CANTONESE VOWEL MERGER-IN-PROGRESS

Roxana S. Y. Fung and Chris K. C. Lee

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
roxana.fung@polyu.edu.hk; kckrislee@gmail.com

ABSTRACT

This study investigates an unreported ongoing sound change in Hong Kong Cantonese. Cantonese is arguably the only variety of Chinese that contains long-short contrast in its vowel system, which is essentially a contrast in vowel quality. However, results from a production experiment with 60 gender-balanced native Hong Kong Cantonese speakers of three age groups suggests that this contrast is disappearing. Extracted from a read passage, the bark-normalized formant values of the three pairs of vowels with length contrast ([a]-[ɐ], [ɛ]-[e], and [ɔ]-[o]) were compared across the three age groups and gender using linear mixed modelling. While the length contrast is largely retained across age groups, the acoustic difference is diminishing, especially among the young group. This merger-in-progress is actualized by increasing proximity in vowel height. All the vowel pairs seem to adopt the unidirectional merger-by-transfer, but they are realized differently: for [a]-[ɐ] pair, the long vowel transfers to the short one, but the other way around for the other two pairs.

Keywords: Hong Kong Cantonese, Diachronic sound change, Vocalic long-short contrast

1. INTRODUCTION

Unlike all other varieties of modern Chinese, duration has long been considered a distinctive characteristic of the vowel system in Cantonese. Apart from duration, the vowels also contrast in quality. Hence, there are different analyses on the Cantonese vowel contrasts. The system adopted in this study recognizes seven long vowels: [i, y, ɛ, œ, ɔ, a, u], four short vowels: [e, ø, ɐ, o] and four pairs of long-short contrast. The vowels [a]-[ɐ], [ɛ]-[e], [ɔ]-[o], and [œ]-[ø] are each long-short pairs with the short vowels higher than the long ones. [1] As shown in the Cantonese rime inventory listed in Table 1, the [a]-[ɐ] pair is the only long-short vowel pair that displays a full-fledged overlapping distribution in Cantonese. As for the other three pairs, the contrastive distributions are quite limited. Most Cantonese speakers have no problem in maintaining the contrast, but interestingly, more and more undergraduate students learning Cantonese phonology claim that they have much difficulty in distinguishing the [a]-[ɐ] contrast. It has been observed that their pronunciation for the long [a] tended to be merged with the short [ɐ].

This study aims to conduct an acoustic investigation on three age groups of HKC speakers to investigate if there are any on-going mergers in the Cantonese vowel system. In particular, this paper aims to answer the following questions:

1. Are the long-short contrasts in the three vowel groups in the process of merging?
2. If yes, what is the mechanism of the merger-in-progress? and
3. What are the possible causes of the merger?

Table 1: The rime inventory of Cantonese

	Front						Central		Back		
V#	i	y	ɛ		œ		a		ɔ		u
VG				ej		øj	aj	ɐj	ɔj		uj
	iw		ɛw				aw	ɐw		ow	
VC	im		ɛm				am	ɐm			
	in	yn	ɛn			øn	an	ɐn	ɔn		un
			ɛŋ	eŋ	œŋ		aŋ	ɐŋ	ɔŋ	oŋ	
	ip		ɛp				ap	ɐp			
	it	yt	ɛt		œt	øt	at	ɐt	ɔt		ut
			ɛk	ek			ak	ɐk	ɔk	ok	

(shaded columns contain short vowels)

2. METHOD

2.1 Speakers and task

60 native speakers of Hong Kong Cantonese in three age groups (young: 18-25, middle-aged: 35-45, and elder: 55-65) with balanced gender (5 speakers × 3 age groups × 2 genders) were selected. They were asked to self-rate their Mandarin proficiency and were divided into three proficiency groups accordingly: 15 rated themselves 'fluent', 33 rated 'average', and 12 'limited'. Their English proficiency were rated by two native speakers of English based on their recorded passage reading, and the participants were again divided into three proficiency groups (20 speakers per group) according to the marks given by the two raters. Participants were asked to record a 350-character descriptive passage written in accordance with the syntax of colloquial Hong Kong Cantonese. All participants finished reading the passage in about 90 seconds.

