ACOUSTIC ANALYSIS OF THE NEW RISING TONE IN HONG KONG CANTONESE

Roxana S.Y. Fung & Cathy S.P. Wong

Hong Kong Polytechnic University, Hong Kong
ctsyfung@polyu.edu.hk; egcathyw@polyu.edu.hk

ABSTRACT

Hong Kong Cantonese has six contrastive tones. A previous study, however, has indentified a group of speakers who merged the high rising tone with the low rising tone in both production and perception. This study investigated the acoustic characteristics of the rising tones produced by three mergers and three non-mergers identified in the previous study, and consequently, characterized the mechanism of this tone merger. Through the F0 measurements of ten time points of the vocalic segments of 140 syllables, the two rising tones were found merging into one tonal category with only one phonetic realization. The height of the new rising tone resembled that of the canonical high rising tone; however, the slope of the new rising tone resembled that of the canonical low rising tone. A bidirectional mechanism was adopted in this sound change by the mergers.

Keywords: Cantonese, tones, mergers, acoustic analysis

1. INTRODUCTION

Cantonese stands out from other tone languages in the world by having a rich system of tonal contrast. There are six contrastive tones in standard Cantonese, namely high level, high rising, mid level, extra-low level, low rising and low level tone. However, this highly complex system is in the process of merging in Hong Kong Cantonese (HKC). Among the alleged tone mergers, the conflation of the high rising tone (HR) and the low rising tone (LR) has attracted most of the attention. According to traditional Cantonese phonology, the HR differs from the LR by having different onset and endpoint F0 values. The HR has been transcribed as (35) whereas the LR as (23) or even (13). The HR has considered to belong to the high register whereas the LR the low register [3, 4]. In recent years, acoustic studies on tone production have revealed that the two rising tones in contemporary HKC begin at around the same onset but differ only in their endpoint values [1]. The LR is then represented as (23). In short, the two rising tones are believed to be distinguished mainly by its F0 slope: the HR is marked by a very steep pitch gradient while the LR only a mild one. On the perception side, [5] also found that the magnitude of F0 change (i.e. the F0 slope) was an important perceptual correlate in distinguishing HR from LR in HKC. The acoustic similarity of the two rising tones may render them particularly susceptible to merging.

Previous studies (such as [2, 7, 8]) have identified some individual speakers who merged the two tones. However, conclusive evidence supporting the tone merger in the community of HKC is only found in a recent study conducted by the authors. In the study, discrimination and production tasks were administered to 120 randomly chosen subjects of age ranging from 20-58 years. It was found that the average discrimination score and production score of the two rising tones were significantly lowered than that of the other tone pairs. As for the direction of the production errors, the overall picture was quite complicated. Speakers of different age groups demonstrated different error patterns. In the oldest group, the HR was well preserved whereas the LR targets were pronounced as HR. In the younger groups, some HR targets were pronounced as LR whereas some LR targets as HR. Intermediate forms were also attested. In some cases, the three raters of the study failed to agree on each other's judgment. Thus, the phonetic realization of the new rising tone category remains puzzling. Are there two different phonetic forms of the category or just one? What are the phonetic features of the new form(s) as compared with the canonical tones? By what route do the two tones become one? This paper takes on the above questions by examining the acoustic characteristics of the rising tones produced by the potential mergers and the non-mergers. Due to the page limit, this paper reports the preliminary findings on the speech outputs of the speakers from the youngest age group only.

2. METHODOLOGY

2.1. Subjects

The data of this paper were based on the speech samples produced by six subjects of age ranging from 21 to 24 years collected in the previous study. The subjects formed two groups: the mergers and the non-mergers. Each group comprised one male and two female speakers. The subjects were selected based on their production and perception scores in the previous study. The backgrounds of the subjects are listed in Table 1.

Table 1: Backgrounds of the subjects.

