Citation information:
Tse, Holman (2019). Vowel shifts in Cantonese?: Toronto vs. Hong Kong. *Asia-Pacific Language Variation*, *5*(1), 67-83. https://doi.org/10.1075/aplv.19001.tse

Published online: June 13, 2019

Vowel shifts in Cantonese?: Toronto vs. Hong Kong

Holman Tse

University of Pittsburgh

This paper addresses Labov's Principles of Vowel Chain Shifting in Toronto and Hong Kong Cantonese based on sociolinguistic interview data from the Heritage Language Variation and Change in Toronto Project. The analysis is based on the F1 and F2 values of a total of 33,179 vowel tokens from 11 monophthong categories produced by 32 speakers (8 from Hong Kong and 24 from Toronto). Results show the retraction of [y] among Toronto speakers when generational group (immigrated to Canada vs. grew up in Canada) is modeled as a main effect. When age is modeled instead, [i] is shown to be fronting in apparent time. In the Hong Kong group, age is a significant predictor for the lowering of [ɪ], [ʊ], [ɔ], and for the fronting of [ɔ] and [i]. Overall, results show more vowel shifting in Hong Kong than in Toronto as well as vowel shifting consistent with Labov's Principles.

Keywords: heritage languages, bilingualism, sociophonetics, contact linguistics, Chinese language – Yue; sound change

1. Introduction

Labov (1994, 2011) proposes several Principles of Vowel Chain Shifting. In one version, these principles are stated as follows:

1) "Tense nuclei rise"
2) "Lax nuclei fall"
3) "Back nuclei move to the front" (Labov, 2011, p. 151)

These principles have motivated many studies of vowel variation and change in progress especially in English dialects. Relatively few studies, however, consider the extent to which these principles apply to change in progress in non-Indo-European languages. The goal of this paper is to address this research gap by presenting results from a comparative study of vowels in Hong Kong and Toronto Cantonese (Sino-Tibetan Family). This study is based on acoustic data collected as part of sociolinguistic interviews from speakers in both Hong Kong and Toronto. Like English, Cantonese also has a tense/lax distinction. This phonological feature makes a study of Cantonese well-suited for addressing the applicability of Labov's (1994, 2011) Principles in a non-Indo-European language. The question that is specifically addressed is as follows: Is there evidence for vowel shifting in apparent time in either Toronto or Hong Kong Cantonese?

2. Background

Cantonese belongs to the Yue subgroup of the Sino-Tibetan (Chinese) language family. According to Yue-Hashimoto (1972, 1991), the term "Cantonese" has been used ambiguously in the English-speaking world to refer to both the dialect of the city of Guangzhou (or "Canton" following the Portuguese spelling) and a group of dialects spoken in Guangdong Province (which has also been transliterated in English as “Canton”). In this paper, “Cantonese” is used to refer to the former while “Yue” is used to refer to the latter. This follows Yue-Hashimoto’s (1972, 1991) usage. Some Cantonese linguists have referred to the Guangzhou dialect as “Standard Cantonese” to make it more clearly distinguishable from other Yue dialects, not all of which are mutually intelligible with each other (Bauer & Benedict, 1997; Yue-Hashimoto, 1972, 1991). Included in the geographical reach of “standard” are the Special Administrative Regions of Macau and Hong Kong. The Hong Kong variety has widely been recognized as the prestige form due to the global economic and cultural importance of Hong Kong (Matthews & Yip, 2011, p. 3). According to the most recently available survey, 88.1% of Hong Kong’s population of 7.3 million are mother tongue speakers of Cantonese (Census and Statistics Department, Hong Kong Special Administrative Region, 2016).

Cantonese speakers are also found in diasporic communities throughout the world in countries such as Malaysia, Australia, the United States, and Canada. According to the most recent census, 565,270 Canadians speak Cantonese as their mother tongue (Statistics Canada, 2017). About 44% (or 247,710) of these 565,270 speakers live in the Toronto Census Metropolitan Area (Ibid.). This makes Cantonese the second most spoken mother tongue in the Toronto Census Metropolitan Area after English. Unlike in Hong Kong where the majority of the population speaks Cantonese, in Toronto Cantonese mother tongue speakers represent less than 5% of the population of 5,883,670. All of the second-generation Toronto speakers analyzed in this study describe English as their preferred language. Thus, influence from Toronto English is a possibility that needs to be considered for second-generation Cantonese speakers.

One relatively uncontroversial issue in Hong Kong Cantonese is the number of surface (or phonetically distinct) vowels. Whether or not these surface vowels are derived from a smaller set of vowel phonemes is a controversial question[1]. This paper follows the description presented in Zee (1999) which describes 11 surface monophthongs and 11 surface diphthongs. The focus of this paper is on the acoustic production of the 11 monophthongs in terms of midpoint F1 and F2 measurements. These 11 vowels include seven tense vowels (**Table 1**) and four lax vowels (**Table 2**). Lax vowels occur only in closed syllable context while tense vowels can occur in either open or closed syllables. Acoustic studies have also shown that Cantonese tense vowels are longer in duration than lax vowels (Zee, 2003).

<INSERT TABLE 1>

<INSERT TABLE 2>

Recent research on variation in Cantonese vowels, however, has focused on Toronto speakers rather than on Hong Kong speakers. Tse (2016a), for example, shows the emergence of allophonic splits in [ɛ] and [ɔ] and suggests that these splits may be contact-induced. What is missing in this study is comparative acoustic data from Hong Kong speakers. Thus, it is uncertain whether these splits have also developed among younger Hong Kong speakers or among Cantonese speakers in other communities. Similarly, Tse (2016b) showed that the acoustic difference between the vowels [i] and [ɪ] is greater among GEN 2 than it is among GEN

[1] Descriptions of Standard Cantonese vary in reporting anywhere between five and 11 vowel phonemes (Yip, 1996). This is due to different ways of analyzing multiple complementary distribution relationships in structuralist analyses of the vowel system. For a full discussion of this topic, see Yue-Hashimoto (1972), Bauer and Benedict (1997), and Barrie (2003). Yip (1996) and Yu (2000) propose that a surface-based description may in fact be more appropriate for Cantonese given the lack of inflectional morphology and hence lack of evidence from morpho-phonological alternations to posit abstract or underlying phonemes that differ from their surface or phonetic representations. This is the approach taken in this paper contra the approach taken in previous studies of Toronto Cantonese (Tse, 2016a, 2016b).