2.2 Acoustic analysis and data processing

Only the first half (about 45 seconds) of the recordings were analyzed. All syllables of the read passage were segmented manually. Each syllable has formants tracked in 11 subsegments by Snack toolkit in VoiceSauce [2]. Outputted formant values were then bark-transformed by (1) where *F* is the formant value:

(1) [26.81× *F* / (1960 + *F*)] -0.53 [3]

Since all short vowels must be in a rime structure that is either VC or VG, only the third subsegment were considered in the statistical analysis. To measure whether there is a merger in the three pairs of vowels [a]-[ɐ], [ɛ]-[e], and [ɔ]-[o] ([œ]-[ø] is not discussed due to insufficient data) [i], the Euclidean distance (in bark scale) is computed for the three pairs of vowels for each speaker using (2), where *i* and *j* denote different vowels.

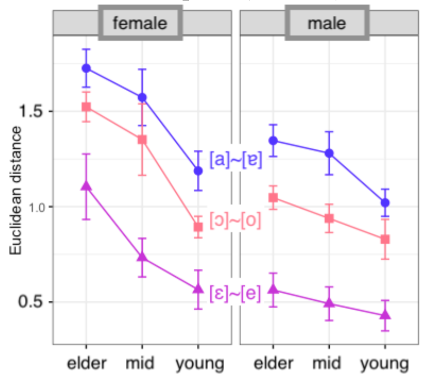
(2) $\sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2}$

3. RESULTS

3.1 Are the long-short contrasts merging?

Figure 1 shows the Euclidean distances between the long and short vowels in the vowel pair across generations. The distance between the two vowels of the same pair is reducing across generations, which is a sign of vowel merger-in-progress.

Figure 1: Euclidean distances between long and short vowels in the three vowel pairs (with SE)



The distance measures were compared by separate 2 (gender) × 3 (age groups) ANOVAs. Interaction effect between age and gender is not significant for all three pair. However, for all three pairs, main effect of age is significant ([a]-[ɐ]: $F(2,54)=9.06$, p=0.0004; [ɛ]-[e]: $F(2,54)=4.89$, p=0.01, [ɔ]-[o]: $F(2,54)=8.71$, p=.0005). Post-hoc pairwise comparisons of age group adjusted by Tukey HSD method suggest that the young speakers are producing the long-short contrast with a reduced Euclidean distance when compared to the elder speakers ([a]-[ɐ]: *p*=.0004; [ɛ]-[e]: *p*=.009, [ɔ]-[o]: *p*=.0004), and, to a lesser extent, the middle-aged speakers ([a]-[ɐ]: *p*=.009; [ɛ]-[e]: *n.s.*, [ɔ]-[o]: *p*=.02). While Figure 1 suggested that the reduction in Euclidean distance between the two members of the three vowel pairs start from the middle-aged group, these speakers produced no significant differences from the elder speakers. Hence, the significant main effect of age and the significant pairwise comparisons between age groups backs our impression the long and short vowels are in the process of merging among the young speakers.

On the other hand, all three pairs also show significant main effect of gender ([a]-[ɐ]: $F(1,54)=10.54$, p=0.002; [ɛ]-[e]: $F(2,54)=11.74$, p=0.001, [ɔ]-[o]: $F(2,54)=14.11$, p=.0004). In general, male speakers produce the long-short contrast with a smaller Euclidean distance between the two vowels. Considering the ANOVA results together, the distance measures suggest that while male speakers in general are more advanced in this sound change, there is a vowel merger-in-progress in apparent time.

3.2 What is the mechanism of the merger?

While Euclidean distance is a good indicator of vowel merger-in-progress, it does not tell us the mechanism of the merger: whether the merger is actualized by one member of the pair shifting toward the other (unidirectional transfer), or both members shifting toward each other (bidirectional approximation). [4]

To better understand how the vowel merger-in-progress were actualized in the acoustic space, the bark-normalized values of the first two formants of the three vowel pairs were checked with linear mixed models, using the *lme4* package [5] in R [6]. Bark-normalized F1 and F2 values were put into separate models, and given the relatively independent vowel pairs, they were analyzed in separate models. For each model, the fixed effects considered were vowel length, gender, age group, and their interactions; the random effects considered were speaker, and word. The statistical significance of all fixed effects was established through Likelihood Ratio Test [7], and that of pairwise comparisons were determined using the *lmerTest* package [8].

