Production and perception of the rising tones in Hong Kong Cantonese

Peggy Pik-Ki Mok & Peggy Wai-Yi Wong

Chinese University of Hong Kong
peggymok@cuhk.edu.hk

Abstract

Cantonese has six lexical tones (T), but the two rising tones T2 (high-rising) and T5 (low-rising) appear to be merging in Hong Kong Cantonese. Both production and perception experiments were carried out using monosyllables. Results show that the potential mergers generally performed less well than the control group in tone perception, with a lower accuracy rate and longer reaction time, but they could still distinguish the T2/T5 pair with above 90% accuracy. Both groups found the T2/T5 pair more difficult to distinguish than other tone pairs. The production data show that while the potential mergers still have two rising tone categories, examples of both categorical and intermediate tone merge can be observed. There is also much individual difference in tone merge. The results indicate that the merging processing of the two rising tones is still in progress in the language as a whole and in individual speakers. Possible reasons for these patterns are discussed.

1. Introduction

Cantonese has a complex tone system. There are six contrastive lexical tones (T1 to T6) and three allotones (the 'entering tones', T7 to T9) which are shorter versions of T1, T3 and T6 in syllables ending with an unreleased final stop consonant (/p t k/). Each syllable, even function words, carries a distinct lexical tone [1]. 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. Tone 1 is distinguished from the other five tones by being at the top of the speaker's normal pitch range. It is also perceptually more salient than the other tones. The 'tonal space' is very crowded in the lower pitch range. Four tones (T2, T4, T5, T6) share a similar starting pitch level. Several tone pairs are particularly similar. The two rising tones T2 and T5 have the same starting point, but one (T2) rises to a higher pitch level while the other (T5) only rises to a mid pitch level. The two level tones T3 and T6 have a much smaller pitch difference (around 30 Hz for the female speaker featured in Figure 1) than that of T1 and T3. 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 could be confusable, especially when they are produced in isolation.

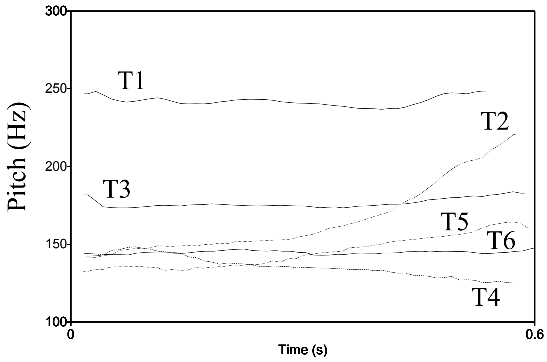

Figure 1. F0 traces of the six lexical tones.

The Cantonese tone system in Hong Kong described above is undergoing changes in recent years in that some young speakers no longer distinguish some of the six tones in their speech, most notably T2 and T5. 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. [2] studied Cantonese tone production by 56 speakers acoustically, without focusing on tone change. 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. [3] investigated the above phenomenon of T2/T5 confusion 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. In addition to production, Cantonese speakers also often confused the two rising tones perceptually [4]. A recent study [5] using 15 speakers reported that four speakers appear to be merging the two rising tones in perception with the same pattern: T5 merging into T2. In production, however, some of them merged T2 into T5 while others merged T5 into T2. Taken together, these studies 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.

[2, 3, 5] show that their speakers who merged the two rising tones did not merge all words with the two target tones. They were still able to differentiate some T2/T5 words. It

shows that these speakers probably still have two separate rising tone categories, although one category is shrinking. It is unclear what factors contribute to such phenomenon. Yiu [5] suggested that lexical diffusion may explain why only some words are affected. As reasonable as this argument may be, it is hard to verify empirically. Another possible factor which can affect the merging pattern may be word frequency. High frequency words are shown to be more susceptible to surface variations and are often more reduced in speech [e.g. 6, 7]. Zhao and Jurafsky [8] showed that word frequency can affect tonal production in Cantonese. Low frequency words (i.e. less familiar words) are hyperarticulated and are produced with a higher pitch. Low frequency words also have a more expanded pitch range than that of high frequency words, i.e. the tones of less familiar words are more dispersed in the tonal space. It is conceivable that more tone mergers may be found in high frequency words (i.e. familiar words) since they have a more compressed pitch range, and are more susceptible to surface variations and sound change.

