
A SHORT INTRODUCTION TO THE PERCEPTION-PRODUCTION ASYMMETRY OF CANTONESE TONE MERGERS

LANGUAGE PROCESSING

Man Ho Wong

Department of Neuroscience
Department of Linguistics
The University of Pittsburgh
Pittsburgh, PA, USA.
m.wong@pitt.edu

December 19, 2023

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords Cantonese · Tones · Tone merger · Perception · Production · Language processing · Speech processing · Phonology · Sound change · Psycholinguistics · Cognitive linguistics

1 Introduction

The phonological tone is an interesting linguistic phenomenon with characteristics distinguished from other phonological segments such as consonants and vowels. Tones have been shown to be autonomous to the segments they are associated with (i.e., the underlying representation and operation of tones are independent of the associated segments). This suggests that tone and other segments have distinct structures of mental representation. While the emergence of tones in some languages is not completely understood, it has been shown that tonogenesis in some languages was likely induced by contacts with other non-phylogenetically related tonal languages in neighboring areas (e.g.,), and that tonogenesis and tone exodus (loss of tones) could happen in a relatively short period of time (e.g., Korean). Such an areal feature of tone and its relatively high fluidity compared to other segments raise the question of how the brain represents and processes tones and other segments differentially.

The unique characteristics of tone discussed above make tonal languages promising targets for studying speech processing and gaining a more comprehensive understanding of how language works. Indeed, the

study of tonal languages revolutionized the traditional approach to segment analysis, leading to the development of autosegmental phonology.

Cantonese is a tonal language spoken in the Guangdong Province of China, Hong Kong, Macau, and many other Han Chinese communities around the world. Compared to other equally-resourced tonal languages, it is particularly well suited for the study of the cognitive processing of tones since most native Cantonese speakers have little metalinguistic awareness of Cantonese tones¹, and therefore, expectedly, less influenced by top-down processing in tone perception.

It is widely accepted that standard Cantonese² has six canonical tones. However, it has been observed recently that some of the tones seem to be merged and are no longer distinguished among younger Cantonese speakers in their perception or production. In this paper, I will first summarize the current findings about tone merging in Cantonese. Using Cantonese tone merging as an example, I will also review some of the proposed mechanisms to shed light on an interesting linguistic phenomenon –the asymmetry between perception and production. Tone merging can happen in perception, production, or both language modalities. Different studies may use different wordings to describe their findings.

To facilitate discussion, I will first define a few terms that I will use in the following text. First, merging participants exhibit tone-merging behavior, regardless of the language modalities involved. Non-merging participants are those who are capable of distinguishing all target tones (tones under investigation) in both language modalities. Second, because perception and production tasks are normally performed separately, I may also use the terms “listener” and “speaker” instead of “participant” when I mention the experiment results from perception and production tasks to avoid confusion between the two types of tasks.

2 Tone merging and sound change

In the current literature, studies on Cantonese tone merging mainly look at two different cases of tone merging: the merging of two canonical tones and the merging of a canonical tone and a morphological tone (i.e., a tone that has morphological functions). The former case reflects a recent ongoing sound change in Cantonese, particularly among the younger population in Hong Kong; It is purely phonological without affecting the meaning of a word. The latter case is observed between a canonical tone and a *pinjam* 變音 (i.e., “changed tone”) which reflects a morphological change (e.g., nominalization of a verb); It does not involve changes of the phonological system of the language and is generally not considered a sound change.

Despite their different roles in Cantonese phonology, the two cases share some similarities in language processing (e.g., a loss of contrast only in perception but not in production can happen in both cases). Currently, the former case is more often studied in cognitive or psycholinguistics probably due to simpler and more straightforward experimental designs (e.g., more minimal pairs available), but the latter case allows us to study the mechanisms of tone merging from an alternative perspective. In this paper, most of

¹Most native Cantonese speakers have difficulties in categorizing syllables based on lexical tones, regardless of their ability to distinguish tones. This is likely due to the lack of formal Cantonese education: Even in Hong Kong where Cantonese is the de facto official language, essentially no native Cantonese speakers have received formal education on the Cantonese language in school (except for students who study Cantonese as a linguistics subject in university). In China, including Cantonese-speaking regions, Mandarin Chinese is the only Chinese language taught in school. In Hong Kong, while Cantonese is the teaching medium of the Chinese language in most schools, students only study the official variety of written Chinese (i.e. written Chinese based on Mandarin syntax). As a result, most native Cantonese speakers do not possess the same level of linguistic knowledge about tones in their native language compared to other tonal language speakers who study their native language formally in school, such as Mandarin speakers (e.g. most Mandarin speakers can categorize tones in Mandarin without much effort).

