Perception of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables

Peggy Pik-Ki Mok , Peggy Wai-Yi Wong

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

peggymok@cuhk.edu.hk, peggywong@cuhk.edu.hk

Abstract

Traditionally, there are six lexical tones (T) in Cantonese, but some tone pairs appear to be merging in Hong Kong Cantonese. Some young speakers do not distinguish the two rising tones T2/T5, or the two level tones T3/T6, or the low falling and low level tones T4/T6. 16 potential mergers and 11 control subjects participated in a perception experiment with an AX discrimination task using monosyllables. Both accuracy rate and reaction time were measured. Results show that the potential mergers generally performed less well than the control group in having a lower accuracy rate and longer reaction time, but they could still distinguish the merging tone pairs with above 90% accuracy. Both groups found the T2/T5 pair difficult to distinguish. The results indicate that the merging processing of the tones is still in progress in the language as a whole and in individual speakers. Possible reasons for these patterns are discussed.

Index Terms: Cantonese, tones, mergers, perception, sound change

1. Introduction

Cantonese has a complex tone system. There are six contrastive lexical tones (T1 to T6) and three allotones (T7 to T9) which are shorter versions of T1, T3 and T6 in syllables ending with an unreleased final stop consonant (/p t k/). Unlike Mandarin, there is no 'neutral' tone for unstressed syllables in Cantonese. Each syllable, even function words, carries a distinct lexical tone [1,2]. Table 1 shows all Cantonese tones with examples. Figure 1 shows the F0 traces of the six lexical tones with the syllable [ji] produced by a female speaker.

Table 1. *Cantonese tones with examples.*

Tone number	Tone category	Example	Gloss
T1	high-level	ji^{55}	To cure
T2	high-rising	ji^{25}	Chair
T3	mid-level	ji^{33}	Idea
T4	low-falling	ji^{21}	Suspicious
T5	low-rising	ji^{23}	Ear
T6	low-level	ji^{22}	Two
T7 (T1)	high-stopped	jɪk^{5}	Benefit
T8 (T3)	mid-stopped	jak^{3}	Eat
T9 (T6)	low-stopped	jɪk^{2}	Also

It can be seen from both Table 1 and Figure 1 that the tonal distinction in Cantonese is based on both pitch height and pitch contour. The tones also differ in duration, but they were produced with a similar duration in Figure 1 for easy comparison. Tone 1 is separated from the other five tones by being at the top of the speakers' normal pitch range. The 'tonal space' is very crowded in the lower pitch range. Four tones (T2, T4, T5, T6) share the same starting pitch level. Besides, several tone pairs are particularly similar. The two rising tones T2 and T5 have very similar starting point, but one rises to a higher pitch level (T2) while the other only rises to a mid pitch level (T5). The two level tones T3 and T6 have a relatively small difference in pitch (around 30 Hz for the female speaker who produced the tones in Figure 1). T6 and T4 differ only in the slight fall towards the end in T4. Given such subtle differences in a narrow pitch range, these several tone pairs can be difficult to distinguish, especially when they are produced in isolation.

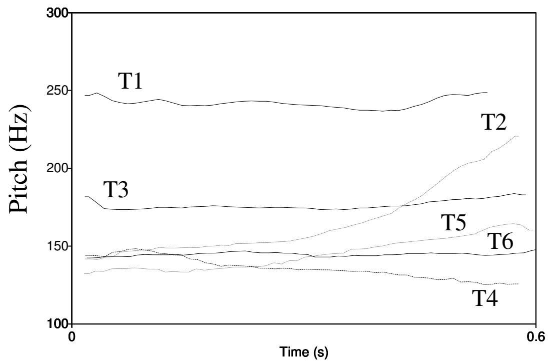

Figure 1. *F0 traces of the six lexical tones.*

The complex tone system in Cantonese attracts a lot of research attention in both production and perception [e.g. 3, 4]. These studies usually adopt a stable six-tone system as described above. Nevertheless, the Cantonese tone system in Hong Kong is undergoing changes in recent years in that some young speakers no longer distinguish some of the six tones in their speech. This is a fairly recent development because only few studies documented this phenomenon, although it is not too uncommon to notice these changes impressionistically. Kei et al. [5] studied Cantonese tone production by 56 speakers acoustically. They found that 6 of their speakers merged the two rising tones (T2 and T5) which they considered as 'tone production errors'. One speaker realised his T2 and T5 midway between the two canonical tones. Two speakers realised most of their T5 tokens as T2, while three speakers exhibited the opposite pattern. Bauer et al. [6] followed on their finding by replicating [5] using 8 male speakers. They also found that 2 of their speakers produced the rising tones unconventionally. One speaker merged T2 into T5; the other speaker merged T5 into T2. Perceptually, Cantonese speakers also often confused the two rising tones [7]. Taken together, these studies clearly suggest that some Hong Kong Cantonese speakers are merging the two rising tones, with three possible merging patterns: low-rising merging into high rising (T5 → T2); high rising merging into low rising (T2 → T5); and having a novel intermediate realisation.