		Percent Correct in Production			Percent Correct in Perception	
Subject No.	Sex	HR	LR	Mean	HR/LR Pair	Overall
Non-Mergers						
25	M	100	100	100	100	91.7
36	F	100	100	100	87.5	99.4
16	F	100	100	100	87.5	99.4
Mergers						
30	F	91.7	41.7	66.7	12.5	92.9
21	F	100	33.3	66.65	25	96.4
27	M	100	33.3	66.65	25	96.4

The selection criteria were not very straightforward since dissociation of perception and production was commonly found among the subjects. In other words, a subject may be good at producing the two rising tones but bad at discriminating them in perception or vice versa. Since this study focused on the production side, the mean production score of the two rising tones was taken as the primary indicator, the perception score of this tone pair the secondary indicator, and the perception score for all tone pairs the tertiary indicator. The non-merger group was comprised of subjects who scored 100% correct in the production of all six contrastive tones and over 85% correct in the perception of HR/LR contrast. The merger group was comprised of subjects whose mean scores of the production of HR and LR tones were less than 67% and their scores of the perception of HR/LR contrast were less than 30%.

2.2. Materials

Twelve different words of eight different CV syllables, six HR and six LR (see Table 2), were selected as the targets. It was necessary to include a wide variety of CV structures in the study since there were great individual variations on which syllables or which words of the two rising tones to merge among the merger speakers. The stimuli were generated by embedding the 12 different words into two sentence carriers:

Table 2: Speech materials.

Full Tone Set				
	/fu/	/sɛ/	/si/	/ji/
HR	'bitter'	'write'	'history'	'chair'
LR	'woman'	'society'	'city'	'ear'
Deficient Tone Set				
	/ku/	/p^hɔ/	/ja/	/jɛ/
HR	'ancient'	'quite'	-	-
LR	-	-	'also'	'wild'

In the first sentence carrier, [ŋɔ^{23} ji^{21} ka^{55} tuk^{2}___tsi^{22}] 'I am now reading the ___character', the target word was embedded in the sentence internal position in between of two words of low level tones. In the second carrier, [ni^{55} kɔ^{33} tsi^{22} hai^{22}_____] 'This character is ____', the target word was embedded in the sentence final position after two words of low level tone.

2.3. Procedures

The subjects were asked to read out the sentences which were recorded at a sampling rate of 44.1 kHz. In total, 144 sentences (12 words x 2 sentence carriers x 6 subjects) were collected, but only 140 sentences were analyzed after discarding the four samples in which the target words were misrecognized by the subjects. The target syllables were then segmented from the sentence carriers. The F0 of each target syllable was exacted at 11 equal distant points using a Praat script. The initial position of the vocalic segment was set at the second upward-going zero-crossing of the voicing cycle. The final position was set to the maximum point of the rising trajectories near the final vocalic portion. The measured F0 values in Hz were normalized to reduce cross-speaker variation using the following T-value formula adopted by [9]. The formula, as shown in (1), included a multiplier of 5 to make the normalized F0 values comparable to the traditional five-level tone system.

$$(1)\quad Ti = [(\lg P_i - \lg P_{min})/(\lg P_{max} - \lg P_{min})]\text{x}5$$

where Pi was the pitch value at the percentage time point i, P_{max} was the highest pitch value and P_{min} was the lowest pitch value in the speech samples produced by a speaker.

A discriminant analysis using four parameters: F0 Contour, Endpoint F0, F0 slope and F0 Height, was performed to characterize the produced tones.