1 speakers and argued that this change is also contact-induced. Comparative acoustic data from Hong Kong Cantonese speakers is also missing from this study.

Some existing studies suggest that Hong Kong Cantonese speakers have also increased the acoustic difference between [i] and [ɪ]. Lee (1983), for example, shows that Hong Kong speakers have more peripheral variants of [i] and lower variants of [ɪ] than Guangzhou Cantonese speakers. Zee (2003) shows an even greater acoustic separation between these two vowels in a study of 100 Hong Kong speakers between the ages of 18 and 21. The separation is so great that [ɪ] has lowered to the point of almost complete overlap with [ɛ] in the vowel space, although the two remain phonologically distinct with [ɪ] being lax and short and [ɛ] being tense and long. Based on a comparative study of different Yue dialects involving textual data, Lau (2003) showed that the lowering of [ɪ] and [ʊ] are, in fact, part of vowel chain shifts that have developed in the history of Cantonese. What is missing from the existing literature on Hong Kong Cantonese vowels are studies examining speakers from a wide range of ages.

3. Data and Methods

The data for the current study comes from a sub-set of interviews collected as part of the Heritage Language Variation and Change (HLVC) in Toronto Project (Nagy, 2011). The project corpus includes hour-long sociolinguistic interviews (following the best practices described in Labov, 1984), an Ethnic Orientation Questionnaire (EOQ), and a picture naming task. Cantonese is one of nine heritage languages included as part of this project. The Cantonese interviews were conducted from 2009 to 2010. To complement the recordings of Toronto speakers, a set of interviews of Hong Kong speakers conducted in 2015 are also included. These interviews were conducted following the same procedures as the interviews conducted in Toronto.

The speakers analyzed include: Homeland speakers (n = 8), GEN 1 Toronto speakers (n = 12), and GEN 2 Toronto speakers (n = 12) for a grand total of 32 speakers. The Homeland group includes speakers who were born in Hong Kong (HK) and have since lived continuously in Hong Kong. GEN 1 includes speakers who grew up in Hong Kong, moved to the Greater Toronto Area (GTA) as adults, and have lived in the GTA for at least 20 years. GEN 2 speakers include those whose parents would qualify as GEN 1 speakers (even if those parents are not in the corpus). The HLVC Project criteria included those who arrived in the GTA before the age of six to be included as part of the GEN 2 group. Of the three speakers analyzed in the current study who were not born in the GTA, two arrived at the age of 2 while the third arrived at the age of 4. **Table 3** lists the speakers included in this study.

<INSERT TABLE 3>

To facilitate processing of the data, speech to segment forced alignment was performed on ELAN (Sloetjes & Wittenburg, 2008) transcriptions for each speaker and the accompanying .wav files using the program Prosodylab-Aligner (Gorman, Howell, & Wagner, 2011). The output was a set of Praat (Boersma & Weenink, 2016) textgrids for each corresponding .wav file. These textgrids had all phonemes automatically labeled (based on Jyutping transcription) and automatically aligned to the waveform and spectrogram. Each textgrid and .wav file pair was manually reviewed to ensure accuracy and corrected if needed[2].

[2] Peters and Tse (2016) showed that Prosodylab-Aligner has an accuracy rate of about 80% for Cantonese data. Thus, about 20% of the generated textgrids required correction.

During review of each textgrid and .wav file pair, specific tokens containing one of the 11 Cantonese monophthongs were identified and labeled for analysis. A maximum of 10 tokens for each word per speaker were selected. Tokens with onset glides ([j] and [w]) or labio-velar co-articulated stops ([kʷ] and [kwʰ]) were excluded. Vowels in open syllable context immediately followed by a syllable with an onset glide without an intermediate pause (ex: [hɔ2 ji5][3], 'able to') were also excluded. Tokens were also excluded in cases involving undershoot, overlapping speech, laughter, singing, too much background noise, unusually rapid speech, or other problems that make reliable formant measurements difficult or impossible to obtain.

After reviewing each textgrid and .wav file pair, a Praat script was run on all usable file pairs to automatically extract the values of the first two formants for the selected vowel tokens. The Praat formant tracker was set to five formants and a window length of 0.025 seconds. The maximum number of formants set was either 5000 Hz for males or 5500 Hz for females. The output file included a list of all tokens extracted along with vowel category, word, and F1 and F2 measurements for each token.

After reviewing each output file for possible errors (ex: unusually high or low F1 or F2) and correcting or discarding tokens as appropriate, the data was normalized using the Lobanov technique in the NORM suite (Thomas & Kendall, 2007). This was the same normalization method used in previous studies of Toronto Heritage Cantonese (Tse, 2016a, 2016b). Having all data normalized together makes it possible to compare speakers from different generational groups as well as speakers from two different places. The output of NORM was a new tab

[3] IPA transcription is used for the segmental transcription here. The "2" after [hɔ] indicates Tone 2 (a mid-rising tone) while the "5" after [ji] indicates Tone 5 (a low-rising tone).

delimited text file with normalized values for the first two formants for each token along with transcriptions and speaker identifiers.

Finally, the last processing step was to merge the NORM output into a spreadsheet with all of the independent variables examined. The spreadsheet was saved as a tab delimited text file and uploaded to the program Rbrul (Johnson, 2009) for mixed effects modeling. The independent variables considered in the models presented below include speaker and word as random effects and either age, group (GEN 1, GEN 2), or city (Hong Kong or Toronto) as the fixed effect.

4. Results

In Section 4.1 is a general overview of the vowel space of both the Hong Kong and Toronto groups based on F1 and F2 measurements from 33,179 vowel tokens. The possiblity that Toronto speakers are diverging from Hong Kong speakers due to influence from Toronto English is considered through an inter-generational comparison in Section 4.2. In Section 4.3 are results addressing whether or not vowel shifts have developed in either location by considering age as a continuous fixed effect.

