Production 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

Cantonese has six tones (T), but some appear to be merging in Hong Kong Cantonese, namely, the two rising tones T2/T5, the two level tones T3/T6, and the low falling and low level tones T4/T6. 17 potential mergers participated in a production experiment. This paper reports data from the production of high and low frequency monosyllabic words. F0 values were measured in the subjects' production. Results show that there is much individual difference in tone merge. While there is tone merging into various tonal categories, the six tones are kept intact. More interestingly, novel intermediate tones emerge.

Index Terms: Cantonese, tones, mergers, production

1. Introduction

Cantonese is a tone language with 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]. 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. Some tones also differ in duration, but they were produced with a similar duration in Figure 1 for easy comparison. 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 the same 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 who produced the tones 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.

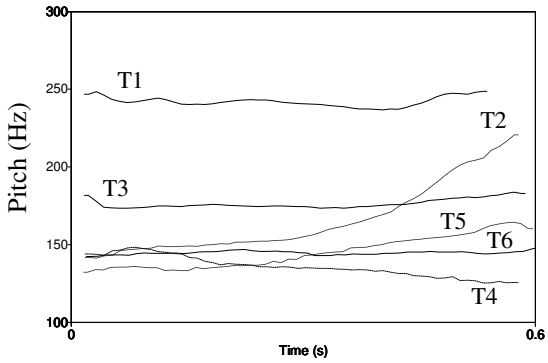


Figure 1. F0 traces of the six lexical tones.

The production and perception of the complex tone system in Cantonese attracts a lot of research attention. 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. Kei et al. [2] 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. [3] replicated [2] 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 [4]. 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.

In addition, [2, 3] show that their speakers who merged the two rising tones did not merge all words with the two target tones. They were able to differentiate a small number of T2/T5 words. It shows that these speakers probably still have two separate rising tone categories, although one category is shrinking rapidly. One possible factor contributing to this phenomenon 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. 5, 6]. Zhao and Jurafsky [7] 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. Although they did not investigate tone mergers, 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.

[2, 3] only documented the production of T2/T5 mergers. It is unclear how well their speakers could distinguish the two rising tones in perception. Thus far, there is also 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 (T3) and the low level tones (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. Word frequency of the target monosyllabic and disyllabic words is included as a factor in the production experiment. In this paper, only preliminary production data on monosyllabic words is presented. Perception 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 production of tone mergers, it is important to ensure that we use 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 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. 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

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

2.3. Procedures

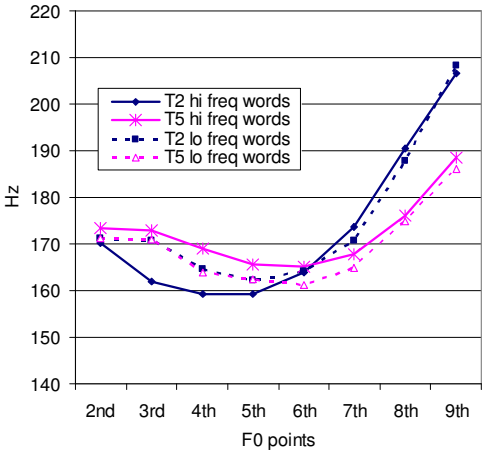
The 17 potential merger subjects participated in the production experiment first. They were invited back for the perception experiment at least two weeks after the production experiment. The production experiment was divided into two parts. The subjects were recorded reading the monosyllabic materials 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.

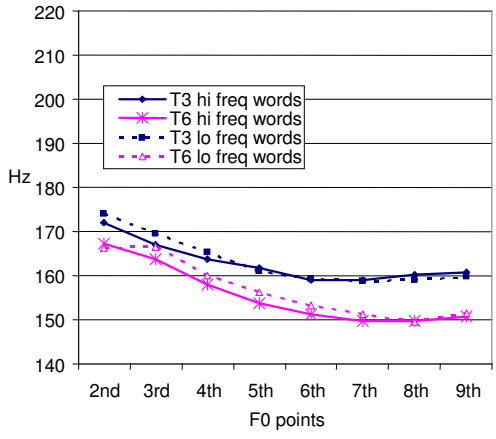
3. Results

The main research question is whether the potential mergers distinguish the merging tone pairs in their production. Figures 2 to 4 show the average F0 values separately for the high and low frequency words of the three groups of potential mergers (T2/T5, T3/T6, T4/T6). The table under each Figure shows the t-tests results comparing the two tones separately for high and low frequency words (an asterisk means $p < 0.05$). Since the second half of T4 (21) reaches the lowest pitch level, it often becomes very creaky, thus resulting in many anomalous or undefined F0 values. Therefore, only data from the second to the sixth measuring points of T4 are shown in Figure 4.



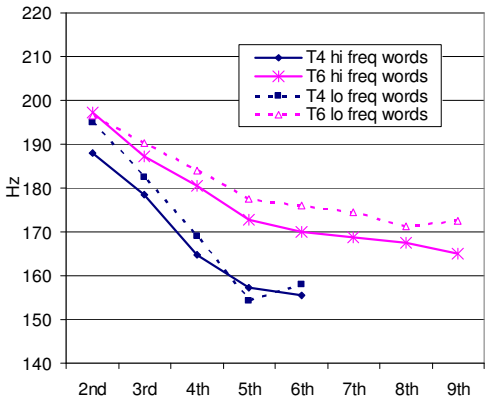
Word Frequency	2nd	3rd	4th	5th	6th	7th	8th	9th
hi	--	*	*	--	--	*	*	*
lo	--	--	--	--	--	*	*	*

Figure 2. Average F0 values and t-test results by potential T2/T5 mergers.