The two studies on rising tones [5, 6] only investigated the production of T2/T5. It is unclear whether their speakers who merged T2/T5 can still distinguish the two rising tones in perception. In addition, to the best knowledge of the authors, there is no study documenting the production or perception of other potential tone mergers in Hong Kong Cantonese. However, impressionistically, some young speakers also mix up the mid level and low level tones (T3/T6), the low level (T6) and the low falling tones (T4). Clearly, more studies are needed to investigate the merging tone patterns in Cantonese and the factors contributing to such patterns. Our study investigates both the production and perception of potential tone mergers in Hong Kong Cantonese. Monosyllabic and disyllabic words as well as non-linguistic pure tones were used. In this paper, only preliminary perception data on monosyllabic words is presented. Production data on monosyllabic words can be found in [8]. Analysis of other data and the relationship between production and perception is currently underway.

2. Methods

2.1. Subjects

Since this study investigates the perception of tone mergers, it is important to ensure that we include speakers who do merge the tones. In order to recruit these speakers, a simple screening test was conducted. Each potential participant was recorded reading a word list with 18 words (3 different words × 6 tones) embedded in a short carrier phrase. Their recordings were auditorily checked by both authors to determine who was likely to be a merger. 129 participants were screened in total. 16 potential mergers were recruited. Table 2 shows the number of recruited speakers who showed signs of merging different tones. The numbers are not balanced because it was quite difficulty to locate these potential mergers. The recruited speakers participated in both production and perception experiments.

Table 2. *Number of potential tone mergers*

No. of speakers	Merging tone pairs
5	T2/T5
5	T3/T6
2	T4/T6
3	T2/T5; T3/T6
1	T3/T6; T4/T6

An additional 11 speakers who clearly distinguish all six tones were used as a control group. Thus, there were 27 subjects in total participating in the perception experiment. They were undergraduate students in the Chinese University of Hong Kong, aged between 18 to 22, with no history of hearing problems. They were paid to participate in the experiment.

2.2. Materials

The perception experiment using monosyllables was an AX discrimination task with two types of materials: 120 AA pairs and 120 AB pairs. There are several criteria for choosing the monosyllables: 1) 10 different syllables of each tone were included. 2) 5 of the 10 syllables are with all six tones attested in Cantonese, e.g. [ji] as shown in Table 1, while the other 5 syllables do not appear in all six tones, e.g. [wan] with T3 missing. 3) The syllables mostly consist of all sonorants. Altogether 60 target monosyllables (6 tones × 10 syllables) were chosen which were also used in the production experiment. They all appeared in 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. 1 syllable has all 6 tones attested while the other syllable does not. 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). A female speaker qualified as a speech therapist in Hong Kong produced all the monosyllables. Unattested matching tones for the AB pairs were presented to her in transcription. She had no difficulty in producing these tokens. In fact, many Cantonese speakers can produce the six tones of any syllable in a reciting manner easily with no problem. The 120 AA and 120 AB pairs were randomized in the perception experiment.

2.3. Procedures

The 16 potential merger subjects participated in the production experiment first. They were invited back for the perception experiment at least two weeks after the production experiment. The 11 control subjects only participated in the perception part. The perception experiment was divided into three sections: monosyllables, disyllabic minimal pairs and non-linguistic pure tones. A rest was given between sections. There was a short practice session before each section to familiarise the subjects with the tasks. The question for the monosyllable section was ‘Is the second syllable the same as the first?’ The response option was ‘same’ or ‘different’. The monosyllable section was divided into 4 blocks. The stimuli within each block were randomised for each subject. A short break was scheduled between blocks.