4.1 Vowel Space Overview

Figure 1 shows the total number of tokens included for each vowel and for each group. The grand total of usable tokens is 33,179. Table 4 shows the percentage of total tokens for each group that is represented by each vowel category. Vowel categories are listed based on the percentage ranking for the GEN 1 group. The ranking of the most common vowels is similar across all three groups. The only difference between GEN 1 and GEN 2 is the relative ranking of the two least frequent vowels, [u] and [ɵ], while the only difference between GEN 1 and the Homeland group is the relative ranking of [ɔ] and [ɐ]. These vowels are indicated in bold. In both cases, the relative ranking is switched around. With similar rankings across all three groups, the speech samples analyzed across all three speaker groups appear to be comparable.

<INSERT FIGURE 1>

<INSERT TABLE 4>

F1 and F2 means for each vowel category across the three groups are included in the vowel plots that follow with Figure 2 showing the tense vowels and Figure 3 showing the lax vowels. Each ellipse represents one standard deviation from the mean F1 and F2 of each vowel for each group. The mean F1/F2 is represented with a red dot for the GEN 1 group, a blue empty square for the GEN 2 group, and a green triangle for the HK group. Of the 33,179 total tokens included, 22,346 are for the tense vowels while 10,833 are for the lax vowels. Most of these 11 vowels are acoustically distinct in F1/F2 across all three groups. The two notable exceptions involve round vowels overlapping with unrounded vowels. For example, [i] and [y] show overlap for the GEN 1 and GEN 2 groups. Similarly, [ɐ] and [ɵ] also overlap. Since [y] and [ɵ] are round and [i] and [ɐ] are unrounded, the overlap in F1/F2 values for these vowels does not indicate merger.

The relative similarity between GEN 1 and GEN 2 in contrast to the HK group is immediately visible in these plots. In many cases, the GEN 1 and GEN 2 ellipses are closer together than either of them is to the HK ellipses. This is especially the case for [i], [y], [u], and [ɛ]. These four vowels appear to be more peripheral[4] for the HK group than they are for either GEN 1 or GEN 2. They are raised, fronted, or both raised and fronted in comparison to their counterpart vowels in the Toronto groups. The mean formant values shown in Figure 2 and Figure 3 also show similarity between the two Toronto groups.

<INSERT FIGURE 2>

<INSERT FIGURE 3>

GEN 1 and GEN 2 have the exact same mean F1 values (rounded to the nearest whole number) for [a], [i], and [y]. The GEN 2 and Homeland group also share some of the exact same F1 means but only for two vowels ([ɐ] and [ɵ]) as opposed to three. In general, the inter-group differences between GEN 1 and GEN 2 are smaller than they are for differences between the HK and Toronto groups. For instance, while the difference between the HK and GEN 2 group for the F2 of [i] is 67 scaled Hertz, the biggest GEN 1 vs. GEN 2 difference is only about 24 scaled Hertz (for [y]).

In **Table 5** are results from a set of mixed effects models run for each vowel/formant pair. Each model includes either F1 or F2 as the dependent variable, "speaker" and "word" as random effects, and city (Hong Kong vs. Toronto) as a main effect. These models show statistically significant differences between Hong Kong and Toronto speakers for the F1 of [y], [ɪ], [ɔ], [u], and [ʊ] and for the F2 of [ɐ], [ɛ], and [i].

<INSERT TABLE 5>

[4] The peripherality of three of these vowels ([i], [y], and [ɛ]) was also observed in Hong Kong in a study comparing Hong Kong and Guangzhou speakers (Lee 1983).

4.2 Generational group as the main effect

As mentioned in the previous section, the biggest difference between GEN 1 and GEN 2 is for the F2 of [y]. In Table 6 are the results of a mixed effects model for [y] with F2 as the dependent variable, speaker and word as random effects, and group (GEN 1 or GEN 2) and gender as fixed effects. According to this model, the GEN 1 group has a tendency of producing higher F2 (mean of 1634 Hz) than the GEN 2 group (mean of 1608 Hz). This means that the GEN 2 group produces significantly more retracted variants of [y] than the GEN 1 group. This model also shows the lack of an effect based on gender. No other vowel shows a significant inter-generational difference.

<INSERT TABLE 6>

4.3 Age as the main effect

Given substantial acoustic similarity between the two groups and no overlap in the age range included in each group (20-44 for GEN 2, 46-87 for GEN 1), one might ask if the retraction follows a shift initiated among GEN 1 speakers. In other words, could the retraction of [y] be an internally motivated change that started among GEN 1 speakers rather than a contact-induced change initiated by GEN 2 speakers. Similarly, could there be other internally-motivated changes that are not evident from models that include generational group as a categorical fixed effect?

To address these two questions, mixed effects models were run with the same factors as for the ones run to produce the results shown in **Table 6** but with “age” included as a continuous fixed effect instead of “group” as a categorical fixed effect. The model for the F2 of [y] did not come out significant, which suggests that the retraction of [y] is not an internally motivated change initiated by GEN 1 speakers. The only model that came out significant was the model for the F2 of [i]. This model (details shown in Table 7) shows an inverse relationship between age and the F2 of [i], which means the younger the speaker, the more likely they are to produce [i] with higher F2 (more fronting). Thus, unlike the retraction of [y], the fronting of [i] may be an internally-motivated change initiated by GEN 1 speakers. Gender, however, does not show a significant effect for the fronting of [i].

<INSERT TABLE 7>

To address whether Homeland speakers show evidence of the same changes in apparent time, a set of mixed effects models with F1 or F2 as the dependent variable, speaker and word as random effects, and age as the fixed effect were run on the Hong Kong data. The results from these models are summarized in Table 8. The model of “age” as a dependent variable for the F2 of [y] did not come out significant. Since the retraction of [y] is not an apparent time change in Hong Kong, the retraction of [y] in Toronto seems more likely to be a contact-induced change. The fronting of [i], identified above as an apparent time change in Toronto, is also an apparent time change in Hong Kong. As shown in Table 9, there is an inverse relationship between age and F2 values which means that younger speakers produce [i] with higher F2 (more fronted articulations). The other models of “age” as a continuous fixed effect that came out significant were the models for the F1 of [ɪ] (Table 10), the F1 of [ɔ] (Table 11), F2 of [ɔ] (Table 12), and the F1 of [ʊ] (Table 13). All of them show an inverse relationship between age and formant values. This means that younger speakers produce vowels with higher F1 (articulatorily lowered) and higher F2 (fronted). Thus, lowering of [ɪ] (Table 10), lowering of [ɔ] (Table 11), fronting of [ɔ] (Table 12), and lowering of [ʊ] (Table 13) are all apparent time changes in the HK group. These directions of movement are summarized in Figure 4.