Word Frequency	2nd	3rd	4th	5th	6th	7th	8th	9th
hi	--	*	*	*	*	*	*	*
lo	--	*	*	*	*	*	*	*

Figure 3. Average F0 values and t-t-test results by potential T3/T6 mergers.



Word Frequency	2nd	3rd	4th	5th	6th	7th	8th	9th
hi	--	--	*	*	*			
lo	--	--	--	--	--			

Figure 4. Average F0 values and t-tests results by potential T4/T6 mergers.

The overall patterns seem to suggest that there was no tone merging for these potential merger groups, as the tones are significantly different. 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 Figures 2 to 4. Nevertheless, it does not mean that there is no frequency effect at all. Figure 5 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.

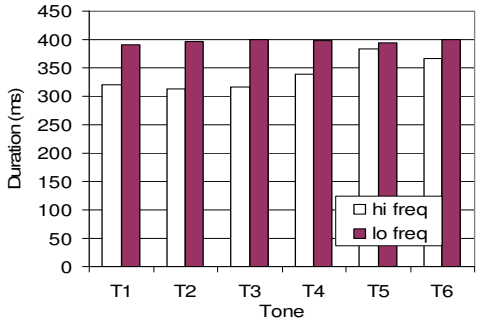


Figure 5. Duration of the target syllables in six tones.

In fact, the averaged data above has concealed the real picture of the merging process since there are quite a lot of individual differences. To further investigate, we selected 4 subjects, 2 as T2/T5 and 2 as T3/T6 potential mergers, who seemed to confuse the target tones more often than others, and studied them in greater details. These 4 subjects happened to be all female.

Each token of the 4 subjects were auditorily checked by the second author and an independent transcriber who is a native Cantonese speaker clearly distinguishing all six tones. Any discrepancy between the two transcribers was further checked by the first author. The majority decision was adopted for these tokens. Table 3 shows the distribution of the target tones T2 and T5 produced by the two subjects. 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 that frequency seems to have opposite effects on T2 and T5. More low frequency words are affected for T2, while more high frequency words are affected for T5.

Table 3. Distribution of T2 and T5 tokens produced by two female speakers (%).

	Tone	Subj T	Subj J
Hi freq	T2=T2	77.8	100
	T2->T5	22.2	0
Lo freq	T2=T2	66.7	83.3
	T2->T5	33.3	16.7
Hi freq	T5=T5	81.8	90
	T5->T2	18.2	10
Lo freq	T5=T5	100	100
	T5->T2	0	0

Table 4 shows the distribution of the target tones T3 and T6 produced by two speakers. Much individual difference can be observed. While subject J seems to be more conservative in merging, subject S is a more advanced merger. Note that subject S produced the low frequency T3 words as T6 (82.4%) even more often than as T3 (17.6%) itself. In fact, for both speakers, low frequency words were produced as other tonal categories more often than their high frequency counterparts for both T3 and T6. The results contradict the expectation that low frequency words are less susceptible to change.

Table 4. Distribution of T3 and T6 tokens produced by two female speakers (%).

	Tone	Subj S	Subj J
Hi freq	T3=T3	66.7	94.4
	T3->T6	33.3	5.6
Lo freq	T3=T3	17.6	88.9
	T3->T6	82.4	11.1
Hi freq	T6=T6	77.3	70.8
	T6->T3	22.7	29.2
Lo freq	T6=T6	66.7	38.9
	T6->T3	33.3	61.1

4. Discussion

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 has been documented so far. Like the subjects in Bauer [3] and Kei et al's [2] studies, speakers who merge tones do not merge all words that have the target tones. For simplicity sake, Table 3 and Table 4 only show the majority decision among the three transcribers. In fact, there are some tokens which appear to be produced between the two canonical tones. For example, some T2 tokens were heard to be between T2 and T5. The three native transcribers can also differ in their perception of these tokens. Such interesting cases show that the potential mergers do not merge tones unidirectionally. 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. So far, our data suggests that T1 is not affected. 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 or expanding of the tonal inventory. Further investigation is needed to examine these interesting cases more carefully.

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 further study of the 4 speakers, there does not seem to be a consistent word frequency effect on tone merging. Contrary to expectation, low frequency words appear to be affected more than their high frequency counterparts.

One limitation of this study is the use of a constant carrier phrase for all monosyllabic words, which was chosen for better experimental control. However, it might have geared the subjects towards more conscious tonal articulation similar to citation form. This may have reduced the likelihood of tone mergers in their production.

Our future work is to study all the individual speakers in detail to get a more comprehensive picture of what individual differences there are, and how word frequency plays out in the process of tone merging. We would also analyse the subjects' production of disyllabic words and the production and perception of minimal pairs. Lastly, how production is related to perception in the tone merge process would also be investigated. The results of tone mergers can shed more light on 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.

6. References

[1] Bauer, R. S. and 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. 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.

[3] 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.

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

[5] Jurafsky, D., Bell, A., Gregory, M., Raymond, W., "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. 2001.

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

[7] Zhao, Y. and Jurafsky, D., "The Effect of Lexical Frequency on Tone Production", Proceedings of ICPhS, 477-480, 2007.

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

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