

Perception of Cantonese lexical tones by Japanese non-learners and learners of Mandarin

YeePing Wong

Sophia University
kookihee@yahoo.co.jp

ABSTRACT

This study examined the perception of the six Cantonese lexical tones by Japanese listeners who are naïve to tone languages and Japanese-speaking L2 learners of Mandarin. Their performance on an AXB discrimination task was analyzed. The results showed that the L2 learner group performed slightly better in general, which suggests the L2 experience of lexical tones can be advantageous for the perception of L3 tones. However, the A-prime scores revealed that their performance only differed significantly for some acoustically easy pairs. They performed no better than the naïve group for pairs with a high degree of phonetic similarity. Specifically, both groups showed high perceptual confusion of acoustically hard pairs within the level-level pairs (T3-T6 and T1-T6) as well as the rise-level tone pair (T2-T5). In addition, the discrimination was generally better when the speaker gender was male. The L2 learners appeared to be more resistant to speaker gender variation.

Keywords: Cantonese, Mandarin, Japanese, lexical tone, speech perception

1. INTRODUCTION

Both Cantonese and Mandarin use F0 as the primary perception cue to distinguish lexical meanings. There are 6 lexical tones in Cantonese—T1, a high level (55) tone; T2, a high rising (25/35) tone; T3, a mid-level (33) tone; T4, a low falling (21) tone; T5, a low rising (23) tone; T6, a low level (22) tone [1]. These 6 tones are unevenly distributed, with 5 of them crowded into the lower part of the tone space. Some tones are harder to discriminate than others due to acoustic similarities. There are 3 level tones, making them internally ambiguous and difficult to perceive in isolation [13]. Besides, T2 and T5 were found to be confusing even for some native listeners as they only differ in the magnitude of the rise in F0 [12]. Tone merging is indeed undergoing between T3 and T6, as well as T2 and T5. On the other hand, there are 4 tones in Mandarin [10]—T1, a high level (55) tone; T2, a rising (35) tone; T3, a dipping (214) tone; T4, a falling (51) tone. The 4 Mandarin tones

are distributed evenly and more distinguishable. T1 and T2 of the two languages are similar in terms of F0 height and contour shape, but there is no contrast between level tones in Mandarin.

Most previous studies showed that listeners of tone languages outperformed listeners of non-tone languages in perception of lexical tones (e.g. [9, 23]). Native Mandarin listeners discriminated two Thai tones (mid vs. low) better than English speakers did in [23]. Native English listeners tended to rely on F0 height while Cantonese and Mandarin listeners focused on both F0 height and direction [5, 6]. However, some other studies showed different findings (e.g. [5, 14, 17, 21]). For instance, native Mandarin listeners showed difficulty in perceiving the three Cantonese level tones, especially between acoustically hard pairs such as T3 and T6. This suggests that the perception performance was affected by the acoustic properties of the tonal stimuli [14]. While F0 is the primary cue in tone perception, the weighting on F0 direction and height may be different according to listeners' L1. Taiwan Mandarin listeners assigned more weight to F0 direction than F0 height, whereas Cantonese listeners were sensitive to both cues [5]. Given the differences between the two tonal systems, they confused even more than the English listeners. Moreover, [2] suggests that listeners of non-tone languages perceived lexical tones mainly on a psychoacoustic mode. In [20], native speakers of English and French were able to categorize Mandarin tones in terms of their intonational categories. Although non-tone language native listeners perceived tones acoustically as non-linguistic units, they could somehow show greater sensitivity to subtle F0 differences within a tonal category [14].

Apart from the influence of L1 and psychoacoustic factors, linguistic experience also plays a role in L2 sound perception. Effects of L2 learning experience or training treatment are well-documented (e.g. [4, 7, 17, 22]). [7] found that late advanced L1 English learners of Mandarin were approaching the perception pattern of native Mandarin listeners compared to the control group by attending to both F0 contour and average F0. In [4],

native Mandarin and native English listeners were trained to recognize Cantonese lexical tones. While L1 Mandarin listeners began to behave like native Cantonese listeners by showing nearly equivalent cue weighting of F0 direction and height after training, L1 English listeners attended solely to F0 height.

Japanese is a pitch-accent language and it may be expected that Japanese can be skilled at processing pitch variations. It is possible that Japanese listeners may assimilate T2 in Mandarin and Cantonese to phonetically similar Low-High (LH) accent pattern in Japanese [18, 21]. Indeed, [18] reported that Japanese speakers were able to categorize the Mandarin tones into their native pitch-accent categories. It is assumed that their selections depended on the phonetic similarities between Mandarin tone categories and Japanese pitch-accent categories. However, there are only 4 tones in Mandarin and the pitch height and direction are fairly distinctive, so the assimilation results are actually rather predictable. Furthermore, pitch accents in Japanese is said to be phonetically different from lexical tones as they are not realized within a single syllable [19, 21]. Pitch range in Mandarin, Cantonese and Japanese may be different as well.