In addition, the merging speakers were considered making 'production errors' by Kei et al. [2], while Bauer et al. [3] investigated only two merging speakers. The two studies also only examined the production of T2/T5. It is unclear how well their speakers who appeared to merge T2/T5 can distinguish the two rising tones in perception. Yiu [5] found a discrepancy between the production and perception of T2/T5 in her speakers. Besides, all three studies found the merging speakers serendipitously. Clearly, more studies targeting the merging speakers are needed to investigate the merging patterns of T2/T5 in Cantonese more thoroughly. Our study investigates both the production and perception of the T2/T5 merger in Hong Kong Cantonese, as well as other potential tone mergers (T3/T6, T4/T6). Monosyllabic and disyllabic words as well as non-linguistic pure tones were used. In this paper, only preliminary data on monosyllabic words of T2/T5 is presented. Analysis of other data is currently underway.

2. Method

2.1. Subjects

Since this study investigates the production and perception of tone mergers, it is important to ensure that we use speakers who do merge the tones. In order to recruit these speakers, a small screening test was conducted. Each potential participant was recorded reading a word list with 18 monosyllabic 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. 17 potential mergers were recruited.

Table 2. Number of potential tone mergers

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

Table 2 shows the number of recruited speakers who showed signs of merging different tones. The numbers are not balanced because it was quite difficult to locate these potential mergers. The recruited speakers participated in both production and perception experiments, except one speaker in the T2/T5 group who only participated in the production experiment. An additional 11 speakers who clearly distinguish all six tones were used as a control group for the perception experiment only. Thus, there were 27 subjects in total participating in the perception experiment, but only 17 subjects in the production experiment. The speakers were undergraduate students in the Chinese University of Hong Kong, aged between 18 and 22, with no history of hearing problems. They were paid to participate in the experiment.

2.2. Materials

For the production materials, both monosyllabic and disyllabic words were selected from an electronic database of around 33,000 Cantonese word types extracted from a 1.7 million character corpus of Hong Kong newspapers (details see [9]). The database is part of a larger Segmentation Corpus. It is a corpus of segmented Chinese texts, including Mandarin newspapers from both the PRC and Taiwan. The three databases were created using word-segmentation criteria developed by researchers at the Chinese Language Centre and Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University. These criteria were intended to be applicable to texts in all three varieties. For this study, we used the wordlist proper of the Cantonese database. It is a file containing a separate entry for each word type identified using the segmentation criteria. Each entry has three fields: the orthographic form(s), the pronunciation(s) in Jyutping, and the token frequency in the segmented newspaper corpus.

For each word entry, the log frequency of the token frequency was taken. For monosyllabic words, 6 high frequency words and 6 low frequency words were selected for each of the 6 tones, which resulted in 72 words (6 words × 6 tones × 2 frequencies). Syllables comprised of all sonorants were chosen as far as possible for each word. Where not available, extra monosyllabic words were added to the tone set. Syllables with a stop coda were not selected. As a result, 12 extra words were added, which yielded a total of 84 words, including the all-sonorant targets. The monosyllabic words were embedded in a short carrier phrase: [ŋɔ^{23} tok^{2} ___ tsi22] 'I read the word __'.

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 (see below). 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 17 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 experiment. The production experiment was divided into two parts. The subjects were recorded reading the monosyllabic materials first before the disyllabic materials, with a rest between the two sessions and short breaks within each session. Two randomised lists of the materials were used for counterbalancing. 9 subjects read list one and 8 subjects read list two. Three repetitions of the materials were recorded. Before the actual recording, the subjects practised by reading the materials for as long as they liked. The recording took place in a sound-treated booth in the Language Acquisition Laboratory at the Chinese University of Hong Kong. Their speech was recorded directly onto disk with a sampling rate of 22050 using Praat via a condenser microphone placed approximately 20 cm away from the subjects. They were recorded reading the materials with a normal speech rate.

The F0 values of each target syllable were automatically tracked using a Praat script. The beginning and end of periodicity of the target syllable (if the whole syllable consists of sonorants) or the vowel (if the syllable has a non-sonorous initial consonant) were marked manually. Ten equidistant measurement points were taken between these two marks, so the data were time-normalised for each tone. Since the automatic tracking program sometimes yielded anomalous values or failed to yield any F0 values for the first and last measuring points (i.e. the beginning and end of periodicity) because of the perturbation from initial consonants or creakiness, data from these two measuring points were excluded from analysis. Data from the other eight measuring points were averaged among tokens for each speaker. In addition, duration of the target syllables was also measured.