<INSERT TABLE 9>

<INSERT TABLE 10>

<INSERT TABLE 11>

<INSERT TABLE 12>

<INSERT TABLE 13>

5. Discussion

To summarize, the results presented above show:

1) Only one inter-generational change in Toronto: the retraction of [y]
2) Only one age-based change in Toronto: the fronting of [i]
3) Five age-based changes in vowel formant values in Hong Kong: fronting of [i], lowering of [ɪ], lowering of [ɔ], fronting of [ɔ], and lowering of [ʊ]

In terms of vowel shifting, it is clear that the HK group is the more innovative group. Furthermore, all of the age-based changes observed in Hong Kong appear to be consistent with Labov's Principles of Vowel Chain Shifting. For instance, Principle I states that "tense vowels rise along a peripheral track" (Labov, 1994, p. 176). The triangular shape of the vowel space means that vowel raising co-occurs with vowel fronting. This triangle is described in Labov (1994, p. 177) and illustrated for changes in Cantonese in Figure 4. The fronting of [i], a tense vowel, would thus be consistent with Principle I. Principle II states that lax vowels lower and that is exactly the direction of movement shown by the two high lax vowels, [ɪ] and [ʊ]. Finally, Principle III states that back vowels move to the front. The fronting of [ɔ] observed would be consistent with Principle III. The simultaneous lowering movement coincides with a downward movement along that side of the triangular vowel space as illustrated in Figure 4.

<INSERT FIGURE 4>

It is worth noting that the fronting of [i] and the lowering of [ɪ] observed in the HK group result in an increase in the acoustic distinction between these two vowels. An increasing distinction between [i] and [ɪ] was also observed among GEN 2 speakers in Tse (2016b). With a larger set of data in the current study, the fronting of [i] is also observed among Toronto speakers based on age, but no change is observable in [ɪ]. Thus, even without the lowering of [ɪ], the fronting of [i] in Toronto appears to reflect internally motivated tendencies in the language rather than contact-induced change. Furthermore, the increasing acoustic distance between [i] and [ɪ] reported in Tse (2016b) does not appear to be unique to Toronto speakers.

In contrast to the HK group, Toronto speakers appear to be relatively conservative. The only age-based change observed is the fronting of [i]. When modeling generational group rather than age as the main effect, the only change observed is the retraction of [y]. This is a change that is not consistent with Principle III. This and the fact that [y] retraction is a change based on generational group rather than on age suggests that it may be a contact-induced change influenced by Toronto English. While Cantonese has two high round tense vowels, Toronto English only has one. This phonological difference could mean that some GEN 2 speakers are collapsing the Cantonese distinction between [y] and [u]. Thus, although Hong Kong speakers may be more innovative in terms of vowel shifting, Toronto speakers may be more innovative in terms of structural change influenced by English phonology.

6. Conclusion

To conclude this paper, the results show evidence for only one inter-generational vowel shift among Toronto speakers: the retraction of [y]. A model of "age" as a continuous fixed effect showed that the retraction of [y] is not likely an internally motivated change initiated by GEN 1 speakers. Rather, it seems more likely to be a contact-induced change influenced by Toronto English. The lack of the same change in the Hong Kong data further supports a contact-induced change account. The only other change identified in the Toronto data was the fronting of [i] among younger speakers, which was also a change identified in the HK data. Thus, the fronting of [i] seems to be a change already in progress among GEN 1 speakers that has been further advanced by younger Toronto speakers. The results also show more innovative vowel shifting in apparent time in the HK group than in the Toronto groups. Other changes observed in the HK data that are absent in the Toronto data include the lowering of the lax vowels [ɪ] and [ʊ] as well as the lowering and fronting of [ɔ]. A graphical summary of all Toronto and Hong Kong vowel shifts reported is presented in Figure 4.

While it is clear that younger Hong Kong speakers are leading in vowel shifting, what is not clear is how these changes interact with other social factors. For instance, could these changes be led by women or by speakers from a middle-class background? The sample of Hong Kong speakers (n=8) included in this study is too small to be able to address such factors. Future studies of Hong Kong Cantonese with a larger set of speakers would be worthwhile to pursue. In the Toronto data, generational group accounts for the retraction of [y], a change that appears to be contact-induced. Could there be other social factors that facilitate this change within the GEN 2 group? For instance, could it be the case that speakers who use more English or who identify more with mainstream Canadian culture be the ones who lead in retracting this vowel?

Other factors not considered in this paper are phonetic context and tone. These linguistic factors present additional complications to the study. For example, Cantonese has co-occurrence restrictions that limit the possible tonal categories that occur in a given syllable based on phonetic environment. This makes tone and phonetic environment difficult to completely separate. Furthermore, there may be a different story of how tone and phonetic environment condition variation for each of the 11 vowels examined. Phonetic context for [ɛ] and [ɔ], however, is addressed in Tse (2019). For the purpose of this paper, the goal has simply been to establish whether or not there are any vowel shifts in Hong Kong or in Toronto Cantonese. Having established that there are indeed geographical differences in Cantonese vowel production patterns in two cities, this paper, thus, sets the groundwork for identifying specific parts of the vowel space worth further investigation in both Toronto and in Hong Kong.

Acknowledgements

The HLVC Corpus is funded by a research grant from the Social Sciences and Humanities Research Council of Canada 410-2009-2330 (2009-2012) and is developed by Naomi Nagy, Alexei Kochetov, Yoonjung Kang, and James Walker. I would like to thank James Walker for organizing the Melbourne workshop and for putting this special volume together, feedback from audience members at the workshop and from the two anonymous reviewers, all of those involved with the HLVC Project, and my dissertation committee. The results presented in this paper are a

subset of results presented in my dissertation, which is available at: http://d-scholarship.pitt.edu/35721/.

