in response to Mandarin. The interlanguage phonology of NJL2M is in-between NJ and NM. Apart from psychoacoustic cues, NJL2M learned to use F0 direction (and duration) as cues in tone perception. Yet, the Mandarin learning experience of NJL2M group might not be sufficient to facilitate distinction of difficult Cantonese tone contrasts. Comparing to [26], given the fact that the two newly added Mandarin learners were still at the beginner stage (one year learning experience, they could easily feel confused when they perceive unfamiliar tone contrasts. Although the present findings do not support the claim that exposure to L2 Mandarin tones could positively modulate L3 (Cantonese) tonal perception, still, it is believed that tonal experience in Mandarin has strengthened the sensitivity of tone contour of NJL2M. In fact, the result of a Pearson correlation test showed that there is weak positive correlation ($r(358) = .1$, $p = .01$) between the length of learning and A' scores of NJL2M group. With more learning experience, Mandarin learners could potentially show additional advantages over novice listeners. To test this claim in future studies, data from more native Japanese learners of Mandarin with different proficiency levels are needed.

Similar to [26], the performance of all four groups were significantly higher when the speaker's voice was male. One justification may be that the two female speakers in this experiment spoke in very different pitches (as illustrated in Figure 1). The F0 of the lower register tones of one female speaker actually overlap with the F0 of the higher tones of the other female speaker. Although it has been noted in [13] that Cantonese tone identification can be heavily influenced by intertalker variations, the performance of NJL2C and NM were more robust even though the discrimination was conducted using female voices. This result could be accounted for by the linguistic experience of L2 Cantonese learning and the higher sensitivity to lexical tones of native Mandarin learners respectively. In other words, these two groups were less adversely affected by the speaker variations than the other two groups.

5. ACKNOWLEDGEMENTS

The author would like to thank Keiichi Tajima, Albert Lee and the reviewer for their helpful advice and constructive comments, as well as all the participants for their time. Any errors are, of course, my own.

6. REFERENCES

[1] Best, C. T. 1995. A direct realistic view of cross-language speech perception. In: Strange, W. (ed), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 171-204.

[2] Burnham, D., Mattock, K. 2007. The perception of tones and phones. In: Bohn, O. S., Munro, M. J. (eds), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 259-280.

[3] 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.

[4] Gandour, J. 1981. Perceptual dimensions of tone: evidence from Cantonese. *Journal Of Chinese Linguistics* 9(1), 20–36.

[5] Gandour, J. T. 1983. Tone perception in far Eastern languages. *Journal of Phonetics* 11, 149-175.

[6] Gandour, J. T. 1984. Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics* 12, 235-261.

[7] Guion, S. G., Pederson, E. 2007. Investigating the role of attention in phonetic learning. In: Bohn, O. S., Munro, M. J. (eds), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 57-77.

[8] Khouw, E., Ciocca, V. 2007. Perceptual correlates of Cantonese tones. *Journal of Phonetics* 35(1), 104–117.

[9] 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.

[10] Li, C. N., Thompson, S. A. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

[11] Liu, S., Samuel, A. G. 2004. Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech* 47(2), 109–138.

[12] Mok, P., Lee, C. W. T., Yu, A. C. L. 2018. Perception and production of Cantonese tones by South Asians in Hong Kong. *Proc. Speech Prosody 2018* Poznán, 458-462.

[13] Peng, G., Zhang, C., Zheng, H. Y., Minett, J. W., Wang, W. S. Y. 2012. The effect of intertalker variations on acoustic–perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research* 55, 579-595.

[14] Qin, Z., Jongman, A. 2016. Does second language experience modulate perception of tones in a third language?. *Language and Speech*, *59*(3), 318-338.

[15] Qin, Z., Mok, P. 2013. Discrimination of Cantonese tones by speakers of tone and non-tone languages. *Kansas Working Papers in Linguistics* 34, 1-25.

[16] Snodgrass, J. G., Levy-Berger, G., Haydon, M. 1985. *Human experimental psychology*. New York: Oxford University Press.

[17] So, C. K. 2010. Categorizing Mandarin tones into Japanese pitch-accent categories: The role of phonetic properties. *Proc. Second Language Studies: Acquisition, Learning, Education and Technology* Tokyo.

[18] 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.

[19] So, C. K., Best, C. T. 2011. Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics* 47(1), 133-145.

[20] 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, 95–221.

[21] Sugiyama, Y. 2017. Perception of Japanese Pitch Accent without F0. *Phonetica* 74(2), 107–123.

[22] 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. *Journal of Liberal Arts (The School of Political Science and Economics, Waseda University)* 126, 79-98.

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

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

[25] Wong, Y. 2019a. Categorization of Cantonese Lexical Tones by Japanese-speaking Novice Listeners and Learners of Chinese Mandarin. *Proc. International Conference on Phonetics and Phonology* Tokyo, 77.

[26] Wong, Y. 2019b. Perception of Cantonese lexical tones by Japanese non-learners and learners of Mandarin. *Proc. 19th ICPhS* Melbourne, 3867-3871.

[27] Wu, X., Tu, J., Wang, Y. 2012. Native and nonnative processing of Japanese pitch accent. *Applied Psycholinguistics* 33(3), 623-641.