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 one?' 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 control subjects. 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 could 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 unclear what causes such pattern.

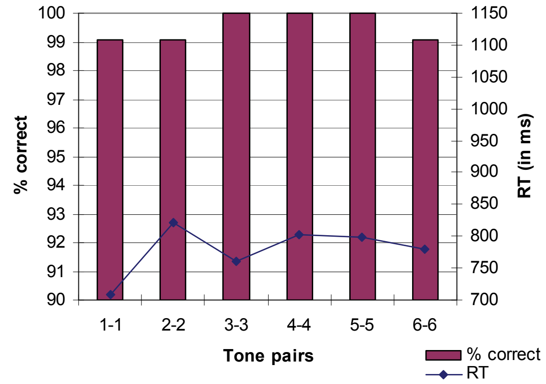

Figure 2. *Accuracy rate and reaction time for the AA pairs by the control group.*

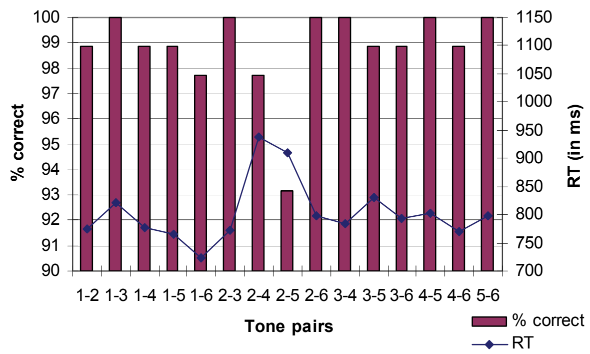

Figure 3. *Accuracy rate and reaction time for the AB pairs collapsed across sequences by the control group.*

A central question is whether the potential T2/T5 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 of the potential mergers. In general, the potential mergers had a lower accuracy rate and longer reaction time than the control group for all the AA pairs, indicating that they had more difficulty in distinguishing the tone pairs. Similar to the control group, they also found the T1/T1 pair easier to distinguish (with the shortest reaction time and a relatively high accuracy rate). The two rising tone pairs (T2/T2, T5/T5) appear to be of similar degree of difficulty for the potential mergers.

For the AB pairs, again, in general, the potential mergers had lower accuracy rates and longer reaction time than the control subjects. Similar to the control group, the potential mergers appear to perform better in tone pairs with T1. Although they had a higher accuracy rate for some tone pairs than the control group, they had a longer reaction time. It is interesting to note that the potential mergers were 100% correct for the T2/T5 pairs, only with a very long reaction time (1104 ms), which is the longest among all tone pairs. This indicates that perceptually the potential mergers could still distinguish the two rising tones, but they had to do so very carefully. Besides, they also found the other two similar tone pairs (T3/T6, T4/T6) difficult to distinguish, resulting in the lowest accuracy rates and relatively long reaction time. Thus, compared to the control group, the potential mergers had a poorer perception of tones in general, and especially for the pairs which are physically similar.

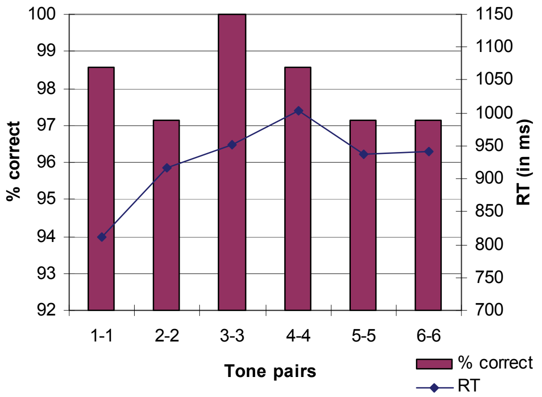

Figure 4. *Accuracy rate and reaction time for the AA pairs by the potential T2/T5 mergers.*