This study expanded the discrimination of Cantonese contrasts to novice native Japanese listeners and Japanese L2 learners of Mandarin. The effect of Mandarin learning was predicted to be minimal and the 2 groups of Japanese listeners would not substantially differ as the contour tones and level tones are indeed quite different in the two languages (Mandarin and Cantonese).

2. METHOD

2.1. Participants

21 Japanese-speaking novice listeners and 10 learners of Mandarin participated in this study. They are either university students or working adults who grew up in the Greater Tokyo Area in Japan (3 learners speak another dialect apart from standard Japanese). All of them had not been exposed to Cantonese before. The non-learner group reported that they have no or minimum exposure to any tone languages, while the Mandarin learners reported the same except their substantial exposure to Chinese Mandarin. The L2 learners self-reported their Mandarin proficiency as beginner-upper intermediate. Their average age of acquisition was 19.20 years (SD: 3.79) and they had received an average of 2.31 years (SD: 1.21) of Mandarin instruction. The participants all passed a pure-tone hearing screening prior to the experiment. None of

them reported hearing problems and language disorders. Information about their musical experience was collected. Some of them had played a musical instrument in the past but none are playing as a professional constantly in the past 2 years.

A Hong Kong Cantonese native speaker checked the stimuli were in proper quality. Three Cantonese native speakers identified the stimuli and the accuracy was 96% approximately.

2.2. Stimuli

All combinations of the 6 tone pairs were included to allow more thorough comparison. The two test stimuli of [jəu] and [se] (/jau/ and /se/ in Jyutping used in Hong Kong), which were used in previous studies (e.g. [8, 21]), were selected for the present study. The 6 tones with these 2 syllables all form real Cantonese words. Two female and two male native speakers of Hong Kong Cantonese produced all the stimuli in a quiet room. They read 2 sets of words (1 set for /jau/ and 1 set for /se/) in the carrier sentence *ngo5 duk6__zi6* “I read the word __” at a normal speech rate. They read many times and the exemplars with duration which were closest to the average of the 6 tones were selected and excised for the perception experiment. Intensity of all the stimuli was normalized at 70dB. The F0 values of the stimuli of the 2 male speakers were 124.6Hz and 128.2Hz, and that of the 2 female speakers were 184.5Hz and 132.3Hz on average.

2.3. Procedure

Participants were run on an AXB discrimination test using Praat. The AXB trials had four possible combinations (AAB, ABB, BAA and BBA). There were 15 A–B pairs for the 6 tones. 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. 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 number ‘1’ (X=A) or the number ‘3’ (X=B) on the keyboard. The process was self-paced and lasted for around 60 minutes, 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).

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

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

A mixed-effects model ANOVA was then conducted on the A' scores with Tone Pair and Speaker Gender as within-subjects factors, and Group as between-subjects factor. The model included by-subject random intercepts and slopes.

3. RESULTS

Normal repeated measures ANOVA model showed that the main effect of group was significant. However, according to the results of likelihood ratio tests, with the random effects of subjects in the mixed-effects model, the effect of Group was only marginally significant ($\chi^2(1) = 3.53, p = .06$). The effect of Tone Pair ($\chi^2(13) = 318.86, p < .0001$), the effect of Speaker Gender ($\chi^2(0) = 150.07, p < .0001$), as well as the two-way interaction of Pair x Gender ($\chi^2(14) = 447.87, p < .0001$) were significant. No other main effect or interactions were found significant. Similar results were obtained using Satterthwaite approximation. Paired t-tests (with Tukey adjustment) were conducted to further compare the effects and to explore the interactions.

First, as predicted, the overall scores across the 15 tone pairs were significantly lower for the 4 difficult tone pairs, T1-T3, T1-T6, T2-T5, and T3-T6, with T3-T4 and T4-T6 in-between, and the scores of other pairs were significantly higher. With regards to the performance of the 2 groups, averaged A' scores across the 15 tone pairs were 0.835 and 0.876 for the novice Japanese listeners and native Japanese listeners who were learners of Mandarin. Generally, the L2 Mandarin learners' scores were higher even though no significance was reported. The mean discrimination scores for each tone pair of the 2 groups are given in Table 1. The differences between the 2 groups were marginal significant at an alpha level of 0.1 for the pairs T1-T2, T1-T3, T1-T5, T3-T4, T3-T5 and T5-T6. However, although not significant, the Mandarin learners actually did not outperform for T1-T6 and even did worse than the novice group for T2-T5 and T3-T6.

