Laboratory Phonology
Journal of the Association for Laboratory Phonology

Ren, X., & Mok, P. (2025). The effects of contextual tonal variation on Cantonese tone merging. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 16(1), pp. 1–44. DOI: https://doi.org/10.16995/labphon.17746

Open Library of Humanities

# The effects of contextual tonal variation on Cantonese tone merging

**Xinran Ren,** School of Foreign Languages, Sun Yat-sen University, China, renxr3@mail.sysu.edu.cn

**Peggy Mok*,** Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong S.A.R., China, peggymok@cuhk.edu.hk

*Corresponding author.

Previous studies on Cantonese tone merging have examined monosyllabic materials so far, yet disyllabic words are common in daily conversation. Sound change often originates from a pool of synchronic variations, and coarticulation from neighbouring units is a common source. The current study examined how tonal coarticulation in disyllabic words contributes to Cantonese tone merging by examining both monosyllabic and disyllabic data from 17 merging speakers and 2 reference speakers. Materials with well-controlled tonal context for the target tones appearing as the first and the second syllables in disyllabic words were used. Results showed that tonal coarticulation and tone merging coexist, with carryover coarticulation in extreme tonal contexts being the most vulnerable condition for change. Large amounts of individual variation were observed, both in terms of cross- and within-speaker variations which could blur the difference between merging and non-merging speakers. The disyllabic data reveal both the independence and interaction of coarticulation and sound change, and allow us to consider various factors in sound change from a wider perspective at the suprasegmental level.

# 1. Introduction

It is generally accepted that sound change, whether it is segmental (consonants or vowels) or suprasegmental (tone) in nature, usually originates from a pool of synchronic variations, although in the literature there have been more studies on segmental than suprasegmental changes. Since it is impossible to wind back the clock to examine the early stages of sound change development, investigating the synchronic variations of a sound change in progress is a good way to help us understand the processes and mechanisms of sound change in general (Beddor, 2023). One such type of synchronic variations is caused by coarticulation. In natural conversation, speech sounds do not appear in isolation but as parts of syllables and words or even larger units, which means that coarticulation with neighbouring units is common and unavoidable for both segmental and suprasegmental features. Coarticulation is a prominent reason for synchronic variations, and it underlies the phonetics of sound change (e.g., Beddor, 2009; Yu, 2020). Nevertheless, previous studies on the incipient tone merging in Hong Kong Cantonese mainly used monosyllabic materials (e.g., Mok et al., 2013), and did not consider the effects of contextual variation or coarticulation on tone merging. It is conceivable that tonal coarticulation will render the phonetic realizations of the merging tone pairs to be more similar in disyllabic than in monosyllabic words in general, but it is unknown if the degree and extent of contrast reduction is similar across merging speakers and/or merging tone pairs. It is also unknown if the phonetic context and coarticulation direction have any effect on the reduction of tonal contrast. The present study expands the scope of the investigation of tone merging in Hong Kong Cantonese by examining the effects of contextual variation on the merging tone pairs in disyllabic words within well-controlled phonetic contexts. It is hoped that this study can contribute to a better understanding of the relationship between synchronic variation and sound change in progress.

The remainder of the paper is organized as follows. Studies on Cantonese tone merging will be reviewed first. Then the findings of tonal coarticulation in various tone languages and in Cantonese will be discussed before the hypotheses of the current study are explained. Details and the results of a production experiment using real disyllabic words will be presented. Finally, the findings of the current study will be discussed in relation to general issues in sound change.

## 1.1. Tone merging in Cantonese

Cantonese has a complex tone system. There are six lexical tones appearing in open syllables or syllables with nasal endings [-m, -n, -ŋ]: T1 (high-level [55]), T2 (high-rising [25]), T3 (mid-level [33]), T4 (low-falling [21]), T5 (low-rising [23]), and T6 (low-level [22]). The numbers in brackets represent the relative initial and final pitch levels of each tone following Chao (1930, 1947), with 5 being the highest and 1 the lowest pitch level. In addition, there are three short tones appearing in checked syllables ending with unreleased stops [-p, -t, -k]: T7 (high-stopped [5]), T8 (mid-stopped [3]), and T9 (low-stopped [2]), which are considered allotones of the three

corresponding unstopped level tones T1, T3, and T6 respectively (Bauer & Benedict, 1997; Chao, 1947), although acoustically the short stopped tones have a slightly falling contour (Rose, 2004; Wong & Chan, 2018).