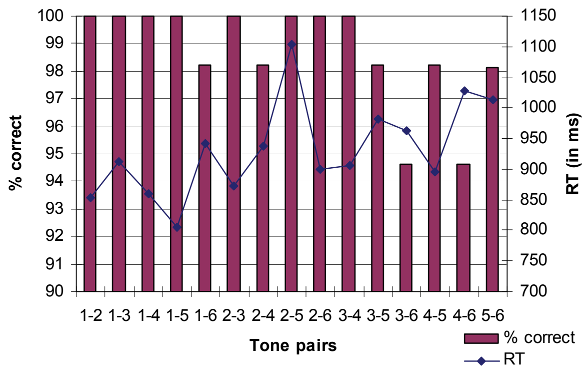

Figure 5. *Accuracy rate and reaction time for the AB pairs collapsed across sequences by potential T2/T5 mergers.*

The next question is whether the potential mergers distinguish the two rising tones in their production. Figure 6 shows the average F0 values separately for the high and low frequency words in T2 and T5 produced by them. The overall patterns seem to suggest that there was no merging for these potential mergers, as the two tones are still clearly separated in the second half of the tones. Also, word frequency did not appear to affect the F0 values of the tones. The F0 contours of high and low frequency words generally parallel each other quite closely in Figure 6. Nevertheless, it does not mean that there is no frequency effect at all. Figure 7 shows the duration of the target syllables in different tones averaged across all subjects according to word frequency. It is obvious that high frequency words were produced with a shorter duration across tones. Although word frequency did not affect the realisation of the tones, it has a robust effect on syllable duration.

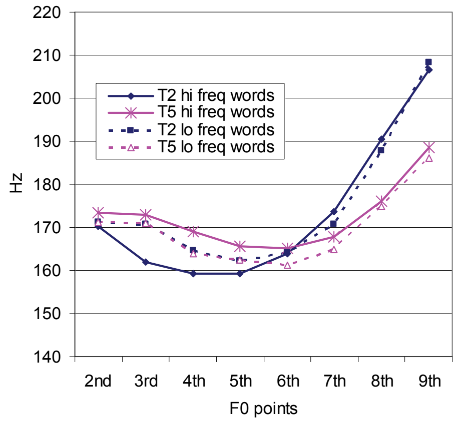

Figure 6. *Average F0 values by T2/T5 potential mergers.*

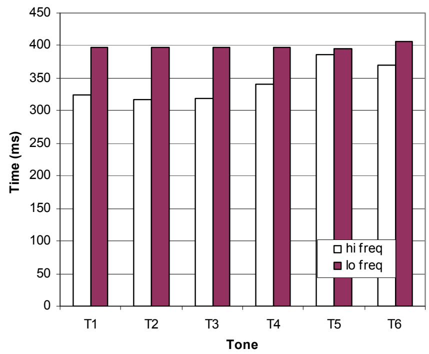

Figure 7. *Duration of the target syllables in six tones.*

In fact, the above data averaged across speakers has concealed the real picture of the merging process since there are quite a lot of individual differences. To further investigate, we selected 2 subjects who appeared to confuse the target tones more often than others, and studied them in greater details. These 2 subjects happened to be both female.

Each token of the T2 and T5 words produced by the 2 subjects were auditorily checked by the second author. Table 3 shows the distribution of the target tones T2 and T5 produced by them. Overall, the data do not show a massive merge between the two tones for both subjects, although merging does occur to different degrees. For both subjects, T2 was produced as T5 more often than the other way around. It is interesting to note that merging does not necessarily happen categorically. Subject T produced the target tones somewhere in between the two tones 6% of the time (5 tokens / 81 tokens). Also, T5 (mid-rising) could be produced as a level tone somewhere between the mid- and the low-level. (See subject J.) Frequency effect can be seen from the results in that in general, high frequency words are comparatively more susceptible to merging, except for T2 for subject J. These observed patterns are indicative of the general picture as data from other subjects also show similar patterns. Nevertheless, more detailed inspection of individual data checked by more listeners is needed to further explore the merging patterns.

Table 3. Distribution of tonal production for T2 and T5 for two female speakers.

	Tone	Subj T	%	Subj J	%
Hi freq (n=27)	T2=T2	21	77.8	26	96.3
	T2->T5	5	18.5	1	3.7
	T2->T2/T5*	1	3.7	0	0
Lo freq (n=18)	T2=T2	13	72.2	15	83.3
	T2->T5	4	22.2	3	16.7
	T2->T2/T5	1	5.56	0	0
Hi freq (n=18)	T5=T5	15	83.3	14	77.8
	T5->T2	0	0	2	11.1
	T5->T2/T5	3	16.7	0	0
	T5->T3/T6	0	0	2	11.1
Lo freq (n=18)	T5=T5	17	94.4	18	100
	T5->T2	1	5.56	0	0
	T5->T2/T5	0	0	0	0

* T# -> T# / T# means that the target tone was heard to be produced between two tones.