References

Barrie, M. (2003). Contrast in Cantonese vowels. *Toronto Working Papers in Linguistics*, *20*, 1–19.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program] (Version 6.0.15). Retrieved from http://www.praat.org/

Census and Statistics Department, Hong Kong Special Administrative Region. (2016). *Thematic Household Survey Report No. 59* (p. 171). Hong Kong. Retrieved from https://www.statistics.gov.hk/pub/B11302592016XXXXB0100.pdf

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, *39*(3), 192–193.

Johnson, D. E. (2009). Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass*, *3*(1), 359–383.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: readings in sociolinguistics* (pp. 28–53). Englewood Cliffs, NJ: Prentice Hall.

Labov, W. (1994). *Principles of linguistic change, Volume 1*. Oxford, UK; Cambridge, MA: Blackwell.

Labov, W. (2011). *Principles of linguistic change, cognitive and cultural factors* (Vol. 3). John Wiley & Sons.

Lau, C. (2003). Labovian Principles of Vowel Shifting Revisited: The Short Vowel Shift in New Zealand English and Southern Chinese. In *Historical Linguistics 2001: Selected Papers From The 15th International Conference On Historical Linguistics, Melbourne, 13-17 August 2001, Blake, Barry J., & Burridge, Kate [Eds]* (pp. 293–301). Amsterdam: John Benjamins.

Lee, T. (1983). The vowel system in two varieties of Cantonese. *UCLA Working Papers in Phonetics*, *57*, 97–114.

Matthews, S., & Yip, V. (2011). *Cantonese: A Comprehensive Grammar*. Routledge.

Nagy, N. (2011). A Multilingual Corpus to Explore Variation in Language Contact Situations. *Rassegna Italiana Di Linguistica Applicata*, *43*(1/2), 65–84.

Peters, A. & Tse, H. (2016). Evaluating the efficacy of Prosody-lab Aligner for a study of vowel variation in Cantonese, presented at Workshop on Innovations in Cantonese Linguistics (WICL-3), The Ohio State University, Columbus, OH. http://d-scholarship.pitt.edu/27237/

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (pp. 816–820).

Statistics Canada. (2017). *Toronto, C [Census subdivision], Ontario and Canada [Country] (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001*. Ottawa. Retrieved from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E

Thomas, E., & Kendall, T. (2007). *NORM: The vowel normalization and plotting suite*. Retrieved from http://lingtools.uoregon.edu/norm/norm1.php

Tse, H. (2016a). Contact-induced splits in Toronto Heritage Cantonese mid-vowels. *Linguistics Atlantica*, *35*(2), 133-155.

Tse, H. (2016b). Variation and change in Toronto heritage Cantonese: An analysis of two monophthongs across two generations. *Asia Pacific Language Variation*, 2(2), 124-156.

Tse, H. (2019). *Beyond the Monolingual Core and out into the Wild: A Variationist Study of Early Bilingualism and Sound Change in Toronto Heritage Cantonese.* Unpublished Doctoral Dissertation, University of Pittsburgh, Pittsburgh, PA.

Yip, M. (1996). Lexicon Optimization in Languages without Alternations. *Rutgers Optimality Archive*. https://doi.org/10.7282/T33X84P0

Yu, D. (2000). *The Underlying Representation of Cantonese Vowels* (Unpublished Undergraduate Honors Thesis). Stanford University, Stanford, CA.

Yue-Hashimoto, A. O. (1972). *Phonology of Cantonese.* Cambridge [England]: Cambridge University Press.

Yue-Hashimoto, A. O. (1991). The Yue dialects. *Languages and Dialects of China. Journal of Chinese Linguistics Monograph Series*, *3*, 294–324.

Zee, E. (1999). Chinese (Hong Kong Cantonese). In *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet* (pp. 58–60). Cambridge: Cambridge University Press.

Zee, E. (2003). Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1117–1120). Universitat Autònoma de Barcelona Barcelona. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_1117.pdf

TABLES

Table 1. Cantonese tense vowels

Front		Central	Back
Unrounded	Rounded		
i	y		u
ɛ	œ		ɔ
		a	

Table 2. Cantonese lax vowels

Front	Central	Back
ɪ		ʊ
	ɵ	
	ɐ	

Table 3. List of speakers analyzed

	Male	Female	Total
Hong Kong	CXM20A CXM27A CXM52A	CXF16A CXF19A CXF43A CXF49A CXF77A	= 8
GEN 1	C1M46A C1M52A C1M52B C1M59A C1M61A C1M87A	C1F50A C1F50B C1F54B C1F58A C1F78A C1F83A	= 12
GEN 2	C2M21B C2M21C C2M21D C2M22A C2M27A C2M44A	C2F20A C2F21B C2F21C C2F22A C2F24A C2F41A	= 12
	= 15	= 17	TOTAL = 32 speakers

Note. The speaker codes indicate the following: “C” = Cantonese, “1” or “2” = Generation Group (“X” = Hong Kong Group), “M” or “F” = Male or Female, Two-digit number indicates age, Last character is used to distinguish between multiple speakers with same demographic characteristics (i.e., A, B, C, etc.)

Table 4. Percentage of total tokens for each vowel

Vowel	GEN 1	GEN 2	HK
[a]	20.57%	18.46%	19.82%
[ɔ]	**14.88%**	16.78%	**16.69%**
[ɐ]	**14.78%**	16.18%	**18.10%**
[i]	13.66%	14.69%	9.00%
[ʊ]	8.02%	8.61%	8.45%
[ɛ]	7.36%	8.14%	7.32%
[ɪ]	6.55%	6.09%	6.61%
[œ]	5.39%	3.90%	5.75%
[y]	4.04%	3.42%	3.68%
[u]	**2.82%**	**1.61%**	2.74%
[ɵ]	**1.92%**	**2.12%**	1.84%

Table 5. Is "city" a significant predictor of variation for each vowel/formant pair?

	Separate models for each vowel/formant pair For each model Random Effects: Speaker and Word Fixed Effect: City (Hong Kong vs. Toronto)	
	Is city significant for F1 as dependent variable?	Is city significant for F2 as dependent variable?
[y]	**	n.s.
[ɐ]	n.s.	*
[a]	n.s.	n.s.
[ɛ]	n.s.	***
[ɵ]	n.s.	n.s.
[i]	n.s.	*
[ɪ]	*	n.s.
[ɔ]	**	n.s.
[œ]	n.s.	n.s.
[u]	***	n.s.
[ʊ]	***	n.s.