The merging of some of the Cantonese tones is a relatively recent phenomenon. The first study documenting such a change was Kej et al. (2002) who reported that some of their Hong Kong Cantonese participants made "tone production errors" as they did not clearly distinguish the two rising tones (T2 [25] and T5 [23]) in their production. Bauer et al. (2003) and Yiu (2009) followed this up from a tone-merging perspective. They found that some speakers produced the two rising tones unconventionally with different possible patterns (T5 [23] →T2 [25]; T2 [25] →T5 [23] and a novel intermediate realization). Yiu (2009) also showed that some participants had perceptual confusion of this tone pair.

Mok et al. (2013) systematically studied the merging patterns of the six Cantonese tones in both production and perception. They screened a large number of young Hong Kong Cantonese participants (169) in order to identify the potentially merging (28) and non-merging control participants (30) for their experiments. In addition to the two rising tones (T2/T5), they also examined two other acoustically similar tone pairs: the level tone pair (T3 [33] and T6 [22]) and the low tone pair (T4 [21] and T6 [22]). Various analyses were used to examine the data, which showed large individual variation in tone production. T1 [55] was stable, merging with no other tone, while some speakers were variably merging T2 with T5, T3 with T6, and T4 with T6 in their production. Using Discriminant Analysis on four measurement points of the F0 contours, they found clear evidence in the misclassification rates that in production the three tone pairs were merging at different rates. The data also demonstrated that the merging of tones was not complete, as most speakers showed only partial overlap (albeit to a substantial degree for some) in the merging tone pairs, and the tones were generally classified correctly above chance level. Their data also showed that the merging of tones was not symmetrical.[1] The misclassification rates revealed that T2 tokens were more often misidentified as T5 than the other way round, and T4 tokens were more often misclassified as T6 than vice versa. The misclassification rates of T3 and T6 were more comparable. Comparing the F0 values at the offsets of the merging tone pairs, they found that the tones produced by the merging participants were acoustically more similar than those of the non-merging participants, i.e., their "tone space" was reduced.

[1] The auditory judgements by two native transcribers instead showed that T2 appeared to be more stable than T5 and T4 to be more stable than T6, in that more T5 and T6 words were judged to be produced with other tones (including intermediate realizations), while T3 and T6 were again of comparable variability. Mok et al. (2013) reasoned that human and machine (discriminate analysis) recognition were based on different sets of data which may explain the discrepancy. Since the current study did not involve auditory judgements, we focused on the misclassification rates. See Mok et al. (2013) for a detailed discussion of the discrepancy.

As for perception, there was no significant difference between merging and non-merging participants in terms of accuracy in an AX discrimination task. However, the reaction time data revealed a different picture. The merging participants were significantly slower than the control participants across the board, and this was not confined to the identified merging tone pairs, which illustrated that the merging participants were careful in the perception task, reflecting their general difficulties with tone perception. Mok et al. (2013) argued that both the production and perception data demonstrated that the merging of tones was at an incipient stage with much individual variation, and that the merging participants still had six tone categories.

Adopting similar experimental designs to Mok et al. (2013), Fung and Lee (2019) investigated the tone production and perception of 120 Hong Kong Cantonese participants, aged 20 to 58 years old. They also found that the three tone pairs were merging at different rates in production and perception among the participants. Their perception data showed that discrimination accuracy of the T2/T5 pair was the lowest (~70%), followed by the T4/T6 pair (~80%), while that of the T3/T6 pair was close to ceiling, as it was for other tone pairs. There were 35% and 18.3% of their participants having difficulty in clearly discriminating the T2/T5 and T4/T6 pairs, respectively. Their production data demonstrated that the T3/T6, T2/T5 and T4/T6 pairs had the lowest Pillai scores (an index of phonetic distinctiveness between the trajectories of the two tones) and highest variability of all the tone pairs. Among the participants, 46.7% had difficulty in clearly distinguishing the T3/T6 pair, 22.5% the T2/T5 pair and 15% the T4/T6 pair. Their data supported that these tone pairs were mergers in progress.[2]