²Currently, two different dialects of Cantonese are both considered the de facto standard form of Cantonese: the Guangzhou dialect spoken in Guangzhou and the Hong Kong dialect spoken in Hong Kong. The two varieties are nearly identical phonologically, but the vocabulary of each of the varieties is influenced by Mandarin and English respectively. At the moment, Hong Kong is the only place where Cantonese is recognized as a working language and widely used in different settings such as education, literature, and courts, while Guangzhou Cantonese is mainly used between friends and family. This makes Hong Kong Cantonese a good study target for Cantonese as most of its speakers are able to express their ideas on different topics fully in Cantonese.

the discussion centers around the merging of two canonical tones, but I will also discuss the insights we gained from the study of a canonical tone and a morphological tone in the last session.

3 Classifying tone mergers by perception and production

As mentioned earlier, standard Cantonese has six canonical tones, namely T1 to T6. The most common types of canonical tone mergers are between T2 vs. T5, T3 vs. T6, and T4 vs. T6, since these tone pairs are easily confused (in production and/or perception) as shown in Mok et al. (2013). Note that the tones in each of the above tone pairs are not equally confused in both directions: Mok et al. (2013) showed that in production tasks, T2 is misclassified as T5 more often than T5 is misclassified as T2. Similarly, T4 is misclassified as T6 more often than vice versa. In addition, tone merging does not occur in all words where T2 or T5 is present, suggesting that some words are more susceptible to merging. Together, these results showed that merging speakers (i.e. speakers who display tone merging in their speech) still maintain distinct tonal categories for tones in a tone pair, as different words are subjected to different degrees of merging. To explain why different degrees of merging are found in different words, Mok et al. (2013) hypothesized that word frequency may play a role in tone merging. A brief review of the effects of word frequency on tone merging can be found in Mok et al. (2013).

Among the above tone pairs, different accuracies in perception and production have been observed in different individuals. Mok et al. (2013) found that merging speakers had less separated tone space than non-merging speakers (i.e., tones produced by merging speakers were acoustically more similar). Interestingly, both merging and non-merging participants had the same discrimination accuracy in the perception tasks, although the merging group showed significantly longer reaction time than the non-merging group. Mok et al. (2013) suggested that the seemingly intact perceptual ability observed in the merging group could be explained by the experimental setting: participants could have paid more attention to subtle acoustic cues in a forced-choice AX discrimination task than they would in a naturalistic environment. Besides, discriminating tones in an experimental setting could be easier than in a naturalistic environment because experimental speech materials are produced carefully and the environment is much less noisy. Nevertheless, longer reaction time still indicates that the merging participants still had perceptual difficulty in tone discrimination.

Perceptual difficulty may lead to production accuracy. However, the asymmetry between perception and production has been found in Cantonese tone merging. With a larger sample size and more sophisticated statistical analyses, Fung and Lee (2019) not only confirmed the occurrence of tone mergers observed in T2-T5, T3-T6, and T4-T6 pairs by other studies, but their findings also indicate that individual variation exists within each of the three tone pairs in terms of the language modalities. The greatest individual variation in perception accuracy was found in the T2-T5 pair, while the T4-T6 pair was the second most varied pair. Moreover, the results showed that individuals had different patterns of asymmetry between perception and production, depending on the tone pairs involved. At the community level, they suggested that these tone pairs can be classified into three types of mergers-in-progress according to the language modalities involved:

- I. The T2-T5 pair is a full-merger where tonal contrast in both perception and production is collapsed;
- II. The T3-T6 pair is a partial-merger where tonal contrast is collapsed only in production;
- III. The T4-T6 pair is a near-merger where tonal contrast is collapsed only in perception.

