GENERATIONAL DIFFERENCES IN PRODUCTION OF A TONAL CONTRAST IN HONG KONG CANTONESE

Bin Li, Yihan Guan

City University of Hong Kong
binli2@cityu.edu.hk

ABSTRACT

Hong Kong (HK) Cantonese contains rich tonal contrasts with some having been reported merging, which calls for empirical investigation across social dimensions. This study examines production of a contrast between two rising tones (the high rising tone T2 vs. the low rising tone T5) in HK Cantonese by native speakers in various age groups. We analysed tonal correlates of syllables contrasting the two tones in an utterance produced by 50 speakers. A change is observed in production of T2 but not T5, which appears affected significantly by age. Our youngest group has distinct ranges of pitch offsets of T2 compared with tones produced by speakers in their 50s and up, while the middle group's production shows an intermediate pattern. Moreover, gender seems to modulate the age factor to some extent.

Keywords: tone production, tone merger, Hong Kong Cantonese, age effect, gender difference

1. INTRODUCTION

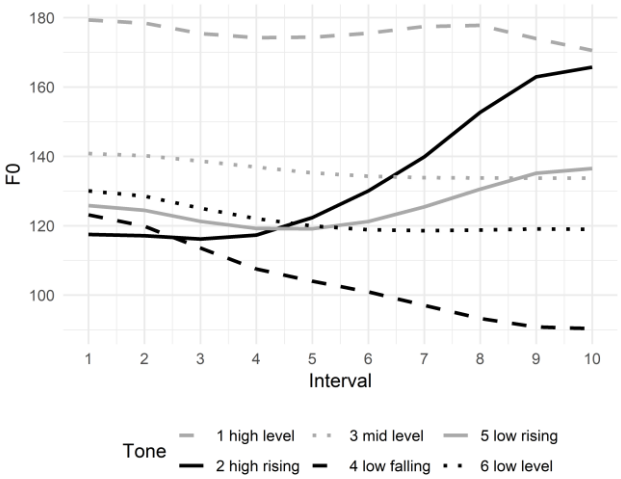
Hong Kong (HK) Cantonese has a rich repertoire of tone contrasts [1], as listed in Table 1.

Table 1: Illustration of Cantonese tones.

Tone	Name	5-point notation
1	high level tone	55
2	high rising tone	25 or 35
3	mid level tone	33
4	low falling tone	21
5	low rising tone	23 or 13
6	low level tone	22

Some tone pair or triplet, T2-T5, T3-T4-T6, are of comparable contours distributed in similar pitch contours (Figure 1). Theoretically speaking, the high number of tones in a packed tonal space provides motivation for tonal change such as tone merger [4]. In fact, merger of the high rising tone T2 and the low rising tone T5 has been observed among young speakers [4, 7, 13]. [4] investigated tones produced by native speakers aged 20-35 years old and proposed that T5 was approaching T2. [13] reported a similar tendency among young speakers whose production of T5 was most variable but remained distinct from T2, suggesting a preliminary tonal change. [7] claimed further that young HK Cantonese speakers have merged the two rising tones into a new category with a pitch offset as high as T2 and a contour slope as mild as T5.

Figure 1: Time-normalised F0 contours of six tones in HK Cantonese, extracted from recordings available on [8].



It is noted that tone merger reported earlier was only among the young generation, while production of other age groups is not examined. However, different linguistic behaviours of the young may be ambiguous as such behaviour could signal language change or the youth's non-conformity with social norms. Disambiguation, then, requires further investigation of tone production in longitudinal studies or at least among more age groups. Additionally, factors like gender may contribute to phonetic variations in HK Cantonese [3].

To identify the occurrence of any tonal change, an investigation on tone production across social groups is warranted. We then set out to answer whether there is any phonetic change in the production of T2 and T5 in HK Cantonese. If the answer is yes, then is the merger still in progress or completed? Is the pattern of change universal or different across social groups?

2. METHOD

All the data used in the current paper were speech recorded as part of student group projects in an undergraduate course. Different from [4, 7, 13], we did not pre-select speakers by the occurrence of tone merger. Our speakers were recruited by students on a voluntary basis. All participants gave written consent to have their speech recorded and used for the course and further analysis, and their demographic and linguistic information collected in a language background questionnaire. The total pool contains 172 valid talkers, but only 50 of them are included in this paper. The selection criteria are stated in Section 2.1 below.

2.1. Participants

Fifty speakers selected out of 172 were all ethnically Chinese, born and raised in Hong Kong. They all self-reported Cantonese as their native language and spoke no other native-like language than Cantonese and English. In addition, their parents spoke Cantonese as (one of) their daily language(s) in Hong Kong. The rest 122 that did not meet either of the above criteria were excluded from the analysis.