Focusing only on the T2/T5 pair, Li and Guan (2019) collected production data from 50 Hong Kong Cantonese speakers aged between 10 to 88, divided into three age groups (young, middle and senior). They found that the F0 contours of the two rising tones were in distinct patterns, suggesting that the two tones were still in separate categories, echoing Mok et al. (2013). The offset slopes of T2 were changing across age groups: The offsets were closer to those of T1 [55] for the senior group, but they became gentler and in a near parallel contour to T5 in the middle and young groups. Since the three age groups did not differ in T5 offsets and slopes but differed significantly in T2 offsets and slopes, they suggested that any change between the merging tone pair was likely to have started with T2 lowering resulting in a shrinking contrast.

Some studies have investigated the tone-merging phenomena in Hong Kong Cantonese with neurolinguistic data. Using the event related potential (ERP) paradigm with both lexical and non-lexical syllables, Law et al. (2013) showed that T4 [21] and T6 [22] can be considered near-mergers in that some speakers could produce them correctly but could not distinguish them in perception. In a later study, Ou and Law (2016), also using ERP, examined the opposite pattern

[2] It should be noted that Fung and Lee (2019) considered T2/T5 a "full merger," T3/T6 a "partial-merger," and T4/T6 a "near-merger." These terms should only be understood analogically, as their data clearly showed that the participants could distinguish the tone pairs at a better-than-chance level.

of participants having intact production but problematic perception of the two rising tones T2/T5, and compared them with those who had intact production and perception. Their reaction time findings match with those of Mok et al. (2013) in that the participants with problematic perception also had generally longer discrimination latency and significant differences in neural responses from those with intact perception.

Zhang (2019) compared tone merging in three varieties of Cantonese: Hong Kong, Macau and Zhuhai. She found that Zhuhai Cantonese was the most advanced in the merging process, as T2/T5 and T3/T6 were already merged, leaving only four lexical tones (T1 [55], T4 [21], a rising tone and a mid-level tone) across the five age groups in her study. The T2/T5 pair was also merged for the two youngest age groups (16–25, 26–35) in Macau Cantonese, while the acoustic distance between T3 and T6 decreased incrementally from the oldest to the youngest group, with the youngest group mixing the two. Both the T2/T5 and T3/T6 pairs were better distinguished in Hong Kong Cantonese across age groups, albeit with decreasing acoustic distance between the tone pairs for the younger groups. However, Zhang (2019) did not find evidence of the T4 and T6 merge in any of the three Cantonese varieties. Her data indicated that the stages of tone merging of the three varieties were in the order of Zhuhai > Macau > Hong Kong, supporting Mok et al.'s (2013) claim that tone merging was at an incipient stage in Hong Kong Cantonese, while at the same time showing patterns of the other two Cantonese varieties which could be predictive of the possible progress patterns of tone merging in Hong Kong Cantonese.

The above studies demonstrated the variations in Cantonese tone merging in terms of production and perception, different tone pairs, age groups and language varieties. Understandably, these studies all only used monosyllabic materials in their experiment materials as they were early investigations of the tone-merging phenomenon. They have helped us to understand the narrowing contrasts in the merging tone pairs, but it is still unclear how much more variation there would be in the merging patterns of disyllabic words which are very common in natural conversation. The next section first reviews some previous studies on general contextual tonal variations before discussing a few studies on tonal coarticulation specifically in Cantonese.

## 1.2. Contextual tonal variations

There are different types of contextual tonal variation. The focus here is on tonal coarticulation, a phonetic and gradient phenomenon of how the realizations of lexical tones are modified by the neighbouring tones. This is different from the phonological and categorical process of tone sandhi. Tone sandhi is about language-specific obligatory (morpho)phonological processes, while tonal coarticulation is attributed to physical constraints, although some studies have demonstrated that the difference between the two may not always be so clear-cut (Chen & Li, 2016; Sun & Huang, 2015). The tone merging phenomenon can be considered to be midway along the phonetic coarticulation and phonological tone sandhi continuum.

Yang and Xu (2019) reviewed 52 tone change studies on 45 diverse tone languages and found surprisingly strong cross-linguistic tendencies in tone change directionality (clockwise, leveling, and regressing to mid patterns) which, they argued, had an articulatory basis in tonal coarticulation and truncation across multiple syllables in connected speech. They concluded that there was a strong match between tone change trends and the patterns of synchronic tonal variation found in connected speech.

