

Effects of Tone Merging and Musical Training on Cantonese Tone Perception

Peggy Mok, Donghui Zuo

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong

peggymok@cuhk.edu.hk, donghui@cuhk.edu.hk

Abstract

This study investigates whether musical training can facilitate lexical tone perception of native speakers of a tone language. Some Cantonese tone pairs, T2/T5 (rising), T3/T6 (level), T4/T6 (falling vs level) are merging in recent years. The merging subjects have poorer general tone perception than the control subjects. Previous studies showed that musical training facilitates lexical tone perception of non-tone language speakers. However, it is unclear if musical training can also influence tone perception of native speakers of tone-languages who are merging tones. Three groups of listeners (normal Cantonese, merging Cantonese, foreign) with and without advanced musical training participated in AX discrimination tasks of Cantonese tones and pure tones. Both accuracy and reaction time data show that while musical training can enhance lexical tone perception of foreign listeners, it has little influence on merging Cantonese listeners. The results indicate that different perceptual mechanisms may be involved in linguistic and musical tone perception, and that the linguistic use of tones is more fundamental and more robust than musical training.

Index Terms: music training, tone perception, tone merging, Cantonese

1. Introduction

Hong Kong Cantonese is well-known for its complex tone system with six lexical tones (T): T1 (high-level [55]), T2 (high-rising [25]), T3 (mid-level [33]), T4 (low-falling [21]), T5 (low-rising [23]) and T6 (low-level [22]) [1]. In the past decade, researchers have noticed that some Cantonese speakers in Hong Kong no longer distinguish all six tones in their production. The most notable merging pair is T2 [25] vs T5 [23] (e.g. [2, 3]). Some speakers also have a tendency to merge T3 [33] and T6 [22], and T4 [21] and T6 [22] respectively [4, 5]. Mok & Wong [4] showed that the merging native speakers have poorer general tone perception of Cantonese tones than their non-merging native counterparts, not only for the merging tone pairs. This shows that the merging speakers are less sensitive to linguistic pitch differences, which suggests that sound change may be caused by poor perception [6]. Further data is needed to confirm this idea, and to explore the effects of other possible factors related to pitch perception.

One such possible factor is musical training. Previous studies showed that musical training facilitates lexical tone perception of non-tone language speakers, e.g. [7-10]. However, it is unclear if musical training also influences tone perception of native speakers of tone languages who are merging tones. Given the functional role of pitch in both linguistic and musical tones, it is legitimate to ask whether musical training can also facilitate the perception of these merging speakers, and whether the same perceptual mechanism is used in both linguistic and musical pitch perception.

This study compares tone perception of merging speakers ('merging' hereafter) with that of the non-merging speakers who clearly distinguish all six Cantonese tones ('normal' hereafter). Only few studies on Cantonese tone merging used perception data, and they were mainly based on accuracy. In the current study, we examine both accuracy and reaction time data to capture the subtle processing of tone perception.

Moreover, this study also investigates the relationship between lexical tone perception and pitch sensitivity by comparing the data of both Cantonese and foreign listeners with advanced musical training ('musicians' hereafter) and those with no or very limited musical experience ('non-musicians' hereafter). Cantonese is an interesting target language because of its complex tone system. Previous studies on similar topics used either the Mandarin 4-tone system (i.e. level, rising, dipping and falling) [7-9] or the Assamese two-tone system (i.e. one falling and one rising) [10] in which the major difference lies in the pitch direction. It is unclear whether the same musical facilitatory effects can be found when the listeners need to distinguish more subtle differences in Cantonese in which both the pitch height and the magnitude of change are important cues for tonal distinction. The comparisons of native and non-native tone language listeners who are musicians and non-musicians in this study can also reveal a more thorough picture of the interactions between language background, musical training and tone perception.

2. Method

2.1. Subjects

Three groups of subjects participated in the experiment. The first group consisted of 34 native speakers of English or French who had not learned any tone languages before (i.e. foreign). The second and the third groups were native speakers of Hong Kong Cantonese. The second group consisted of 30 Cantonese speakers who could distinguish all 6 tones (i.e. normal), and the third group consisted of 28 Cantonese speakers who could not distinguish all 6 tones in their production (i.e. merging). The merging subjects were selected from 161 speakers with a screening process. Each speaker was recorded reading a list of 30 words (5 different words × 6 tones) embedded in a short carrier phrase. The recordings were auditorily checked by two native speakers of Cantonese who clearly distinguish all six tones to determine whether the speaker was likely to merge the tones. Only those who were identified by both judges were included in this study.