Table 6. Best step-down model for the F2 of [y] for Toronto groups only

Random Effects ($R^2 = 0.332$)					
Speaker, Word					
Fixed effects ($R^2 = 0.048$)					
		Coef.	N	Mean (Hz)	*p*-value
Group					< .05*
	GEN 1	24	623	1631	
	GEN 2	-24	351	1607	
Sex					n.s.

Note. R^2 [total] = 0.380

Table 7. Best step-down model for the F2 of [i] with GEN 1 and GEN 2 data included

Random Effects ($R^2 = 0.290$)			
Speaker, Word			
Fixed effect ($R^2 = 0.050$)			
	(Hz)	Coef.	*p*-value
Age (20-87)	+1	-1.067	< .05*
Sex			n.s.

Note. R^2 [total] = 0.340

Table 8. HK results for "age" (continuous) as significant predictor of variation

	Separate models for each vowel/formant pair Only Hong Kong data included in these models <u>For each model</u> Random Effects: "Speaker" and "Word" Fixed Effect: "Age" (continuous), Age range: 16-77

	Significant Predictor for F1?	Significant Predictor for F2?	Direction of Change (if applicable)
[y]	n.s.	n.s.	--
[ɐ]	n.s.	n.s.	--
[a]	n.s.	n.s.	--
[ɛ]	n.s.	n.s.	--
[ɵ]	n.s.	n.s.	--
[i]	**n.s.**	*****	**fronting**
[ɪ]	*****	**n.s.**	**lowering**
[ɔ]	******	******	**Fronting/lowering**
[œ]	n.s.	n.s.	--
[u]	n.s.	n.s.	--
[ʊ]	*****	**n.s.**	**lowering**

Table 9. Mixed effects model for the F2 of [i]

Random Effects ($R^2 = 0.126$)			
Speaker, Word			
Fixed effect ($R^2 = 0.035$)			
	Coef.	Coef. (Hz)	*p*-value
Age (16-77)	+1	-0.953	< .05*

Note. R^2 [total] = 0.161

Table 10. Mixed effects model for the F1 of [ɪ]

Random Effects ($R^2 = 0.230$)
Speaker, Word
Fixed effect ($R^2 = 0.054$)

	Coef.	(Hz)	Coef.	*p*-value
Age (16-77)	+1		-0.310	< .05*

Note. R^2 [total] = 0.284

Table 11. Mixed effects model for the F1 of [ɔ]

Random Effects (R^2 = 0.381)			
Speaker, Word			
Fixed effect (R^2 = 0.044)			
	Coef.	Coef. (Hz)	*p*-value
Age (16-77)	+1	-0.298	< .001**

Note. R^2 [total] = 0.425

Table 12. Mixed effects model for the F2 of [ɔ]

Random Effects (R^2 = 0.234)			
Speaker, Word			
Fixed effect (R^2 = 0.017)			
	Coef.	Coef. (Hz)	*p*-value
Age (16-77)	+1	-0.511	< .001**

Note. R^2 [total] = 0.251

Table 13. Model of the F1 of [ʊ]

Random Effects (R^2 = 0.218)			
Speaker, Word			
Fixed effect (R^2 = 0.019)			
	Coef. (Hz)	Coef.	*p*-value
Age (16-77)	+1	-0.171	< .05*