Table 2: Best models for the three pairs of vowels

	Model for F1	Model for F2
[ɛ]-[e]	Length × age × gender	Gender
[ɔ]-[o]	Length × age; Length × gender	Gender; Age group
[a]-[ɐ]	Length × age; Length × gender	Length × gender

Table 2 summarizes significant fixed effects of all models. In general, while the Euclidean distance is reduced, the long-short contrast is still evident in the F1 dimension. However, the downward shift of the short vowels led to generational differences in the distribution of vowels and possibly the impression of vowel merging. For the fitted Bark-normalized formant values of all vowels, please refer to the footnote.[ii]

3.2.1 [ɛ]-[e] pair

Figure 2: Fitted Bark-normalized F1 of [ɛ] and [e] of different groups of speakers (with SE)

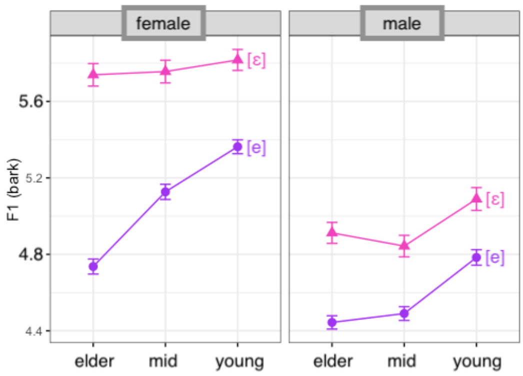


Figure 2 shows the Bark-normalized F1 of [ɛ] and [e] of the six groups of speakers respectively. It suggests that the downward shift of [e] in the acoustic space across generations, with female speakers showing a greater shift than males, causes a decrease in Euclidean distance. This downward shift of [e] across the generations makes [e] produced by young speaker differ significantly from the elder group for both genders (elder male *vs.* young male: $t(69.9)=2.45$, $p=.017$; elder female *vs.* young female: $t(69.4)=4.65$, $p<.0001$). Additionally, middle-aged female speakers produced an acoustically different [e] from elder female speakers ($t(70.1)=2.99$, $p=.004$). On the other hand, while there are changes in the [ɛ], no pairwise comparison is significant.

While there is a significant downward shift of [e] across generation in both genders and [ɛ] is relatively stable in the acoustic space, statistics show that the two vowels are not merged completely in the acoustic vowel space: All within-group comparisons show that the F1 values of [ɛ] and [e] are significantly different ($p<.02$ for all six groups of speakers).

Recall that the best F2 model for this pair contains only gender as fixed effect. Second formant then has no significant role to play in the long-short contrast of this pair, and this is true across generations of speakers. The downward shift of [e] alone primes the impression of merger-in-progress. It seems that the merger is actualized by unidirectional transfer: the short [e] transfers to the long [ɛ].

3.2.2 [ɔ]-[o] pair

Figure 3: Fitted Bark-normalized F1 of [ɔ] and [o] of different groups of speakers

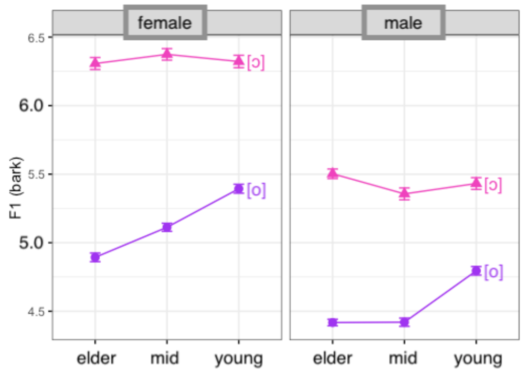


Figure 3 shows the Bark-normalized F1 of [ɔ] and [o] of the six groups of speakers. In essence, the reduction in Euclidean distance between the two vowels has the similar pattern as of the [ɛ]-[e] pair: The short vowel ([o]) shifts downward and causes the two vowels become nearer to each other. This downward shift of [o] again makes young speakers different from the elder speakers ($t(74.6)=4.11$, $p=.0001$) and it is true for both genders. What is different from the [ɛ]-[e] pair is that [o] produced by young speakers again differs significantly from the middle-aged speakers ($t(75.8)=3.12$, $p=.003$). Just like the [ɛ]-[e] pair, the long vowel [ɔ] is again relatively stable across age (all pairwise comparison of [ɔ] $p>.60$). Hence, the vowel merger-in-progress is again a result of the unidirectional transfer of the short [o] to long [ɔ].

However, recall that the Euclidean distance between [ɔ] and [o] is relatively greater than that of [ɛ]-[e] pair, it is not surprising to see that all within-group long-short difference in F1 is significant ($p<.0001$ for all six groups of speakers). This again points to the fact that while the two vowels are becoming more alike in the F1 dimension, the two vowels are not yet completely merged.

For the F2 dimension of this pair, second formant again then has no significant role to play in the long-short contrast of this pair, and this is true across generations of speakers, given only gender and age group are included in the best model.