Each group of subjects was further divided into three categories by their musical background. Subjects with more than 7 years of formal musical training in any instrument or vocal singing and have been playing music in the past two years were classified as musicians. Subjects with no more than 2 years of casual musical experience and had not been playing music in the past two years were classified as non-musicians. Other subjects who fell in-between these two categories were classified as intermediates. The numbers of subjects in each

group are listed in Table 1.

All the subjects were students (local or exchange) at the Chinese University of Hong Kong, except five foreign musicians who were recruited at the University of Chicago. All subjects were paid to participate in the study.

Table 1. Numbers of subjects with different language and musical training backgrounds.

	foreign	normal	merging
Musician	10	10	10
Intermediate	10	10	7
Non-musician	14	10	11

2.2. Materials

The experiment was an AX discrimination task with two sets of stimuli – monosyllables and pure tones.

2.2.1. Monosyllables

The monosyllable stimuli were produced by a female researcher who clearly distinguishes the six Cantonese tones in her production. The tokens were paired up to form two types of stimuli pairs – 120 AA pairs (same-tone pairs) and 120 AB pairs (different-tone pairs). Altogether 60 target monosyllables (6 tones × 10 syllables) were chosen as the AA pairs together with 60 dummy items in order to balance the number of the AB pairs. These dummy items were excluded from analysis. For the AB pairs, 2 syllables of each tone which also appeared in the AA pairs were chosen. These 2 syllables are paired with the other 5 tones to form the AB pairs, for example, T1/T2, T1/T3, T1/T4, T1/T5, T1/T6. The order of the AB pairs is counter-balanced. This resulted in 120 AB pairs (6 tones × 2 syllables × 5 matching tones × 2 orders). The 120 AA and 120 AB pairs were randomized in the perception experiment.

2.2.2. Pure tones

The six pure tone stimuli were resynthesized from the monosyllable [wai] produced by the same female researcher mentioned above using Praat. The pitch contours of these pure tone stimuli were the same as the six Cantonese lexical tones, only that all segmental information is removed. There are in total 48 AA pairs (6 tones × 8 repetitions) and 60 AB pairs (6 tones × 5 matching tones × 2 orders). The pure tone tokens were also randomized in the perception experiment.

2.3. Procedures

The subjects participated in the perception experiment individually in a quiet room at the Chinese University of Hong Kong or the University of Chicago. The stimuli were presented to them via a stereo headphone using E-Prime 2.0 Professional with a desktop computer. A rest was given between the monosyllable section and the pure tone section. Short breaks were included within each section. There was a short practice before each section to familiarise the subjects with the tasks.

The subjects were asked to indicate whether the two tokens in the stimuli pairs carry the same tone or not by pressing different buttons on a serial response box. They were encouraged to respond as accurately and as quickly as possible. Both accuracy and reaction time were collected. The

inter-stimulus interval (ISI) was 500 ms. Time-out time was 10000 ms. No feedback was given. All RT values longer than 3000ms were counted as missing responses and were excluded from analysis. This resulted in 0.6% loss of data.

3. Results

3.1. Effects of language backgrounds

The accuracy (%Correct) and the reaction time (RT) data of the three language groups were analysed regardless of musical training. We log-transformed the RT data (LogRT) for normalisation. The results show that in general the normal subjects were the quickest and the most accurate while the foreign subjects were the slowest and the least accurate (see Fig. 1). The merging subjects' data fell in-between them. The averaged %Correct and LogRT () of the monosyllable task is 98.4% (2.863), 97.7% (2.923), 94.4% (2.967) for the normal, merging, foreign subjects respectively. One-way ANOVAs with language background as a factor shows that both the %Correct [$F(2, 89) = 7.727, p=0.001$] and the LogRT [$F(2, 89) = 11.980, p<0.0001$] are significantly different between groups. Figure 1 gives the results of the post-hoc tests. The asterisks (*) stand for significant difference in %Correct, and the number signs (#) stand for significant difference in LogRT. Both the normal and merging subjects were significantly more accurate than the foreign subjects ($p<0.05$). As for LogRT, the normal subjects were significantly faster than both the merging and foreign subjects ($p<0.05$). However, the normal and the merging subjects do not show significant difference in terms of %Correct ($p=0.829$), while the merging and foreign subjects do not show significant difference in LogRT ($p=0.157$). In brief, the merging subjects resemble the normal subjects in accuracy but resemble the foreign subjects in reaction time.