Among all 50 speakers, 33 were female and 17 were male. Their age ranged from 10-88 (M = 37, SD = 18): 13 speakers in the young group aged 25 and below, 21 in the middle group aged 26-50, and the rest 16 in the senior group were from 51-88.

2.2. Materials

To include all lexical tones in one short utterance, our speech elicitation used a meaningful phrase composed of 6 syllables each bearing a different lexical tone (as shown in Table 2).

Table 2: Transcription and gloss (CN: Cantonese).

CN	三	碗	細	牛	腩	麵
IPA	saːm	wun	sɐi	ŋɐu	nam	min
Tone	1	2	3	4	5	6
Gloss	three	bowl	small	beef	belly	noodle

The current study employed a phrase utterance for three reasons apart from the relative easiness as coursework assigned to students as naïve data collectors. First, there has been few findings on tone mergers in HK Cantonese that used phrase elicitation, as most previous studies used syllables produced in citation forms in isolation or in a carrier sentence. Second, sentence elicitation would be more casual and closer to natural speech [10] where speakers would pay more attention to the meaning expressed than the sound assessed than carrier phrase elicitation where the target syllable usually does not contribute to the meaning of the sentence. Our stimulus is familiar among HK Cantonese speakers as an illustration of Cantonese tonal system, and speakers would attend to tone difference by nature in connected speech. Finally, tones are difficult to differentiate in isolation without other linguistic information [11], and all six lexical tones presented in one phrase would help speakers adjust their pitch range to produce more accurate relative pitch. Each speaker was required to produce the phrase twice.

2.3. Data processing

After data collection and consolidation, a preliminary manual check was carried out. Except for four speakers, each speaker gave two clear repetitions of the target phrase, providing in total 192 tokens for analysis. Transcription and segmentation were completed by trained transcribers who are native speakers of HK Cantonese. We then extracted and measured pitch correlates of the rhyme only in each syllable. Using ProsodyPro [16] in Praat [5], we marked 10 equidistant points automatically and extracted F0 values within 50 Hz and 650 Hz. Manual verification of extraction was carried out to ensure that the first and the last points of the target was the beginning and end of periodicity within the target rhyme.

Values obtained at the first 2 points and the last one (points 1, 2, 10) were excluded to avoid perturbation or creakiness caused by preceding and following consonants, as well as to avoid syllable-edge effects [6]. Therefore, point 3 in the contour would be defined as the new onset, and point 9 the new offset (in Figure 2 of Section 3). Then, F0 values at these 7 points were standardized into z-scores (Z) using speaker-specific means and standard deviations. A pitch slope was calculated using the following method: $(Z_{offset} - Z_{onset})$ / duration. However, since the 10-point extraction already provided time-normalised value, the slope would therefore be simplified as pitch range: $Z_{offset} - Z_{onset}$.

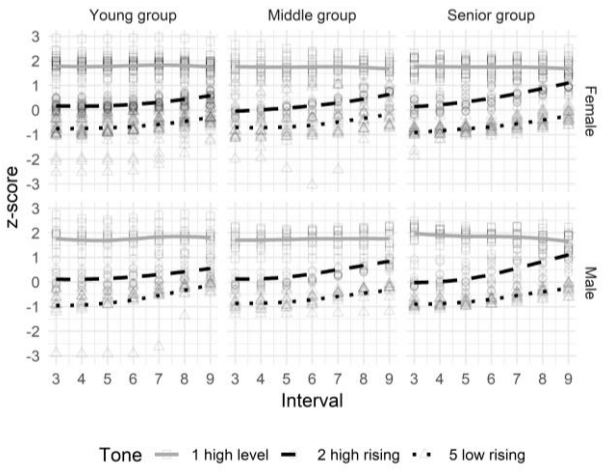
In order to compare pitch correlates of tokens across groups, three linear mixed effects models built using the *lme4* package [2] in R [14] were fitted. In each model, z-scores of onset, offset or pitch range were treated as the dependent variable, whereas age (3 groups), gender (2 groups) and tone type (2 groups) were entered as fixed variables, with T2, female and the young group set as reference levels. In addition, a three-way interaction of variables above was entered in the model. Different speakers were treated as the random intercept. In order to compare main and interaction items, anova function was used. P-values

were obtained using the *lmerTest* package [9]. Multiple comparisons were calculated by using the emmeans function in the *emmeans* package [12] with Tukey adjustment where significance was shown.