Note. R^2 [total] = 0.237

FIGURES

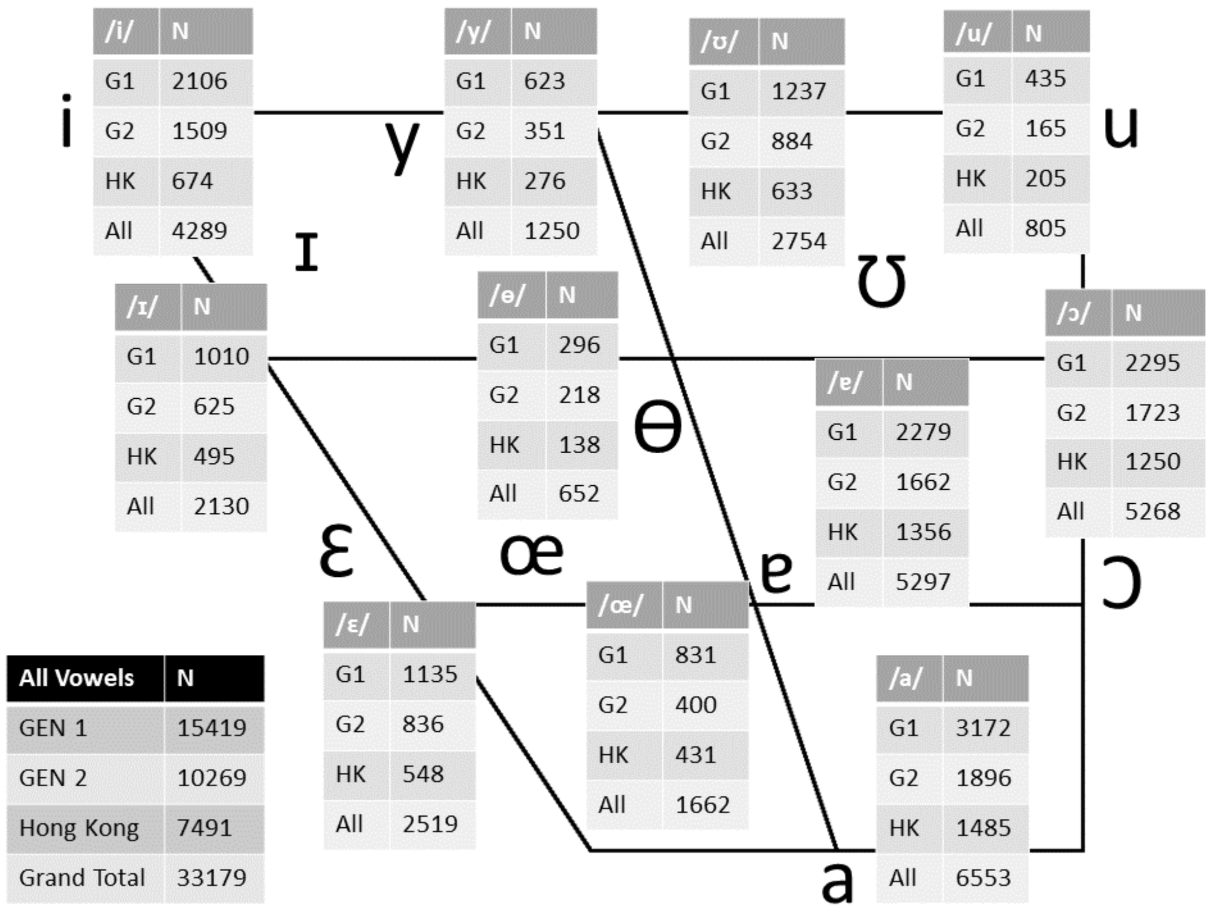
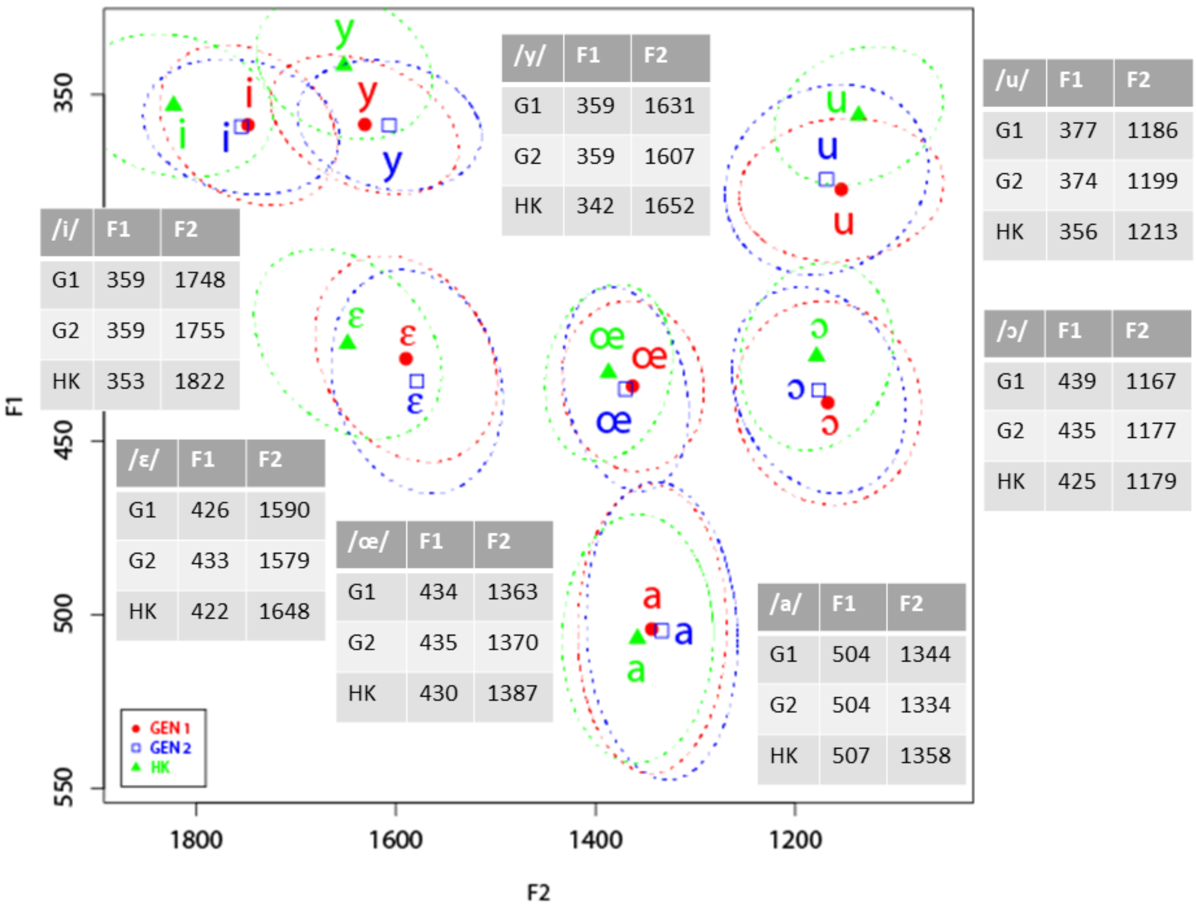


Figure 1. Total number of vowel tokens for each vowel category and for each group