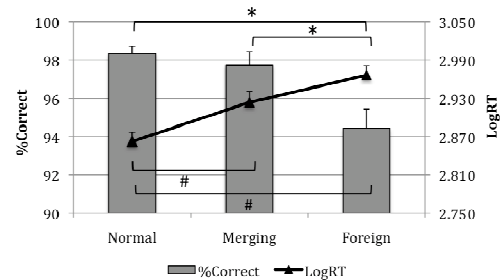


Figure 1: Average %Correct and LogRT of the three language groups in the monosyllable task

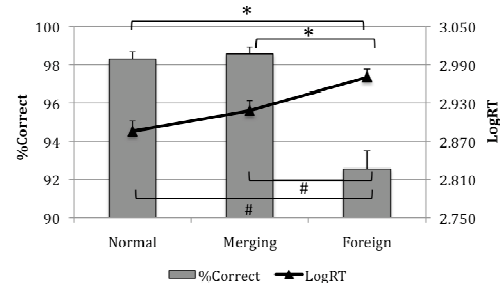


Figure 2: Average %Correct and LogRT of the three language groups in the pure tone task

A language difference emerged in the pure tone task (Figure

2). The normal and merging subjects are not significantly different in either the accuracy or the reaction time, while both groups are different from the foreign subjects. This suggests that the normal and the merging subjects are equally sensitive to pure tones which were resynthesized based on the contours of canonical Cantonese tones.

To further investigate the differences between the normal and merging subjects, we did t-tests for each tone pair. Since only one comparison between two groups of subjects was done for each tone pair, and no multiple comparison was involved across tone pairs, no correction for family-wise Type I error is needed. Due to the page limit, only the results of all the AA pairs and the confusing AB pairs (T2/T5, T5/T2, T3/T6, T6/T3, T4/T6, T6/T4) are presented here. The results show that in the monosyllable task, the merging subjects were significantly slower than the normal ones for all AA pairs (Figure 3) and most of the confusing AB pairs (Figure 4). The %Correct are not significantly different between them for almost all tone pairs, probably due to a ceiling effect as the merging subjects could still distinguish the 6 Cantonese tones.

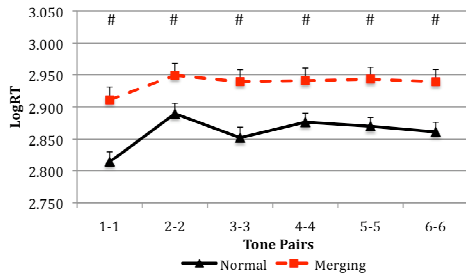


Figure 3: Cantonese listeners' LogRT of confusing tone pairs in the monosyllable task (AA pairs)

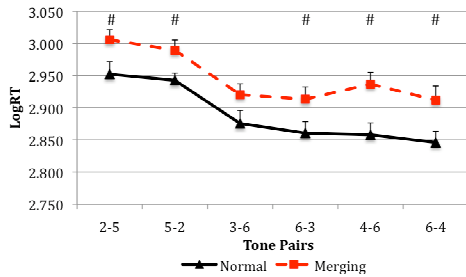


Figure 4: Cantonese subjects' LogRT of confusing tone pairs in the monosyllable task (AB pairs)

The difference in the reaction time indicates that the normal subjects are more sensitive to tones than the merging subjects are in general, not just for those confusing AB tone pairs. However, none of these tone pairs show significant difference in the pure tone task, which suggests that when the linguistic information is filtered out, the normal subjects do not have advantage over the merging subjects in perceiving pitch contours similar to the canonical lexical tones.

3.2. Effects of musical training

To investigate whether musical training affects the perception of tones, we compare the musicians' data and non-musicians' data in each language group.