As tonal coarticulation is related to biomechanical constraints, it is not surprising that a number of studies in various tone languages showed some common tonal coarticulation patterns, e.g., Mandarin (Xu, 1997), Taiwan Southern Min (Peng, 1997), Thai (Gandour et al., 1994), Vietnamese (Han & Kim, 1974), Mizo (a Tibeto-Burman language, Sarmah et al., 2015), Vientiane Lao (Yu, 2011) and Yoruba (Laniran, 1992). These studies mainly focused on two aspects: directionality (anticipatory or carryover coarticulation) and the nature of the contextual effects (assimilatory or dissimilatory). Chen et al. (2018) summarized the major findings of previous studies showing that both anticipatory and carryover tonal coarticulation were found, but carryover effects were usually much greater in magnitude than anticipatory effects, and were typically assimilatory. The weaker anticipatory coarticulation, if present, was mostly dissimilatory. They also reported that high and low tones differed in tonal coarticulation, whether they were the target or the trigger. For carryover coarticulation, high tones were better triggers and targets than low tones, while for anticipatory coarticulation, it was more likely for a low tone to trigger anticipatory coarticulation on a previous high tone target (i.e., pre-low raising). Despite these common patterns, Chen et al. also reported that there could be comparable anticipatory and carryover effects in some languages like Nanjing Chinese and Malaysian Hokkien, and that the coarticulatory asymmetry between high and low tones might not be consistent. In addition, Gandour et al. (1994) reported that coarticulation primarily affected tone height in Thai, while tone slope was relatively unaffected, although other studies did not report such an asymmetry.

Several studies have examined tonal coarticulation in Cantonese. Using the disyllabic sonorous non-word sequence /lau lau/ with all possible tone combinations ($6 \times 6 = 36$) produced by four speakers, Wong (2006b) demonstrated that the dissimilatory anticipatory effect was much weaker than assimilatory carryover coarticulation, agreeing with the previous findings discussed above. Wong provided the F0 contours of all six tones as first and second syllables, which showed that the most consistent part of the F0 contour for each tone was in the second half of the target tone. In addition, the level tones were the least susceptible to carryover effect, while the falling tone was the most susceptible. Also using 36 tone combinations on the disyllabic sonorous non-word sequence /jau wai/ but only one professional speaker, Gu and Lee (2007) had similar findings for Cantonese tonal coarticulation as Wong (2006b). Moreover, they found that focus interacted with coarticulatory patterns in that the increase in F0 caused by focus varied with tonal context (it was larger on higher pitch targets) and that anticipatory assimilation was enhanced in the context of focus.

More relevant for our purpose, Li et al. (2020) focused on carryover coarticulation on the T2/T5 merging pair. Twelve disyllabic real words in the form of Tx + T2/T5, i.e., six minimal pairs between T2/T5, were recorded from 23 young speakers. Tx was one of the six lexical tones in Cantonese. Li et al. found an assimilatory carryover effect in that the onsets of both rising tones became significantly higher when preceded by the highest tone T1 [55] and significantly lower when preceded by the lowest tone T4 [21], but no significant difference was found when preceded by other tones which had offsets mostly in the mid pitch range. Interestingly, carryover coarticulation also affected the T2/T5 offsets. When preceded by all six tones, the offsets of T2 were significantly lower than in citation forms, while those of T5 were significantly lower than in citation forms only when preceded by T4. Li et al. argued that the F0 perturbation of T2 could reflect the ongoing sound change in Hong Kong Cantonese as the T2 offsets were lowered by all preceding tones, regardless of tonal context. Tonal coarticulation, including downdrift (Wong, 1999), could be the cause of the flattening of T2 offset contours in connected speech. They called for future studies to investigate how tonal coarticulation is related to tone merging.