/y/	F1	F2
G1	359	1631
G2	359	1607
HK	342	1652

/u/	F1	F2
G1	377	1186
G2	374	1199
HK	356	1213

/i/	F1	F2
G1	359	1748
G2	359	1755
HK	353	1822

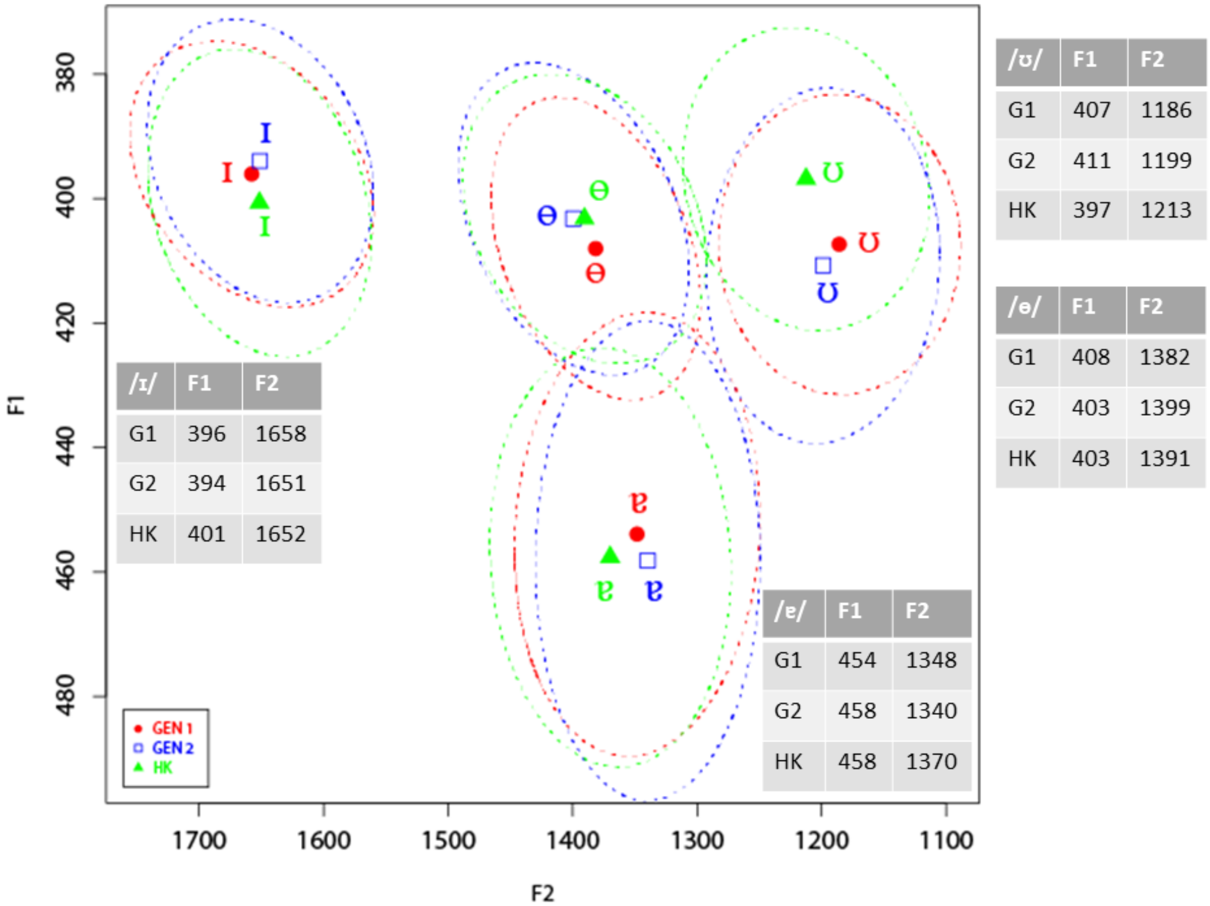
/ɔ/	F1	F2
G1	439	1167
G2	435	1177
HK	425	1179

/ɛ/	F1	F2
G1	426	1590
G2	433	1579
HK	422	1648

/œ/	F1	F2
G1	434	1363
G2	435	1370
HK	430	1387

/a/	F1	F2
G1	504	1344
G2	504	1334
HK	507	1358

Figure 2. Tense vowels, 32 speakers, n=22,346. Ellipses indicate mean F1/F2 ± 1 SD



/ɪ/	F1	F2
G1	396	1658
G2	394	1651
HK	401	1652

/ʊ/	F1	F2
G1	407	1186
G2	411	1199
HK	397	1213

/ɵ/	F1	F2
G1	408	1382
G2	403	1399
HK	403	1391

/ɐ/	F1	F2
G1	454	1348
G2	458	1340
HK	458	1370

Figure 3. Lax vowels, 32 speakers, n=10,833. Ellipses indicate mean F1/F2 ±1 SD (rounded to the nearest Hz)

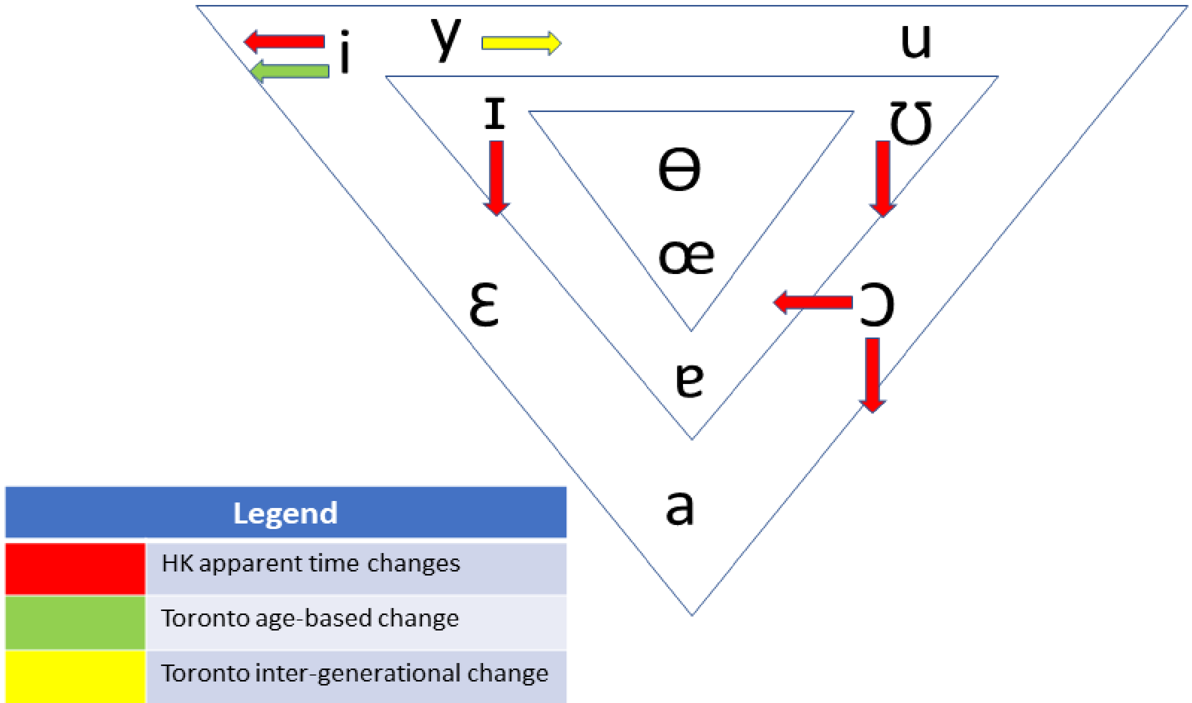


Figure 4. Summary of changes identified in data