3.2.3 [a]-[ɐ] pair

Among the three pairs investigated, [a] and [ɐ] is the only pair that can contrast meaning in both VG and VC structures. Hence it is perhaps not surprising to see a relatively greater Euclidean distance between [a] and [ɐ] compared to the other two pairs. This greater Euclidean distance is echoed by the significant within-group long-short comparisons in all groups of speakers (*p*<.0001). Hence, similar to the other two vowel pairs, while the distance between the long [a] and short [ɐ] is reducing, they are still acoustically distinct in terms of the F1 dimension. Why then is the pair perceived as merged by some of the native speakers?

Figure 4: Fitted Bark-normalized F1 of [a] and [ɐ] of different groups of speakers

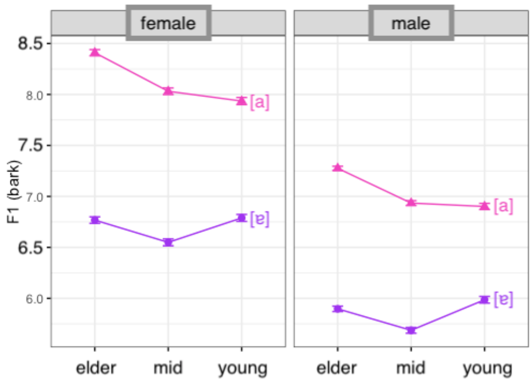


Figure 4 shows the Bark-normalized F1 of [a] and [ɐ] of the six groups of speakers. Unlike the other two vowel pairs, the long vowel ([a]) in this pair has shifted upward. Figure 4 shows that the alleged reduction of [a]-[ɐ] contrast actually starts from the middle-aged speakers. They have produced a raised [a], which has a significantly higher F1 than the elder speakers ($t(74.5)=3.02$, $p=.003$). As for the young group, the height of [a] they produced is, in general, similar to that of the middle-aged speakers, since it is again significantly different from the elder speakers ($t(73.6)=3.24$, $p=.002$), but not different from the middle-aged speakers ($p>.80$), and this trend is true in both genders.

Apart from the upward shift of [a], Figure 4 also shows a tendency of [ɐ]-shifting. On the one hand, [ɐ] produced by middle-aged speakers is slightly raised but such change does not produce statistical significance when compared to the elder speakers ($p=.066$). On the other hand, while the young speakers produced a lower [ɐ] compared to the speakers of elder speakers, it is only significantly lower than the slightly raised [ɐ] produced by the middle-aged speakers ($t(66.5)=2.45$, $p=.017$). Hence, the perceived merger-in-progress of this vowel pair is still essentially unidirectional: The raising of [a] in terms of tongue position of the younger groups compared to the elder group.

Unlike the other two pairs, the best F2 model of this pair contains the factor group vowel length × gender. However, while there is a significant gender difference in F2 values for both [a] ($t(70.0)=-13.9$, $p<.0001$) and [ɐ] ($t(64.0)=-12.3$, $p<.0001$), the within-group long-short comparisons are not significant (male: $p=.614$; female: $p=.253$). Hence, the F2 is again not playing a significant role in the long-short contrast in this vowel pair.

4. POSSIBLE CAUSES OF MERGER

4.1 Vowel space reduction leading to merger?

While all three pairs actualized vowel merger as unidirectional transfer, it is the long vowel that shifted upward for [a]-[ɐ] pair, but the short ones downwards for the other two pairs. This gives the impression of vowel space reduction. To check it is the case across generations, the formant values of [i] and [u] (in Bark) were analyzed (see footnote [ii] for the formant values of the two vowels, §3.1 for the statistical analysis details). Results show that both [i] and [u] remained acoustically stable in the apparent time. Downward shifting of [e] and [o] is hence not due to the pressure from the shifting [i] and [u]. What causes upward shifting of [a], however, requires further investigation.

4.2 Language contact effect?

To assess the possibility of the effect of language contact on the observed Cantonese vowel merger, we ran separate linear mixed models to predict the effect of Mandarin proficiency and English proficiency on the F1 and F2 values of the three pairs of vowels. In each model, fixed effects considered were vowel length, gender, and language proficiency (age is excluded in these models because it is not balanced across proficiency groups), with speaker and words as random effects. Results show that proficiency effects, whether it is Mandarin or English, does not show any influence on either formant values of all vowels. This suggests that the observed vowel merger-in-progress in Hong Kong Cantonese is more likely to be an internally-motivated sound change.