## 1.3. The present study

The above review on Cantonese tone-merging patterns and contextual tonal coarticulation illustrates some possible interactions between the two phenomena. First, since Cantonese tone merging is still at an incipient stage, it straddles the border between phonetics and phonology and thus is still likely to be influenced by phonetic processes like tonal coarticulation. Coarticulatory patterns may also reveal how tone change may have developed, e.g., it has been shown that the offset of T2 has become flatter and closer to T5, probably due to tonal coarticulation (Li & Guan, 2019; Li et al., 2020). Second, as reported in many studies on various tone languages including Cantonese, assimilatory carryover coarticulation is stronger than dissimilatory anticipatory coarticulation. It is possible that the merging tone pairs would be more similar (or more merged) when they are in the second syllable of a disyllabic word than when they are in the first syllable. Nevertheless, whether the merging tone pairs in the first syllable would be affected by coarticulation depends on how strong the anticipatory coarticulation is, as anticipatory coarticulation has been reported to be rather weak in various tone languages. In addition to the coarticulatory influence from the neighbouring tones, the generally richer phonetic context of disyllabic words is conducive to less standard production. It is of interest to examine how much more similar the merging tone pairs would be in disyllabic words in comparison to when they are in monosyllables. Third, it has been shown that tonal context may exert coarticulatory forces differently on the target tones. Some tones, especially those with more extreme onsets and offsets, may trigger stronger coarticulation. Some tones are also more susceptible to coarticulation than others. Therefore, syllable position (for coarticulatory direction) and tonal context were systematically manipulated in the current study to explore their effects on tone merging.

A third factor, word frequency, was also included in the current study. The effect of word frequency is an important factor in language change. There are two types of word frequency effect on sound change: Reductive sound change (i.e., changes involving deletion/weakening of speech sounds) tends to affect high-frequency words first because it originates from the automation of speech production, whereas analogical sound change (i.e., changes happening based on an analogy taken from related patterns) usually starts from low-frequency words first as it stems from imperfect learning (Bybee, 2007). There are also two types of word frequency: token frequency and type frequency. Token frequency refers to the number of times a unit is experienced (e.g., how often a certain word appears in a corpus), while type frequency refers to the number of distinct types that exemplify a certain pattern (e.g., how often T2 words appear in natural conversations) (Kapatsinski, 2023). Mok et al. (2013) explored the effects of word frequency on Cantonese tone merging. They found that although token frequency of monosyllables did not have any consistent pattern in the rates of tone change, type frequency seemed to be at play. T5 (with the lowest type frequency) was more variable than T2 (with a high type frequency) for words with both high- and low-token frequency. T3 and T6 have comparable type frequency, and they also had comparable variability in tone production. They argued that tone merging in Cantonese probably should not be regarded as a reductive sound change, because producing T2 instead of T5 or T3 instead of T6 are not "phonetic shortcuts" due to the automation of speech production. Rather, tone merging in Cantonese should be regarded as an analogical sound change resulting from perceptual difficulty and imperfect learning, and thus low type frequency words would be affected more than high type frequency words as low frequency words are under more pressure to conform.

Nevertheless, a different perspective on word frequency effects may be needed for the current study, which focuses on tonal coarticulation of disyllabic words and tone merging. Since tonal coarticulation is a phonetic assimilatory process, and tone merging is still at an incipient stage, reductive sound change may be more prominent than analogical sound change in disyllabic words. It is possible that disyllabic words with high token frequency allow more tonal coarticulation, thus rendering the merging tone pairs to be more similar than those with lower token frequency. It is difficult, however, to predict the effects of tone-type frequency, as there are only six lexical tones in Cantonese, and there is no separate frequency count for monosyllabic versus disyllabic words. As there are very few studies investigating the effects of word frequency on tone production, let alone tone merging, the current study systematically manipulated the token frequency of disyllabic words to further explore how word frequency affects the realization of merging tones.

The current study is an extension of Mok et al. (2013) who reported on monosyllabic data. The disyllabic production data were collected from the same speakers in Mok et al. (2013) at the same time as the monosyllabic experiment, i.e., recorded over 10 years ago, but have not been

reported. Although the disyllabic data have been collected a while ago, they are still valid and insightful, as the focus of the current study is not on how tone merging has progressed in recent years, but on how the degree of tone merging interacts with tonal coarticulation within the same individuals. Thus, the timing of the data is not a concern. Direct comparison with and reanalysis of the monosyllabic data in Mok et al. (2013) were done for a more comprehensive investigation of the tone-merging phenomenon in Hong Kong Cantonese.