From a historical linguistics and sociolinguistics perspective, we should keep in mind that the above tone mergers are considered mergers-in-progress (the sound change is not complete) and that not all tone mergers are in the same stage of sound change or will undergo the same sound change. In addition, different speech communities may have different progress in the same sound change³.

³See the discussion of Hong Kong South Asian speech communities in next section

4 The two big questions about Cantonese tone mergers

Cantonese tone merging is an ongoing sound change that has been systematically studied only recently. While the current development of tone merging has been comprehensively described in the literature, several enigmas concerning the phenomenon remain unresolved. Among these, the most fundamental question probably relates to the causes of tone merging. Another equally important yet more puzzling question concerns the asymmetry between perception and production in the cases of partial-merger and near-merger.

4.1 What causes tone merging?

The merging of two canonical tones may arise from similar sources as other sound changes. As discussed by Mok et al. (2013), a sound change is often driven by an automation of speech production (e.g. deletion or weakening of speech sounds as phonetic “shortcuts”); it may also originate from perceptual difficulties or imperfect learning. In the case of Cantonese tone merging, a tone is simply replaced by another tone without deletion or weakening, i.e. no automation of speech production is involved. On the other hand, since the tones in a tone merger have very similar F0 (fundamental frequency) values, this increases the perceptual difficulty in tonal discrimination and, as a result, contributes to imperfect learning. Additionally, results from the study of tone merging in South Asian speech communities in Hong Kong indicate that the different perceptual abilities of members from different speech communities may contribute to the diverse tone merging progress observed in different communities. Among South Asian speech communities, participants whose dominant language is Punjabi, a language that also has lexical tones like Cantonese, had the lowest performance in perception tasks. This can be explained by both the Perceptual Assimilation Model (PAM) or the Speech Learning Model (SLM), which suggest that a learner’s L1 (first language) to L2 (second language) linguistic distance inversely influences their ability to maintain the contrast between speech sound categories in L2. Overall, these studies highlight the potential roles of perceptual difficulties and imperfect learning as contributing factors to tone merging.

4.1.1 Insights gained from the South Asian speech communities

As discussed briefly in the previous section, the South Asian speech communities in Hong Kong present a unique opportunity to examine Cantonese tone merging as a sound change. Hong Kong has a sizable ethnic South Asian population mainly consisting of people of Indian, Pakistani or Nepali descent. In addition to their heritage languages, many South Asians are also proficient in the official languages in Hong Kong, Cantonese⁴ and English. The South Asian speech communities in Hong Kong have several distinct characteristics that make them an ideal research subject for studying Cantonese tone merging: First, since Cantonese proficiency is not mandatory in education⁵, many South Asians acquire Cantonese through natural exposure (as a first or second language) with minimal or even no Cantonese education. Therefore, studying tone merging in South Asian speech communities offers us valuable insights into the specific role of perception in Cantonese tone merging; Second, most South Asian Cantonese speakers speak another non-tonal language (e.g., Hindi and Urdu) while some speak Punjabi, an Indo-European language with lexical tones. Such diverse linguistic backgrounds provide further details about tone perception and production; Moreover, in contrast to ethnic Chinese Cantonese speakers, a significant proportion of South Asian Cantonese speakers do not possess written Chinese literacy. Consequently, their tone production is expected to be much less influenced by Chinese orthography.

In a study of Cantonese tone perception and production by South Asians, Mok et al. (2018) found that South Asian participants performed significantly worse than ethnic Chinese participants in perception tasks, despite a similar pattern of discrimination between Cantonese tone pairs observed in both groups, regardless

⁴While the variety of Chinese language is not specified officially, Cantonese is the de facto working language.

⁵Most schools use either Cantonese or English as the medium of instruction; English is the dominant language in higher education.

of their dominant languages. Regarding production, the tone space appeared to be much smaller in South Asian Cantonese speakers than in ethnic Chinese Cantonese speakers, indicating less tonal distinction in production. Not only do these findings reveal that Cantonese tone merging also happens in the South Asian speech communities, but the same sound change also progresses at a faster pace. This suggests that while tone merging is influenced by some socio-linguistic factors, the causes of this sound change are likely associated with fundamental language processing functions regardless of one's socio-linguistic background.