3. RESULTS

Firstly, F0 contours of the high rising tone T2 and the low rising tone T5 were found in distinct patterns from each other, as shown in Figure 2 below. Recall that our speakers were recruited in a quasi-random manner. These patterns suggest that T2 and T5 are largely in separate categories in contemporary HK Cantonese, which echoes findings in [13]. Secondly, offsets of T2 and T5 seem to be drawn apart from each other at various extents across groups. Thirdly, corresponding to the offset variation, slopes of T2 appear in different degree of steepness across groups. In the senior group, the offsets are closer to those of the high level tone (T1), which is shown as references in Figure 2 as they remain static across all groups. But as age decreases, the slope would become gentler and therefore T2 offsets move further away from T1, resulting in a near parallel contour as T5 in the middle and young groups.

Figure 2: F0 contours of T1, T2 and T5 across age and gender groups, using loess regression.



Linear mixed effects analysis performed on the onset Z-scores shows significant differences for T2 and T5 (F = 169, p < .001), indicating that the pitch onsets of the two tones may not be the same. This pattern suggests that the tone notation under the 5-point scale, of T2 as 35 and of T5 as 13 differing in pitch onsets might be more faithful to contemporary tone articulation in HK Cantonese (Analysis using all 10 points of F0 extracts yields similar pitch contours and ranges). Other than the significant effects shown in tone, there is no significant difference in gender or age, or any interaction effects (all F < 1.0, p > .36).

Linear mixed effects analysis performed on the offset z-scores shows significant differences for T2 and T5 (F = 367, p < .001), indicating that pitch offsets of the two tones are different. Statistics reveals that gender does not seem to affect such difference in tone production (F = .23, p > .6), but that age contributes significantly to the variation in pitch offsets (F = 4.02, p < .02). There is also a significant effect of the age and tone interaction (F = 10.11, p < .001), which will be further analysed in Section 3.2, but no other significant interaction is found (age * gender: F = .14, p > .86; gender * tone: F = .57, p > .57; age * gender * tone: F = 2.34, p = .1).

Linear mixed effects analysis performed on the pitch slope shows significance for tone factor (F = 6.00, p < .01), indicating a difference between the two tone slopes. Age is found to contribute significantly to the difference (F = 6.13, p < .005). Interaction terms between age and tone, as well as the three-way interaction among age, gender and tone both contribute significantly to the differentiation of two slopes (F = 41.67 and 3.61 respectively, p < .001 and .05), which will be further elaborated in Section 3.2 and 3.3 below.

3.1. Age effects

As there have been no significant age factors contributing to the difference found in the onset of T2 and T5, no pairwise comparison was run.

However, given the significant difference in pitch offsets of the two lexical tones produced by different age groups, a pairwise comparison using Tukey adjustment was carried out. It reveals that the young group is significantly different from the senior group (t = -4.07, p < .02), while there is no significant difference between the young and the middle, or the middle and the senior groups (both $|t|$ < 2.0, p > .1).

Pairwise comparison on the slopes shows that the young and the senior groups are significantly different in terms of pitch slopes (t = -3.41, p < .005), but no difference is found between the young and the middle groups (t = -.8, p = .70) or the middle and the senior groups (t = -2.4, p = .06).

3.2. Age and tone interaction

As there has been no significant interaction found in T2 and T5 onsets, no pairwise comparison was run.

Due to significant interactions found in the two tone offsets, a pairwise comparison was carried out. The results show that T2 offsets by the senior group would differ significantly from those by the young group (t = -4.82, p < .001) and by the middle group (t = -3.08, p < .07), but no difference is shown between

offsets by the young group and by the middle group (t = -1.47, p = .31). While comparing T5 offsets, three groups show similar performance (all $|t| < .4$, $p > .9$). This indicates that any change in the pitch offset is more likely to start with T2, as we found speakers of the young and the middle groups have lowered their F0 in the offset, while senior speakers maintain a higher F0, which is different from findings in [4, 13] that T5 is merged with T2. However, comparison within age groups shows that all three groups have distinctive pitch offsets between T2 and T5 (all $t >$ 8.8, $p < .001$). This suggests that T2 and T5 may have remained distinctive from each other, even though the categorical difference is shrinking among the young group, similar to findings in [13].

There is also a significant interaction between age groups and tones in terms of slopes, therefore, a pairwise comparison was carried out. In terms of the T2 slope, the senior group performs significantly different from the young group (t = -5.22, $p < .001$) and the middle group (t = -2.68, $p < .05$), but no difference is observed between the young and the middle groups (t = -2.22, p = .07). In terms of the T5 slope, all three age groups show similar results (all $t > .9$, $p > .6$). Comparison within three age groups shows that although the young and the senior groups still have distinctive slopes in T2 and T5 (t = -2.45 and 3.62, $p < .05$ and .001), the middle group has similar slopes for the two tones (t = 1.50, p = .13).

3.3. Age, gender and tone interaction

Since no three-way interaction was found in the onset and the offset of two tones, no pairwise comparison was done.