# 2. Methods

## 2.1. Participants

The participants were a subset from Mok et al. (2013), namely the 17 native Cantonese speakers (14 female, 3 male, aged 18 to 22 at the time of recording, all university students) who participated in the monosyllabic production experiment. These 17 participants were selected as merging participants through a screening test in which two native Cantonese-speaking phoneticians, who clearly distinguished all six tones in Cantonese, judged the tone production of 169 native Cantonese speakers and identified 28 potentially merging speakers with non-standard tone production. Because of time limitation and various logistic constraints in the original study, only 17 of the 28 identified speakers were recorded, while all of them participated in the perception experiment. Therefore, the current study could only examine the disyllabic production data of the 17 recorded speakers. As well as the merging participants, the same two female native Hong Kong Cantonese reference speakers (R1 and R2), who clearly distinguished all six tones, also produced the same sets of monosyllabic and disyllabic materials for comparison.

## 2.2. Materials

The current study investigated tone merging in disyllabic words with three factors: syllable position (whether the target tone is on the first or second syllable in a disyllabic word), tonal context (the tone following or preceding the target tone on the first or the second syllable respectively) and word frequency (token frequency). Both monosyllabic and disyllabic data were examined and compared.

The tonal contexts were divided into three categories based on the tone onsets/offsets of the contextual tones: high, mid and low. For example, if a target tone precedes or follows contextual T2 [25], since T2 starts at a low pitch [2] and ends on a high pitch [5], the tonal context is low for the target tone preceding T2 (i.e., with the target tone on the first syllable), while the tonal context is high for the target tone following T2 (i.e., with the target tone on the second syllable). Correspondence between the six tones and their tonal contexts is presented in **Table 1** for easy reference. It can be seen that the tonal context is balanced among the high, mid and low

categories for target tones on the second syllable, while the low tonal context dominates when the target tones are on the first syllable. This imbalance is unavoidable given the pitch shapes of the six Cantonese tones.

| Contextual tone | Target tones on 1st syllable anticipatory coarticulation | Target tones on 2nd syllable carryover coarticulation |
|---|---|---|
| T1 [55] | high | high |
| T2 [25] | low | high |
| T3 [33] | mid | mid |
| T4 [21] | low | low |
| T5 [23] | low | mid |
| T6 [22] | low | low |

**Table 1:** Tonal contexts with the target tones in different syllable positions.

As for the effects of word frequency, as described in Mok et al. (2013), the frequency of Cantonese words was calculated using the log frequency (base 10) of the token frequency in a written corpus of Hong Kong newspapers (Chan & Tang, 1990). This corpus includes 33,000 Cantonese words, and the high-frequency and low-frequency words were chosen from the higher and lower logged frequency ranges in the corpus.

In order to examine the effects of syllable position, tonal context and word frequency, six high-frequency and six low-frequency disyllabic words were chosen for each of the six tones, with the target syllables appearing as both the first and the second syllables. For example, 經濟 [keŋ$^{55}$ t͡sɐi$^{33}$] 'economy' and 欣慰 [jɐn$^{55}$ wɐi$^{33}$] 'gratified' are high- and low-frequency words respectively for target T1 [55] appearing as the first syllable; 參與 [t͡sʰaːm$^{55}$ jyː$^{23}$] 'participate in' and 已經 [jiː$^{23}$ keŋ$^{55}$] 'already' are high-frequency words for target T1 appearing as the first or second syllable. Tonal contexts were fully crossed, i.e., all six tones were used as the preceding and following contexts for the target syllables. As a result, 144 disyllabic words (2 frequencies × 2 syllable positions × 6 contextual tones × 6 target tones) were chosen as the disyllabic production materials. A sample wordlist for target T1 in high frequency words is given in Appendix A for illustration. These words were embedded in a carrier sentence: [ŋɔ$^{23}$ tʊk$^{2}$ __ t͡sʰɵt$^{5}$ lɐi$^{21}$] 'I read __ out', and each sentence was repeated three times, giving a total of 432 sentences in a randomized order for the participants to read. All the target sentences were presented in Chinese characters, and the participants were recorded reading them with no information about tones provided.