4.2 What contributes to the asymmetry between perception and production?

As discussed earlier, Cantonese tone mergers can be classified as full-merger, partial-merger or near-merger depending on the language modalities involved. Full-merger is less of a problem as there is an obvious link between perception accuracy and production accuracy. However, the causes of asymmetry between perception and production in partial-merger (where contrast is collapsed only in production) and near-merger (where contrast is collapsed only in perception) are still not understood. The asymmetry in near-merger is particularly puzzling as it is counterintuitive.

4.2.1 Partial-merger

Lin et al. (2021) specifically looked at the T3-T6 partial-merger. They showed that partial-merger is reversible via phonetic imitation. In their study, participants (i.e. shadowers) were asked to repeat after a model talker (a person who produces T3 and T6 distinctively). As expected, the merging shadowers were able to discriminate T3 and T6 with high accuracy. They were also able to produce the two merging tones with enlarged distinction by imitating the model talker's production immediately after they heard the sounds. This finding not only shows that partial-merger is reversible, but it also provides a piece of evidence supporting the exemplar theory: a model talker's production may be added to the merging shadower's categorical system as new exemplars, and therefore changes the calculation of production targets. In addition, since the tonal distinction in merging shadowers' production diminished after the shadowing experiment, this indicates that the newly added exemplars during the shadowing experiment are not stored in the categorical system. In other words, this study suggests that the asymmetry between perception and production in partial-merger could be explained by the lack of proper exemplars for T3-T6 distinction.

4.2.2 Near-merger

The near-merger is a more complex phenomenon than the partial-merger. It has been studied from both linguistic and cognitive perspectives. A neuroimaging study investigated the information flow in the speech processing of near-merger by measuring the mismatch negativity (MMN) of participants during experiments. MMN is a component of the event-related potential (ERP) associated with a pre-attentive comparison process in speech processing. It has been used to study sensitivity to speech sound contrasts. A greater MMN amplitude and a shorter peak latency have been linked to larger tonal contrasts among Mandarin and Cantonese speakers. Using the T1-T6 tone pair as a control (there is no dissociation between perception and production), they found that the MMN was only present in response to the lexical syllables of T1-T6 but not the lexical syllables of T4-T6 in the merging participants. In the non-merging participants, they found MMN responses to the lexical syllables of T4-T6 but not to the nonlexical syllables of the same tone pairs. These results suggest that top-down processing may be involved in the processing of near-merger, providing a physiological explanation to the reduced sensitivity to T4-T6 contrast in the merging group.

Going back to a more linguistic perspective, near-merger has also been studied in non-canonical tones. Yu (2007) studied the merging of a canonical tone and a morphological tone: the canonical rising tone, T2, (in the paper: lexical mid-rising tone, R^{lex}) and the morphologically-derived mid-rising tone (R^{morph}). R^{morph} is the changed tone (*pinjam* 變音) that a word acquired after a morphological change. For example:

- (1) Nominalization of the verb “to sweep” (掃)

/sou³³/ → /sou³⁵/ “to sweep” → “a broom”

In example (1), the tone of the target word changes from its lexical tone (i.e. its base tone, [33]) to its R^{morph} , which is realized as the same tone contour as the canonical T2[35] (R^{lex})⁶. However, Yu (2007) showed that the two rising tones actually have different F0 values: R^{morph} has a higher overall F0 profile than R^{lex} despite having the same contour. Moreover, the study also found that listeners performed poorly in discriminating R^{lex} and R^{morph} , although they were able to produce the two rising tones differentially. The result indicates that R^{lex} and R^{morph} belong to a near-merger as the merging happens only in perception but not in production. To understand why there is an asymmetry between perception and production in this tone pair, we need to first understand what the origin of *pinjam* is and why R^{morph} has a higher F0 profile than the corresponding canonical tone T2.