To reduce the influence of F0 perturbation caused by the voiceless obstruent initial of some syllables, the F0 value at the onset position was excluded from the calculation of the values of each parameter. The first parameter, 'F0 Contour' was the F0 trajectories of the rising tones based on the F0 (in T-value) measured at the ten time points. The second parameter, 'Endpoint F0', was the F0 (in T-value) at the 100% position of the vocalic segment. The third parameter, 'Slope of F0 rise', referred to the slope of the best fitted straight line that passed through the set of the F0 (in T-value) measured at the ten positions. It was the coefficient of the linear regression of the ten F0 measurements generated by SPSS. The fourth parameter, 'F0 Height', was defined as the mean of the F0 (in T-value) over the ten positions of the vocalic segment. The values for each parameter measured for each syllable were averaged for each speaker and then for each group. Kruskal-Wallis Test, a non-parametric one-way ANOVA, using speaker group as the independent condition was conducted individually for each parameter.

3. RESULTS

Figure 1 and Figure 2 display graphically the F0 contours of the two rising tones produced by the non-mergers and the mergers respectively. Table 3 displays the values of the other three parameters across group. An examination of each parameter should be in order.

3.1. F0 contour

In Figure 1, two distinctive tone contours are observed among the non-mergers. The two rising tones share similar starting point and remain overlapped at the first 30% of the tonal duration. They then depart each other at the latter portions of the vocalic segment: The HR moves to an endpoint much higher than that of the LR. On the contrary, the contours of the two rising tones produced by the mergers as shown in Figure 2 almost overlap each other for the whole vocalic segment. The two rising tones are visually non-distinctive. In fact, the graphical representations of the contour of the two tones are consistent with the quantitative measurements of the three other parameters as shown in Table 3. The mean values between the HR and LR of the merger group for each parameter were statistically insignificant. In other words, the two rising tones were completely merged into one new rising tone by the subjects in our merger group. Figure 3 displays the contours of the new rising tone and the two canonical tones produced by the non-mergers for comparison.

Figure 1: F0 trajectories of the rising tones produced by the non-mergers.

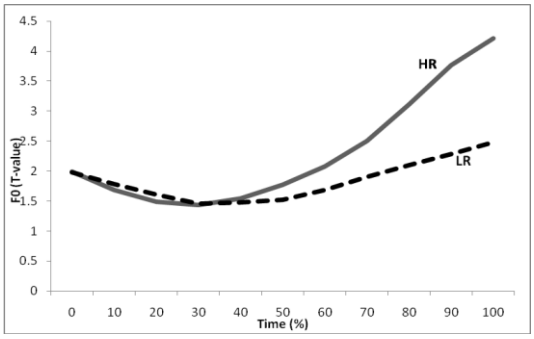


Figure 2: F0 trajactories of the rising tones produced by the mergers.

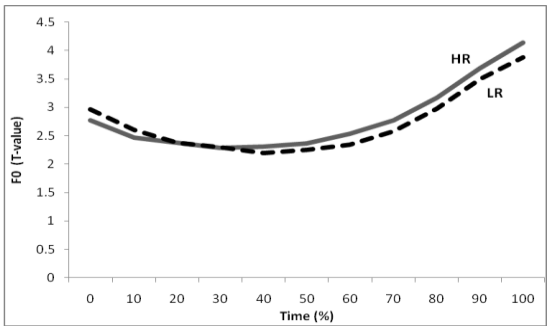
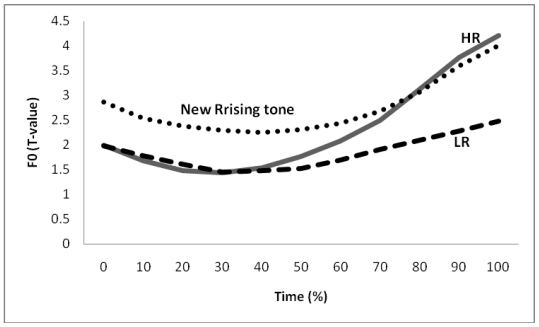


Figure 3: F0 trajactories of the new rising tone and the two cannoical rising tones.



In the following sections, we will compare the mean values of the three other parameters of the tones produced by the mergers to that of the canonical target tones produced by the non-mergers.