4. Discussion

This paper investigates both the production and perception of the two rising tones T2/T5 in Hong Kong Cantonese. Our production data show that the process of tone merging in Hong Kong Cantonese is underway. However, it seems to be more complicated than what have been documented. Like the subjects in Kei et al. [2], Bauer et al.'s [3] and Yiu's [5] studies, speakers who merge tones do not merge all words that have the target tones. The speakers in our study actually retained quite a lot of words in the target tones. Also, they do not merge tones uni-directionally. That is, it is not the case that one tone is always produced as another tone. Rather, the target tone could be produced as more than one other tonal category, e.g. T2 words can be produced as T5, or realised somewhere between the two tones. In terms of tonal inventory, while the six distinct tones in Hong Kong Cantonese are retained among the individual speakers we studied in greater details, novel tones that lie somewhere along the categorical boundary have emerged. This is interesting because it implies that the direction of the tone merge process in Hong Kong Cantonese could proceed in at least two ways: shrinking and expanding of the tonal inventory.

While word frequency having an effect on the duration of word production might not be surprising, how word frequency is related to tone merge requires more investigation. From our detailed study of the 2 speakers, high frequency words show a higher probability of tone merge than low frequency words in general. However, there is also exception (see T2 for subject J). More detailed investigation is needed to examine the effects of word frequency on the production of tone merge.

Perceptually, the results reveal both similarities and differences between the control group and the potential merger group. 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 [10]. 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 [11]. The acoustic similarity between the two rising tones renders them particularly susceptible to sound change.

Although the potential merger group 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 much 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 the six tones are retained in their tonal inventory, despite them being potential mergers. Taken together, the results point to the conclusion that the merging process is still in progress. Merging in production seems to progress faster than merging in perception. 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 pair. 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 pair corresponds well with acquisition data. Both monolingual and bilingual children easily mixed up the two rising tones (T2/T5) [12, 13]. Non-native (Thai and Filipino) speakers learning Cantonese also found these two tones particularly problematic in both production and perception [14]. 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 [15]: subtle differences in a narrow pitch range.

Only some preliminary data on monosyllables is presented in this paper. Our future work is to study all the individual speakers in detail to get a more comprehensive picture of individual differences, and to examine how word frequency plays out in the process of tone merge. We would also analyse the subjects' production of disyllabic words, perception of non-linguistic pure tones and both the production and perception of minimal pairs. It is particularly worthwhile to investigate the relationship between production and perception of the potential mergers. The results will shed further light on the process of tone merging in Hong Kong Cantonese, and contribute to the larger 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, which was supported by a grant from the University Grants Committee (UGC) of Hong Kong to Dr. Y. S. Cheung. The present project is supported by a research fund of the Department of Linguistics and Modern Languages, Chinese University of Hong Kong, awarded to the first author.

6. References

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

[2] Kei, J., Smyth, V., So, L. K. H., Lau, C. C. and Capell, K., 2002. 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.

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

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

[5] Yiu, C. 2009. A preliminary study on the change of rising tones in Hong Kong Cantonese: an experimental study. *Language and Linguistics*, 10:269-291.

[6] Jurafsky, D., Bell, A., Gregory, M., Raymond, W., 2001. Probabilistic relations between words: Evidence from reduction in lexical production, in J. Bybee & P. Hopper [ed], Frequency and the Emergence of Linguistic Structure, Amsterdam: Benjamins.

[7] Wright, R., 2004. Factors of lexical competition in vowel articulation. in J. Local, R. Ogden & R. Temple [eds.], *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.

[8] Zhao, Y. and Jurafsky, D., 2007. The effect of lexical frequency on tone production, *Proceedings of ICPhS*, 477-480.

[9] Chan, S. D. and Tang, Z. X., 1990. Quantitative Analysis of Lexical Distribution in Different Chinese Communities in the 1990's, *Yuyan Wenzi Yingyong* [Applied Linguistics], 3:10-18.

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

[11] Ohala, J., 1981. 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, Chicago Ling. Soc..

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

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

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

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