The monosyllable and the pure tone tasks exhibit the same patterns. The results show that for foreign listeners, the musicians (%Correct Mean = 98.3) are significantly more

accurate than the non-musicians (%Correct Mean = 91.6) in tone perception [$t(13.828) = 14.402, p = 0.006$] (Figure 5). However, the difference in reaction time is not significant, which indicates that musical training helps the foreign subjects perceive the tones more accurately, but not faster. For the normal and merging subjects, however, musical training affects neither the %Correct nor the LogRT. Although the musicians have higher average %Correct than the non-musicians in both the normal and merging groups, the difference does not reach significance.

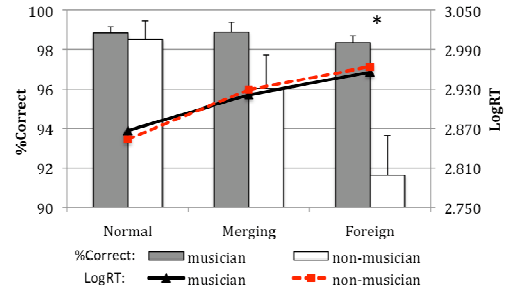


Figure 5: %Correct and LogRT of musicians and non-musicians in the monosyllable task

When individual tone pairs are examined, significant differences are found mostly in %Correct of the confusing pairs perceived by foreign subjects (Figure 6 & Figure 7). It is worth noting that comparing with the non-musicians, the musicians are generally more sensitive to the different magnitudes of change between T2 and T5. Musicians are either significantly more accurate or considerably faster than non-musicians in the T2/T5 and T5/T2 pairs in both the monosyllable and pure tone tasks.

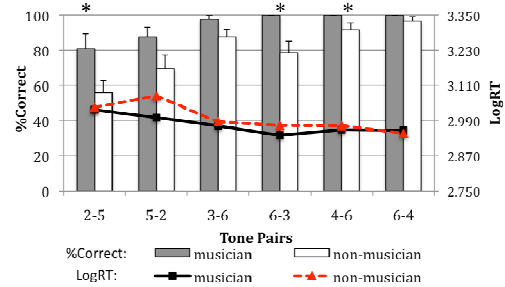


Figure 6: %Correct and LogRT of foreign musicians and non-musicians in the monosyllable task (AB pairs)

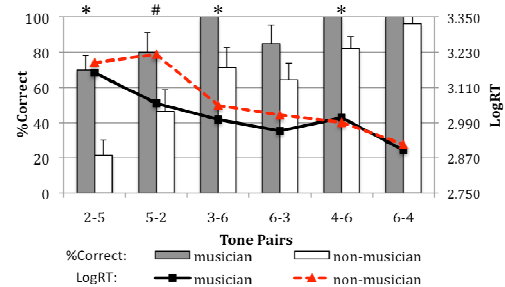


Figure 7: %Correct and LogRT of foreign musicians and non-musicians in the pure tone task (AB pairs)

In contrast with the foreign listeners, neither the %Correct, nor the LogRT of the native Cantonese listeners (i.e. the normal and the merging subjects) is affected by the level of

musical training. Figure 8 & Figure 9 shows the %Correct and the LogRT of the musicians and non-musicians from the native merging group. The difference between musicians and non-musicians in %Correct is very small compared with that of the foreign listeners. The LogRT lines also overlap with each other. None of the differences in these two variables approaches significance in any tone pair.

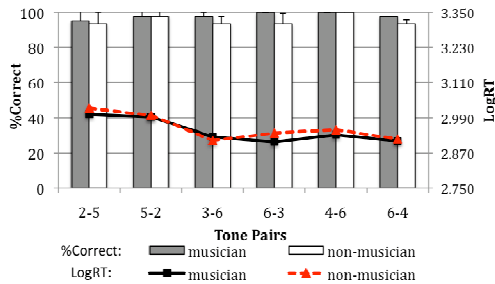


Figure 8: %Correct and LogRT of merging musicians and non-musicians in the monosyllable task (AB pairs)

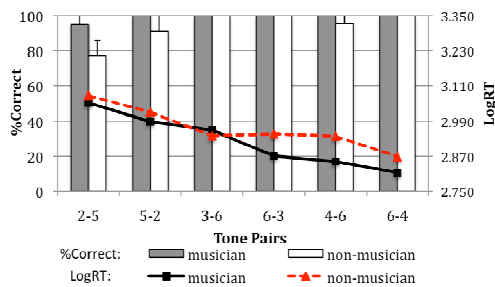


Figure 9: %Correct and LogRT of merging musicians and non-musicians in the pure tone task (AB pairs)

4. Discussion

One interesting finding of this study is that, although the monosyllable task suggests that the merging subjects are significantly slower than the normal subjects in perceiving lexical tones, no significant difference is found between the two groups in the pure tone task. The different performance indicates that the native speakers might be using two different mechanisms to perceive the lexical pitch and the music pitch.