Within-tone comparison of the slopes of T2 and T5 shows that both genders perform similarly with age and tone interaction (all $|t| < .96$, $p > .3$). Same as what we found in Section 3.2, T2 slopes are clearly contrasted between the young and the senior groups for both genders (both $|t| > 4.0$, $p < .002$), but no difference would be shown between the rest groups (all $|t| < 1.8$. $p > .1$). For T5, age and gender do not contribute to any significance (all $|t| < 1.7$, $p > .22$). However, gender effects were observed in cross-tone comparison of slopes produced by different age groups. The young male and the two senior groups have kept two distinct slopes of two tones (all $|t| >$ 2.10, $p < .04$), but within the two middle groups and the young female group, the tonal contrast between T2 and T5 seems to be shrinking (all $|t| < 1.01$, $p > .29$). This may indicate a gender effect within the young group, that young male speakers assimilate the speech style of the more senior groups more than their female counterparts, and the female speakers would likely be leading the change.

4. DISCUSSION

We collected and analysed speech production by native speakers of HK Cantonese in a wide range of age groups. The pitch tracks suggest that the high rising T2 and the low rising T5 are in general kept as distinct categories. The contrast pattern, however, appears to vary across groups.

First, the main difference is observed in the offset and the slope. Onsets of T2 and T5 are comparable in all groups. The offset, however, varies greatly with T2 ending much closer to that of the high level tone T1 in the senior group's production, whereas the offset of other two groups' T2 is drawn away from that of T1. Thus, age seems to be most influential in the variation of T2 production, with the senior speakers maintaining a steeper slope than the young.

Secondly, tonal difference in the onset may be informative in confirming using 35 for notation of the high rising tone (T2) in HK Cantonese. However, the mid-to-high F0 track could be attributed to progressive assimilation of tones: in the reading material, the syllable preceding T2 bears a high level tone, which may give the onset of T2 a boost.

Thirdly, in the middle group and the young females, production of the two tones shows a reduced contrast than the senior or young males, which reveals gender effects in speech variation. The young females may be more sensitive to new trend in speech while the young males are yet to pick up such change.

The tonal variation between the young and the senior, however, should be interpreted with caution. It could be an indication of a change in progress in the high rising tone, as a space packed with tones of comparable contours may have provided motivation and ground for tonal change. The variation could also be a long existing age-grading phenomenon [15], with the young being more resistant to the social norms to produce distinct T2 and T5 contrast. This would require a real-time study to confirm.

Our study discovers two robust cues, F0 offset and slope, to the tonal contrast across the social groups in contemporary HK Cantonese. Our findings, however, are not sufficient to conclude on a change as suggested in previous literature. It may be safe to propose a shrink in tonal contrast between two rising tones among the young. Further disambiguation requires a perceptual assessment to see whether reduction of tone distinction will lead to the confusion in perception and thus contribute to further sound change. Future research could correlate perception with production along social dimensions to attest our findings of a tendency in merging the rising tones in Cantonese.

5. REFERENCES

[1] Matthews, S., Yip, V. 1994. *Cantonese: A Comprehensive Grammar.* London: Routledge.

[2] Bates, D., Maechler, M., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67(*1), 1-48.

[3] Bauer, R. S. 1982. Cantonese sociolinguistic patterns: correlating social characteristics of speakers with phonological variables in Hong Kong Cantonese. PhD dissertation. Berkeley: University of California, Berkeley.

[4] 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*, 211-25.

[5] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43 http://www.praat.org/.

[6] Chasaide, A., Gobl, C. 1993. Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech 36*(2,3), 303–30.

[7] Fung, R. S. Y., Wong, C. S. P. 2011. Acoustic analysis of the new rising tone in Hong Kong Cantonese. *Proc. 17th ICPhS*, Hong Kong, 715-18.

[8] International Phonetic Alphabet (IPA) handbook recording. https://www.internationalphoneticassociation.org/content/ipa-handbook-downloads.

[9] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software 82*(13), 1–26. https://escholarship.org/uc/item/1sn2f75h.

[10] Labov, W. 1972. The isolation of contextual styles. *Sociolinguistic Pattern*. Philadelphia: University of Pennsylvania Press.

[11] Lee, K. Y. S., Chan, T. Y. K., Lam, H. S. J., van Hasselt, C. A., Tong, C. F. M. 2015. Lexical tone perception in native speakers of Cantonese, International *Journal of Speech-Language Pathology*, *17*(1), 53-62.

[12] Lenth, R. 2018. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.0. https://CRAN.R-project.org/package=emmeans.

[13] Mok, P. P. K., Zuo, D., Wong, P. W. Y. 2013. Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change 25*, 341-70.

[14] R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[15] Wagner, S. E. 2012. Age grading in Sociolinguistic Theory. *Language and Linguistics Compass 6*(6): 371-82.

[16] Xu, Y. 2013. ProsodyPro — A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 7-10.