The subjects participated in the perception experiment individually in a quiet room at the Chinese University of Hong Kong. The stimuli were presented to them via a stereo headphone using E-Prime 2.0 Professional with a desktop computer. Both accuracy and reaction time data was collected using the PST Serial Response Box. A fixation point appeared on the screen before each trial. The subjects pressed the leftmost key on the response box with their left index finger for the ‘same’ responses and the rightmost key with their right index finger for the ‘different’ responses. No feedback was given. The inter-stimulus interval (ISI) was 500 ms. Reaction time was calculated from the onset of the second monosyllable. Time-out time was 10000 ms after the offset of the second monosyllable. Missing responses were excluded from the analysis.

3. Results

Figures 2 and 3 show both the accuracy rate and reaction time for the AA and AB pairs averaged across the 11 subjects in the control group. The general accuracy is very high with small variation of reaction time. For the AA pairs, the T1/T1 pair is the easiest for the subjects with a high accuracy rate and the shortest reaction time. The reaction time for the T2/T2 pair is slightly longer than the other tone pairs, but there does not appear to be any consistent pattern for other AA pairs. In general the control subjects can distinguish all AA pairs well. For the AB pairs, the control group could generally distinguish the pairs with T1 better than other tone pairs. They found the

T2/T5 pair most difficult to distinguish, resulting in the lowest accuracy rate and the second longest reaction time. However, the longest reaction time for the T2/T4 pair is unexpected, given that the two tones diverge in their contours (T2 being a high-rising tone and T4 being a low-falling tone). It is interesting to note that for both tone pairs (T2/T5 and T2/T4), the presentation sequence of the tones can affect the perception results. Listeners made more errors and had a longer reaction time when they heard T2 first: T2-T5 (88.6%, 931 ms) vs T5-T2 (97.7%, 889 ms) and T2-T4 (95.5%, 967 ms) vs T4-T2 (100%, 908 ms). This pattern is not found in other AB pairs with T2. In addition, except T2/T5, the control group did not have particular difficulty in distinguishing T3/T6 and T4/T6.

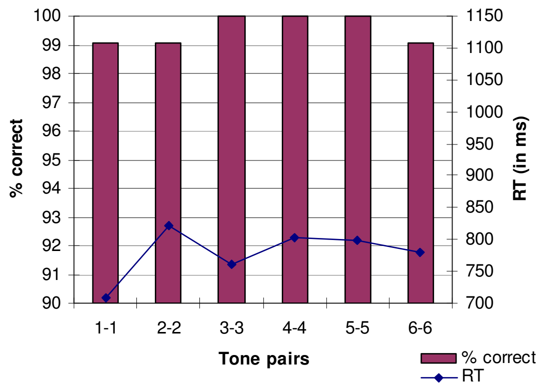

Figure 2. *Accuracy rate and reaction time for the AA pairs by the control group.*

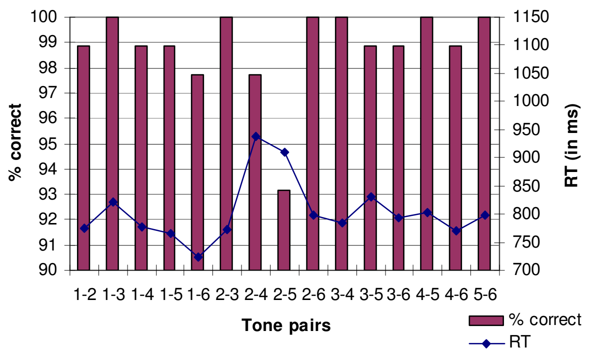

Figure 3. *Accuracy rate and reaction time for the AB pairs collapsed across sequences by the control group.*

A central question is whether the potential mergers differ from the control subjects in their tone perception. Figures 4 and 5 show the accuracy rate and reaction time for the AA and AB pairs averaged across all potential mergers (16 in total), since the number in each merger group varies. In general, the potential mergers have a lower accuracy rate and longer reaction time than the control group for both AA and AB pairs, indicating that they had more difficulty in distinguishing the tone pairs. Similar to the control group, they also found the T1/T1 pair the easiest to distinguish, but they clearly differ from the control group for the T2/T2 pair (high rising), resulting in the lowest accuracy and a longer reaction time. However, no such obvious difference is found for the T5/T5 (low-rising) pair.

For the AB pairs, similar to the control group, the potential mergers appeared to perform better in tone pairs with T1. The T2/T5 and T3/T6 pairs are most difficult for them, with the longest reaction time (T2/T5) and the lowest accuracy rate (T3/T6). The second longest reaction time for the T3/T5 pair implies that they also had troubles distinguishing these two similar tones: T3 (33) vs T5 (23). However, contrary to expectation, they did not show particular difficulty for the T4/T6 pair.