5. CONCLUSION

The present study discussed the unreported ongoing disappearance of long-short vowel contrast in HKC, which is principally a contrast in vowel height, across different age groups. Among the three pairs of vowels investigated, while [ɛ]-[e] has lost the contrast almost completely, the pairs [ɔ]-[o] and [a]-[ɐ] showed a tendency to merge in the younger generations of HKC speakers. Our findings suggest an on-going sound change in the loss of vowel contrasts in Cantonese based on apparent time data. However, due to limited information about the social variables in the current study, the causes of the sound change are yet to be determined.

6. ACKNOWLEDGEMENTS

We would like to thank the two anonymous reviewers for their valuable comments on the previous draft, our participants for providing the data, and acknowledge support from the Dean's Reserve provided by the Hong Kong Polytechnic University (#1-ZVG3).

7. REFERENCES

[1] Li X.-H. *et al.* (1995) *Guangzhou Fangyan Yanjiu* (廣州方言研究). Guangzhou: Guangzhou People's Publishing House.

[2] Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. *Proceedings of the ICPhS XVII*, 1846–1849.

[3] Traumüller, H. (1990) Analytical expressions for the Tonotopic Sensory Scale. *Journal of the Acoustic Society of America,* 88:97-100.

[4] Labov, W. (1994) *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

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

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

[7] Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. URL: http://arxiv.org/pdf/1308.5499.pdf.

[8] 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.

[i] The mean number of tokens considered for statistical analysis per vowel is shown in the table below (with SD). Note that in cases where formant tracking is inaccurate, or mispronunciation, the speaker will have less tokens.

Vowels	Token #	Vowels	Token #
[i]	15 (1.81)		
[ɛ]	11 (1.45)	[e]	14 (1.44)
[a]	17 (2.82)	[ɐ]	19 (1.70)
[ɔ]	11 (1.40)	[o]	13 (1.37)
[u]	8 (0.98)		

[ii] (a) The mean Bark-normalized formant values of each vowel produced by female speakers (with SD)

	Elder		Middle-aged		Young	
	F1	F2	F1	F2	F1	F2
ɛ	4.74 (0.47)	14.0 (0.53)	5.13 (0.47)	14.0 (0.47)	5.36 (0.44)	13.8 (0.57)
e	5.74 (0.59)	13.8 (0.44)	5.76 (0.59)	13.7 (0.37)	5.82 (0.56)	13.5 (0.51)
ɔ	4.89 (0.37)	8.94 (1.09)	5.11 (0.34)	9.12 (1.09)	5.39 (0.38)	9.55 (1.15)
o	6.31 (0.46)	8.97 (0.52)	6.37 (0.44)	9.13 (0.55)	6.32 (0.48)	9.53 (0.62)
a	6.77 (0.49)	11.1 (1.15)	6.55 (0.50)	11.1 (1.19)	6.79 (0.51)	11.3 (1.21)
ɐ	8.41 (0.37)	11.4 (0.51)	8.03 (0.39)	11.3 (0.58)	7.94 (0.44)	11.6 (0.61)
i	3.52 (0.33)	14.7 (0.95)	3.67 (0.35)	14.7 (0.97)	3.79 (0.36)	14.8 (1.01)
u	3.98 (0.28)	7.90 (1.06)	4.13 (0.31)	7.90 (1.09)	4.23 (0.33)	8.17 (1.08)

(b) The mean Bark-normalized formant values of each vowel produced by male speakers (with SD)

	Elder		Middle-aged		Young	
	F1	F2	F1	F2	F1	F2
ɛ	4.44 (0.42)	12.8 (0.49)	4.49 (0.42)	12.6 (0.54)	4.78 (0.48)	12.9 (0.47)
e	4.91 (0.58)	12.6 (0.43)	4.84 (0.57)	12.3 (0.47)	5.09 (0.62)	12.6 (0.40)
ɔ	4.42 (0.27)	8.14 (1.05)	4.42 (0.32)	8.32 (1.07)	4.80 (0.35)	8.41 (1.07)
o	5.50 (0.38)	8.13 (0.45)	5.36 (0.43)	8.32 (0.49)	5.43 (0.44)	8.39 (0.53)
a	5.90 (0.41)	9.99 (1.13)	5.69 (0.43)	9.96 (1.15)	5.99 (0.51)	10.1 (1.15)
ɐ	7.28 (0.27)	10.1 (0.49)	6.93 (0.30)	10.0 (0.50)	6.90 (0.40)	10.2 (0.51)
i	3.14 (0.29)	13.3 (0.98)	3.12 (0.29)	13.3 (0.90)	3.12 (0.34)	13.5 (0.96)
u	3.84 (0.29)	7.79 (1.10)	3.83 (0.32)	7.64 (1.04)	3.91 (0.27)	7.65 (1.06)