Besides, within the Mandarin-learner group, the score differences were significant between all the combinations involved pairs T1-T3, T1-T6, T2-T5 and T3-T6. On the other side, on top of these combinations, those involved T3-T4 and T4-T6 were also found different significantly within the Mandarin-learner group. This is because the average score for pairs T4-T6 and T3-T4 were relatively low, which was not the case within the learner group.

Moreover, as illustrated in Figure 1, although the interaction was non-significant, the effects of speaker's gender on the 2 groups were rather

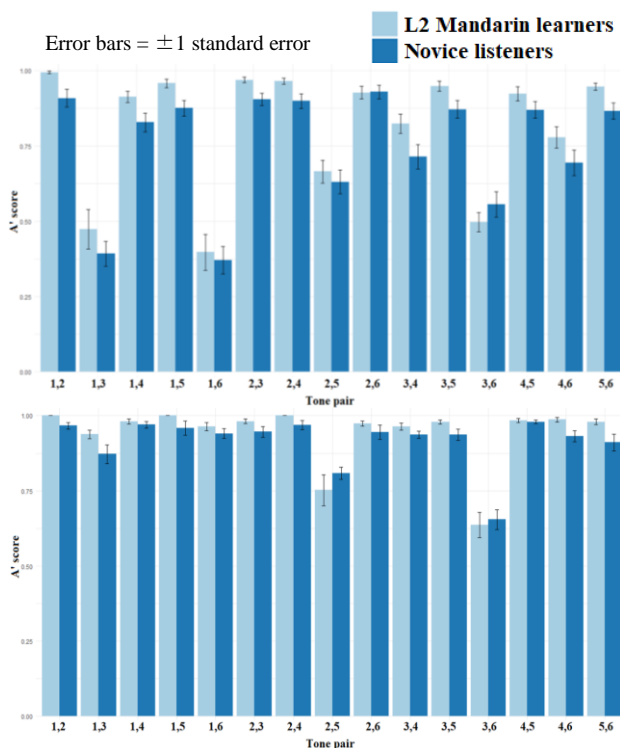
different. The average scores of the novice group for stimuli in a female and male voice were 0.754 and 0.915 respectively, whereas those of the Mandarin-learner group were 0.812 and 0.941. While both groups did significantly better when the speaker's voice was male (both $p < .0001$), the difference of scores between the 2 groups was marginally significant ($t(29) = 2.00, p = .06$) when the speaker's voice was female.

The scores of the stimuli in male voice appeared to be higher for all pairs. The interaction of speaker's gender and tone pairs was further investigated. The results of paired t-tests showed that listeners performed significantly better when the speaker's voice was male for almost all pairs except T1-T2, T2-T3, T2-T6 and T5-T6. The differences were particularly obvious for the tone pairs T1-T3 ($t(292)=19.32, p < .0001$) and T1-T6 ($t(292)=23.27, p < .0001$). The differences were also large for the pairs T2-T5, T3-T4, T3-T6 and T4-T6 (t values ranged from 3.52 – 9.13). It is worth noting that the difference between the pairs T1-T3 and T3-T6 was only significant when the speaker's voice was male (male: 0.91 vs. 0.65, $p < .0001$; female: 0.43 vs. 0.53, $p = .37$ respectively). The score of the pair T1-T3 was much higher when the speaker's voice was male, but the score dropped significantly for the lower level pair T3-T6 (refer to Figure 1). In addition, within the learner group, the score differences between male and female speaker were significant for pairs T1-T3, T1-T6, T2-T5, T3-T4, T3-T6 and T4-T6. In contrast, within the novice group, the differences were significant for all but pairs T2-T3, T2-T6 and T5-T6.

Table 1: Mean A' scores for the 15 Cantonese tones by the two groups of listeners. Standard deviations are in parentheses.

Tone pairs	Novice listeners (n=21)	L2 Mandarin learners (n=10)
T1-T2	0.937 (0.105)	0.997 (0.010)
T1-T3	0.632 (0.293)	0.706 (0.280)
T1-T4	0.899 (0.128)	0.947 (0.056)
T1-T5	0.917 (0.120)	0.979 (0.039)
T1-T6	0.656 (0.326)	0.680 (0.320)
T2-T3	0.925 (0.088)	0.975 (0.026)
T2-T4	0.934 (0.100)	0.982 (0.029)
T2-T5	0.720 (0.169)	0.708 (0.146)
T2-T6	0.937 (0.106)	0.950 (0.055)
T3-T4	0.826 (0.177)	0.893 (0.104)
T3-T5	0.904 (0.118)	0.963 (0.042)
T3-T6	0.605 (0.179)	0.566 (0.137)
T4-T5	0.924 (0.106)	0.953(0.063)
T4-T6	0.813 (0.190)	0.883 (0.132)
T5-T6	0.889 (0.125)	0.963 (0.034)

Figure 1: Mean A' scores of 15 tone pairs uttered by female (top panel) and male (bottom panel) speakers in L3 Cantonese for Japanese-speaking novice listeners and L2 Mandarin learners.