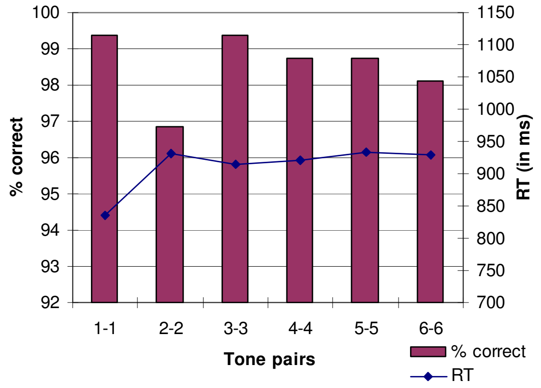

Figure 4. *Accuracy rate and reaction time for the AA pairs by all potential mergers.*

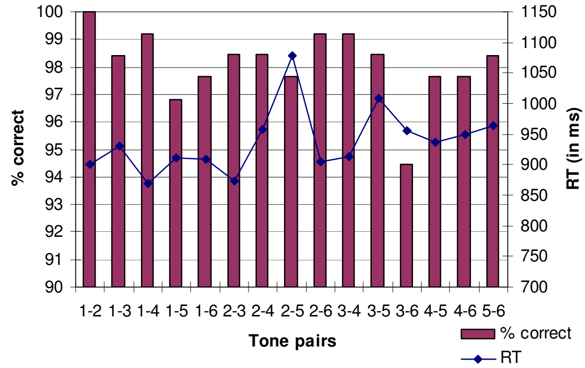

Figure 5. *Accuracy rate and reaction time for the AB pairs collapsed across sequences by all potential mergers.*

Since there are three different potential merger groups, it is necessary to see if the subjects in each group performed differently for the merging tone pairs. Table 3 shows their accuracy rate and reaction time for the three merging tone pairs. The accuracy data for the T4/T6 potential merger group should be interpreted with caution since there were only 3 subjects in the group. The reaction time data should reflect their performance better.

Table 3. *Accuracy rate and reaction time of the three tone pairs by three potential merger groups.*

Potential mergers (no.)	Tone pairs					
	T2/T5		T3/T6		T4/T6	
	%	RT (ms)	%	RT (ms)	%	RT (ms)
T2/T5 (8)	100	1104	94.6	964	94.6	1027
T3/T6 (8)	97.2	1060	95.7	936	98.6	886
T4/T6 (3)	95.8	942	95.8	855	100	805

All three potential merger groups found the T2/T5 pair most difficult to distinguish with the longest reaction time compared with all other AB tone pairs (data not shown). It is interesting to note that the subjects in the T2/T5 group could correctly distinguish all T2/T5 pairs, only with a very long reaction time (1104 ms). They also had difficulty in distinguish T3/T6 and T4/T6, as reflected by a low accuracy rate and long reaction time. The performance of the T3/T6 group is consistent with expectation in that they had a lower accuracy rate and fairly long reaction time for the T3/T6 pair. The pattern of the T4/T6 group is unexpected in that they performed well in distinguishing the T4/T6 pair. In fact, they had 100% accuracy rate for all tone pairs except T2/T5 and T3/T6. Nevertheless, since there were only 3 subjects in the group, their pattern should be treated as tentative only.

4. Discussion

The results reveal both similarities and differences between the control group and the potential merger groups. Both groups found the T1/T1 pair (AA) and tone pairs with T1 (AB) easier to distinguish. This is not surprising given that T1 is well-separated from the other five tones by being at the top of the speakers' normal pitch range. It is perceptually more salient than other tones. The adult perception data agrees with the acquisition patterns of Cantonese-speaking children in that T1 is acquired before other tones [9]. It is conceivable that T1 is the most stable and 'easiest' tone in the Cantonese tone system and most resistant to sound change (after the historical merge of (53) into (55) for Hong Kong Cantonese).

A comparison between the control group and the potential merger group reveals some interesting patterns. First of all, the potential merger group had poorer general performance in tone perception for both the AA and AB pairs, not only for the merging tone pairs. The control group also found the T2/T5 pair difficult to distinguish. This corresponds quite well with the idea that perception difficulty/confusion and listeners can be a source of sound change [10]. The acoustic similarity between the merging tone pairs renders them particularly susceptible to sound change.