It has been proposed that *pinjam* originated from the loss of certain diminutive morphemes through syllable contraction. For example:

- (2) chess piece–DIM (棋子) ~ chess piece (棋)
/k^hɛi²¹ tsi³⁵/ ~ /k^hɛi³⁵/

In example (2), the diminutive morpheme (DIM) /tsi/ is elided in the process. However, its rising tone [35] remains and is reassociated with the preceding morpheme /k^hɛi/, replacing the preceding morpheme’s base tone [21]. According to this hypothesized origin of *pinjam*, it is possible that other morphemes with a higher tone (e.g. [55]) could have also contributed to the development of *pinjam*. The higher F0 profile of R^{morph} could be the result of the addition of exemplars with *pinjam* to the exemplar space of the base tone. To test this possibility, Yu (2007) induced syllable-contraction experimentally by asking the participants to read the target phrases

- (3) a. verb + POT /tək[˥]⁵⁵/ (得)
b. verb + PFV /tsɔ³⁵/ (咗)

(i.e. verb + potential marker 得 /tək[˥]⁵⁵/ or verb + perfective marker 咗 /tsɔ³⁵/) in the carrier phrase with “lazy speech style” (i.e. not carefully articulated). To summarize, they found that the fused syllables in the target phrases induced by the experiment have different F0 profiles, depending on the modal markers involved. The tone of the fused syllable in verb + potential marker /tək[˥]⁵⁵/ has a much higher F0 profile than the *pinjam* R^{morph} , while that in verb + perfective marker /tsɔ³⁵/ has a similar F0 profile as R^{lex} (i.e. T2 [35]). From the experimental data, they hypothesized that *pinjam* may originate from the addition of exemplars with different tones to the exemplar space of a word. If the production target for the tone is the averaged tone of all the exemplars in the exemplar space, the F0 distribution of the resulting tone (R^{morph}) would be higher than the canonical R^{lex} but lower than the highest tone one can find in the exemplars. On the other hand, as the F0 distribution of R^{morph} overlaps with that of R^{lex} in the example space, they cannot be distinguished easily in perception. The above hypothesis not only helps us to understand the historical origin of *pinjam*, but it also provides a possible explanation for the asymmetry of perception and production observed in near-merger.

5 Summary

This paper summarized the current findings about tone merging in Hong Kong Cantonese. In the reviewed studies, the most common types of tone mergers in Cantonese are T2-T5, T3-T6, and T4-T6. In addition to the merging of canonical tones, Yu (2007) also showed that merging also occurs between a lexical tone

⁶The tone contour for mid-rising tone is analyzed as [35] in this paper, but it may also be analyzed as [25] and termed “high-rising tone” in other studies.

and a morphologically derived tone. Current studies have also found that different language modalities are involved in different tone mergers: the T2-T5 full-merger where tonal contrast in both perception and production is collapsed; the T3-T6 partial-merger where tonal contrast is collapsed only in production; and the T4-T6 near-merger where only perception is affected. Different proposed mechanisms were explored in the paper to address the two major puzzling questions about tone merging: the origin of tone merging and the mechanisms contributing to the asymmetry of perception and production in partial- and near-merger. Unfortunately, no definite answers are provided to these questions. More studies will be needed to uncover the underlying mechanisms of tone merging, especially in the case of near-merger. Since tone merging possibly involves multiple linguistic mechanisms, from phonology to sound change and language acquisition to categorization, studying tone merging can help us to understand better these linguistic mechanisms and the interactions between them.

5.1 Citations

Citations use natbib. The documentation may be found at

<http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf>

Here is an example usage of the two main commands (`citet` and `citep`): Some people thought a thing [Kour and Saabne, 2014a, Hadash et al., 2018] but other people thought something else [Kour and Saabne, 2014b]. Many people have speculated that if we knew exactly why Kour and Saabne [2014b] thought this. . .

References

- George Kour and Raid Saabne. Real-time segmentation of on-line handwritten arabic script. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 417–422. IEEE, 2014a.
- Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. Estimate and replace: A novel approach to integrating deep neural networks with existing applications. *arXiv preprint arXiv:1804.09028*, 2018.
- George Kour and Raid Saabne. Fast classification of handwritten on-line arabic characters. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 312–318. IEEE, 2014b. doi:10.1109/SOCPAR.2014.7008025.