3.1. Endpoint F0

The difference between the endpoint F0 values of the HR between the mergers and the non-mergers was not significant. However, a significant sharp rise in the LR endpoint produced by the mergers was found, $\chi 2\ (1)=30.633$, $p < 0.001$.

Table 3: Mean values of the parameters of the two rising tones across groups.

	Endpoint F0 (T-value)		F0 Slope (T-value/s)		F0 Height (T-value)	
	HR	LR	HR	LR	HR	LR
non-mergers	4.21	2.49	2.36	1.83	0.31	0.10
mergers	4.14	3.88	2.78	2.67	0.19	0.15

3.2. F0 slope

A significant sharp decrease in the slope of HR produced by the mergers was attested, $\chi 2$ (1)=17.649, $p < 0.001$. On the contrary, there was no significant change in the slope of LR tones among the mergers.

3.3. F0 height

For the HR, there was a significant rise in the F0 height produced by the mergers, $\chi 2$ (1)=4.932, $p < 0.05$. For the LR, a significant sharp rise in the F0 was attested, $\chi 2$ (1)=12.845, $p < 0.001$. The results suggested that the heights of both rising tones among the mergers have been increased, but the height of LR has increased to a greater extent. As revealed by Figure 3, the increase in the average F0 height of HR is mainly caused by the change of the contour, whereas that of the LR by the raise of the whole register from low to high.

4. DISCUSSION

Our results suggested that two rising tones have been completely merged into one tonal category with one phonetic form among some young HKC speakers. It is a high register tone with high endpoint. However, the magnitude of F0 movement is relatively mild. As compared with the canonical tones, the new rising tone inherits the contour shape from the LR but the register from the HR. Our finding differs from the findings in previous works such as [2], in which the F0 of HR was found intact whereas the trajectory of LR collapsed with that of the HR. Our findings also reveal the mechanism of the tone merger. The following three major mechanisms of sound merger were proposed by [6]: (1) Merger by transfer: A unidirectional process in which words are transferred gradually from phonemic category to another. (2) Merger by approximation: A gradual approximation of two phonetic targets of two phonemes until they are non-distinct. Intermediates between the two original phonemes may be resulted. (3) Merger by expansion: The lexical constraints on the distribution of the two former phonemes are removed. The phonetic range of the new phonemic category is roughly equivalent to the union of the two categories that merged. Our findings show that the rising tone merger is not a unidirectional but a bidirectional process. It resembles the merger-by-approximation model in the sense that both former rising tones have undergone significant phonetic changes and approximate each other's phonetic forms. The HR retains its endpoint while reducing the slope, whereas the LR retains its slope while raising the F0 height. The two tonal variants then gradually become non-distinct giving rise to a new rising tone. However, the process is not exactly equal to the merger-by-approximation since the resulting tone does not show a simple arithmetic mean value intermediate between the two canonical tones but incorporates the phonetic features of both tones.

5. REFERENCES

[1] Bauer R.S, 1998. Hong Kong Cantonese tone contours. *Studies in Cantonese Linguis*tics, 1-33.

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

[3] Chao, Y.R. 1947. *Cantonese Primer*. New York: Greenwood.

[4] Cheung, S.H.-N. 1972. *Studies on Cantonese as Spoken in Hong Kong.* Hong Kong: The Chinese University of Hong Kong.

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

[6] Labov, W. 1994. *Principles of Linguistic Change: Internal Factors.* Oxford: Blackwell.

[7] Mok, P. P.-K., Wong, P. W.-Y. 2010a. Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. *Proc. Speech Prosody 2010*, 100986:1-4.

[8] Mok, P.P.-K., Wong, P. W.-Y. 2010b. Production of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. *Proc. Speech Prosody 2010*, 100916:1-4.

[9] Shi, F. 1986. Bejinhgua de shendial geju [The tonal pattern of Beijingese]. In Shi, F., Liao, R. (eds.) *Yuyin Conggao*. Beijing: Beijing Foreign Language Institute Press, 10-19.