Although the mechanisms may be separate, there can be some interaction too. The higher accuracy of the foreign musicians, as comparing with the foreign non-musicians in both the monosyllable and pure tone tasks, confirms the results of previous studies [e.g. 7-10] that musical training may facilitate the perception of lexical pitch for non-tone language speakers. This hypothesis is valid even when the tone system of the target language makes use of not only the pitch direction as in previous studies, but also the pitch height and the magnitude of change, such as Cantonese. Therefore, we support the argument that the music domain may have some facilitatory effects on the linguistic domain.

However, such facilitatory effect is not found in the native listener groups (i.e., normal and merging groups). The musician and non-musician groups are equally successful and fast in the monosyllable and the pure tone tasks. It is not surprising that these two groups perform equally well in the monosyllable task, as it is sufficient for them to depend solely on their linguistic knowledge to discriminate the tones in their native language. Nonetheless, the fact that the two groups do

not show any significant difference in the perception of pure tones may indicate that speakers of tone languages may process music pitch differently from non-tone language speakers. Yet, there are two alternative accounts for such phenomenon: 1) while musical training only helps listeners process musical pitch more accurately but not faster, the ceiling effect causes the lack of difference in the %Correct between the musician and the non-musician groups; 2) as the pure tones were resynthesized from canonical Cantonese tones, they resemble the linguistic stimuli very much so that the native listeners can also rely on their linguistic knowledge to process the sound. But given the discrepancy between the reaction time difference of the two groups of native subjects (i.e., normal and merging) in the monosyllable task and the pure tone task, the second possibility is unlikely to be a major reason for the observed patterns.

To conclude, the current results show that musical training can facilitate lexical tone perception only if musical training starts before the linguistic use of lexical tones, as in the case of foreign subjects. Musical training has little influence on lexical tone perception of tone language speakers, even for those with poorer general tone perception (merging subjects). The findings imply that different perceptual mechanisms may be involved in linguistic and musical tone perception. The linguistic use of tones is more fundamental and more robust than musical training.

5. Acknowledgements

We would like to thank Prof. Alan Yu for recruiting foreign musician subjects at Chicago for us.

6. References

- [1] Bauer, R. S. & Benedict, P. K. *Modern Cantonese Phonology*. Berlin, New York: Mouton de Gruyter, 1997.
- [2] Kei, J., Smyth, V., So, L. K. H., Lau, C. C. & Capell, K. Assessing the accuracy of production of Cantonese lexical tones: a comparison between perceptual judgment and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing*, 7: 25-38, 2002.
- [3] Bauer, R., Cheung, K. H., & Cheung, P. M. Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change*, 15: 211-225, 2003.
- [4] Mok, P & Wong, P. Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. In *Proceedings of Speech Prosody 2010*, Chicago, 2010.
- [5] Mok, P & Wong, P. Production of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables. In *Proceedings of Speech Prosody 2010*, Chicago, 2010.
- [6] Ohala, J. J. The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives*, 237-278, Longman, 1983.
- [7] Alexander, J., Wong, P., & Bradlow, A. Lexical tone perception in musicians and nonmusicians. In *Proceedings of Interspeech' 2005-Eurospeech-Ninth European Conference on Speech Communication and Technology*, Lisbon, 2005.
- [8] Gottfried, T. L. Music and language learning: Effect of musical training on learning L2 speech contrasts. In O. Bohn & M. Munro (Ed.), *Language Experience in Second Language Speech Learning*, John Benjamins, 221-237, 2007.
- [9] Lee, C. Y., & Hung, T. H. Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, 124: 3235-3248, 2008.
- [10] Wayland, R., Herrera, E., & Kaan, E. Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, 38: 654-662, 2010.