Although the potential merger groups generally performed less well than the control group, their accuracy rate is still quite high (well above 90%). This indicates that they can still distinguish the tones in perception, only with more difficulty as reflected in a lower accuracy and a longer reaction time. These potential mergers were recruited because they could not distinguish the tones clearly in their production during the small scale screening test. Their production data also shows that six tones are retained in their tonal inventory, despite them being potential mergers. This implies that there is a discrepancy between their production and perception. Taken together, the results point to the conclusion that the merging process is still in progress. Nevertheless, the perception data was based on a forced-choice AX discrimination task carried out in time-pressed experimental settings. It is quite possible that the subjects were focusing their attention to the subtle acoustic cues which they normally may not notice in naturalistic and noisy environment. Their performance in the perception experiment could be better compared with their everyday discrimination of the merging tone pairs. At any rate, these subjects should be regarded as incomplete mergers. Therefore, not only are the tone pairs still merging in the language as a whole, they are also still merging within individual speakers.

Again, the adult data on the merging tone pairs corresponds well with acquisition data. Both monolingual and bilingual children easily mixed up the two rising tones (T2/T5) and the two level tones (T3/T6) [11, 12]. Non-native (Thai and Filipino) speakers learning Cantonese also found these two tone pairs particularly problematic in both production and perception [13]. These data suggests that language learning, particularly by children, can be another likely source of sound change paralleling the change in adult patterns. Both processes may stem from the same underlying physical phonetic causes [14]: subtle differences in a narrow pitch range.

Only some preliminary perception data on monosyllables is presented in this paper. More detailed analyses of the monosyllabic data and data on non-linguistic pure tones will shed further light on the process of tone merging in Hong Kong Cantonese. It is particularly worthwhile to investigate the relationship between production and perception of the potential mergers. The results on the merging tones contribute to the bigger picture of sound changes happening in modern Hong Kong Cantonese phonology.

5. Acknowledgements

We thank Dr. S. D. Chan for the use of the Segmentation Corpus supported by a grant from the UGC of Hong Kong to Dr. Y. S. Cheung. The present project is supported by the research fund by the Department of Linguistics and Modern Languages, Chinese University of Hong Kong to the first author.

6. References

[1] Bauer, R. S. and Benedict, P. K., *Modern Cantonese Phonology*. Berlin, New York: Mouton de Gruyter, 1997.

[2] Zee, E., "Chinese (Hong Kong Cantonese)", in *Handbook of the International Phonetic Association*, 58-60, Cambridge University Press, 1999.

[3] Vance, T. J., "Tonal distinctions in Cantonese", *Phonetica,* 34:93-107, 1977.

[4] Khouw, E. and Ciocca, V., "Perceptual correlates of Cantonese tones", *Journal of Phonetics,* 35:104-117, 2007.

[5] Kei, J., Smyth, V., So, L. K. H., Lau, C. C. and 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.

[6] Bauer, R., Cheung, K. H. and Cheung, P. M., "Variation and merger of the rising tones in Hong Kong Cantonese", *Language Variation and Change,* 15:211-225, 2003.

[7] Varley, R. and So, L., "Age effect in tonal comprehension in Cantonese", *Journal of Chinese Linguistics,* 23:76-98, 1995.

[8] Mok, P.,. and Wong, P., "Production of the merging tones in Hong Kong Cantonese: preliminary data on monosyllables", *Speech Prosody 2010*, 2010.

[9] So, L., and Dodd, B., "The acquisition of phonology by Cantonese-speaking children", *Journal of Child Language,* 22:473-495, 1995.

[10] Ohala, J., "The listener as a source of sound change", in C. Masek, R. Hendrick, and M. Miller [Eds], Papers from the Parasession on Language and Behavior, 178-203, 1981.

[11] Ciocca, V., and Lui, J., "The development of lexical tone perception in Cantonese", *Journal of Multilingual Communication Disorders,* 1:141-147, 2003.

[12] Law, C. W., *Tonal characteristics of early English-Cantonese bilinguals.* MA thesis. University of Hong Kong, 2006.

[13] Chung, V., *Perception and Production of Cantonese tones by Thai and Filipino Students in Hong Kong Secondary School.* MA thesis. Chinese University of Hong Kong, 2009.

[14] Greenlee, M. and Ohala, J., "Phonetically motivated parallels between child phonology and historical sound change", *Language Sciences*, 2:283-308, 1980.