4. DISCUSSION

Overall, the findings of this study showed that the Japanese-speaking L2 Mandarin learners as well as the novice listeners were significantly better at discriminating the contour-level tone pairs than the level tone pairs. Furthermore, Japanese Mandarin learners who had L2 experience with a tone language had some advantages over the novice listeners. The advantages were larger with rising-level tone pairs such as T1-T2 and T1-T5. Nevertheless, they were not better than the novice group for the more difficult contrast pairs, that is, the level-level tone pairs, T3-T6, and the rising-rising tone pair, T2-T5. This is in line with results of previous studies (e.g. [15]), which suggest that the acoustic properties of the tonal stimuli robustly influence listeners' perception of tones regardless of their language backgrounds. These 2 pairs are said to be acoustically hard as the F0 direction are the same within each pair. In particular, due to the inherently smaller acoustic difference between T3 and T6 [14], T3-T6 was even more difficult than T1-T3 or T1-T6.

There is no middle level nor real level low-low pitch accent pattern in Japanese. A sequence of high tones is not allowed in word-initial position in Tokyo Japanese as well. There is also no level tone contrast in Mandarin. In fact, [15] has reported that Mandarin speakers had difficulty with the level-level

tone pairs in Cantonese. If listeners are to assimilate the different level tones in Cantonese into the single high tone (T1) in Mandarin, or the closest high-high pitch accent pattern in Japanese, their sensitivity to the different level tones could be reduced.

Interestingly, the performance of pairs T3-T4 and T4-T6 were in the middle of the 'difficult' tone pair group and 'easy' tone pair group. It seems that Japanese naïve listeners had some extra difficulty in distinguishing the low falling tone (T4) with a level tone (middle level (T3)/low level (T6)) which is close to each other in terms of F0 value. In fact, [11] reported that South Asians in Hong Kong also had difficulty in distinguishing the lower register tones (T3-T6, T4-T5, T4-T6 and T5-T6). The perception of non-native sounds may be based on the phonetic similarities between L1 and L2 sounds. Tones which are more dissimilar acoustically will be easier to discern while tones with more similar features may cause more perceptual and learning difficulties for listeners. These claims are consistent with L2 speech perception models such as the Speech Learning Model [3]. Cantonese learners may be advised to focus on tone exercises with those harder pairs.

Additionally, the A' scores of both groups were higher when the speaker's voice was male in general. It may be related to the fact that one of the female speakers in this study has a pitch range much lower (107-179Hz) than the other female speaker (162-244Hz), such that the pitch of her voice was actually quite similar to a male's pitch (94-174Hz). Listeners might have tried to calibrate and adjust their criteria of judgment of F0 based on a low pitch. This could have caused them extra difficulty in discriminating tones with F0 height, which could be even worse when the pairs were level tones or had similar F0 direction. In fact, [14] has found that Cantonese tone identification can be heavily influenced by intertalker variations, especially for T3, T6 and T5. The linguistic experience of L2 Mandarin learners might have helped them so that they were more consistent and performed slightly better in the harder condition (i.e. stimuli in a female voice).

In order to examine Cantonese tone perception by Japanese listeners in detail, it is necessary to compare the performance of Japanese-speaking Cantonese learners in the future. Data from more learners and learners with higher proficiency level are also desirable. Additionally, in light of the error pattern found in the present study, it might also be informative to find out how native Japanese speakers actually perceptually categorize Cantonese tones into their L1 prosodic system (i.e. Japanese pitch accent patterns). It is possible that sound categorization will occur phonologically, with little influence of L2 experience of Mandarin learning.

5. ACKNOWLEDGEMENTS

The author would like to thank Keiichi Tajima for his helpful advice and comments, as well as all the participants for their time. Thanks also go to Peggy Mok and Kimiko Tsukada for their encouragement and generous share of information.

6. REFERENCES

- [1] Bauer, R. S., Benedict, P. K. 1997. *Modern Cantonese Phonology*. Berlin and New York: Mouton de Gruyter.
- [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] Flege, J. E. 1995. Second-language speech learning: theory, findings, and problems. In: Strange, W. (eds), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 233-273.
- [4] Francis, A. L., Ciocca, V. 2003. Stimulus presentation order and the perception of lexical tones in Cantonese. *J. Acoust. Soc. Am.* 114(3), 1611-1621.
- [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] Hallé, P. A., Chang, Y. C., Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of phonetics* 32(3), 395-421.
- [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] 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* Poznań, 458-462.
- [12] Mok, P., Zuo, D., Wong, P. 2013. Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese. *Language Variation and Change* 25, 341-370.
- [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. 2005. The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *J. Acoust. Soc. Am.* 117(4), 2427-2427.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] Wayland, R. P., Guion, S. G. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54, 681-712.