A separate set of materials consisting of disyllabic minimal pairs[3] between T2 and T5, and between T3 and T6, were also recorded for comparison with the above non-minimal pair materials. While the non-minimal tone pairs differed in terms of both segments and tones, this set of minimal-pair materials consisted of identical segments of consonants and vowels, and differed only in the tones of the second syllables, e.g., 不變 [pɐt$^{5}$ pɪn$^{33}$] 'unchanged' vs. 不便 [pɐt$^{5}$ pɪn$^{22}$] 'inconvenient'. In sum, there were 24 target words in this set of materials, as six different disyllabic words were selected for each of the target tones (6 words × 4 tones). The participants read the target minimal pairs in the same carrier sentences with three repetitions, giving a total of 72 sentences. Syllable position, tonal context and word frequency were not considered for the minimal pairs due to practical constraints. The minimal pairs were randomized and mixed with other disyllabic words, so they did not stand out as minimal pairs. The list of all the disyllabic words used in the production experiment can be found online at https://osf.io/86f4m/?view_only=e22e31c0d15c4fb0950b7fa56d5ff9bf.

In addition, the monosyllabic words in Mok et al. (2013) were also included for comparison, with six different words of both high and low frequency for each of the six Cantonese tones (6 words × 6 tones × 2 frequencies) embedded in the carrier sentence [ŋɔ$^{23}$ tʊk$^{2}$ __ tsi$^{22}$] 'I read the word __'.

All three sets of materials are summarized in **Table 2**.

| Word conditions | Combinations |
|---|---|
| Disyllabic words | 2 frequencies × 2 syllable positions × 6 contextual tones × 6 target tones = 144 words |
| Disyllabic minimal pairs | 4 target tones × 6 different words = 24 words |
| Monosyllabic words | 2 frequencies × 6 target tones × 6 different words = 72 words |

**Table 2:** Materials used in the production experiment.

## 2.3. Procedure

Two randomized lists of the same materials were used (nine participants read the first list and eight participants read the second list). The speakers were recorded in a soundproof booth using Praat (Boersma, 2001) at a sampling rate of 22,050 Hz. A condenser microphone was positioned approximately 20cm from the participants. The participants read the monosyllabic materials before the disyllabic materials. Breaks were given when necessary.

[3] No minimal pair for T4 and T6 was included because of the fact that merging between this tone pair is not as prominent as for the other two tone pairs, and data for the other two tone pairs already answer the research question about minimal versus non-minimal pairs.

## 2.4. Acoustic measurements and statistical analyses

Segmentation was done in such a way that was consistent with Mok et al. (2013). For syllables starting with sonorants, the whole syllable was segmented, and for syllables with a non-sonorous onset, only the vowel part was segmented. The onset and offset of segmentation were defined as the start of F1 and the end of F2, respectively. To better capture tonal variation in disyllabic contexts, the F0 values were time-normalized with eleven equidistant measurement points taken for each syllable in the disyllabic words, while only ten equidistant measurement points were taken for the monosyllabic words in Mok et al. (2013). The (F0) values of these syllables or vowels were extracted in semitones using ProsodyPro (Xu, 2013) in Praat to make the F0 data more comparable and interpretable. Due to consonant perturbation from the onset and offset, the first and final measurement points were excluded from the analysis. That is, the first and tenth points for monosyllabic words and the first and eleventh points for disyllabic words. Anomalous F0 values were fixed manually using ProsodyPro. Approximately 1.6% of tokens were excluded from the analysis for either being heavily creaky or mispronounced as other irrelevant tones (neither the target tone nor the merging counterpart).

First, in order to get an overall picture of the tone merging patterns, SS-ANOVAs were conducted using all the data points (except for the onset and offset) to compare the tone contours of the 17 participants for the merging tone pairs of T2/T5, T3/T6 and T4/T6.

Subsequently, following Mok et al. (2013), we used predictive discriminant analysis (also called linear discriminant analysis, LDA) to predict the classification of the merging tone categories in Cantonese by comparing the F0 values of the target Cantonese tones against all Cantonese tone categories. Different from Mok et al. (2013), which used the second, fifth, sixth and ninth measurement points to represent the onset, midpoint and offset F0 values of the tone contours, the current study used three measurement points focusing more on the latter part of the tone contours, since the merging tone pairs mainly differ towards the end of the contours. F0 values at three measurement points were used as predictors for the classification of tone categories, i.e., the points one third of the way along, three quarters of the way along and at the end of the tone contour, minus the very first and last measurement points. This corresponded to the fourth, eighth and tenth data points for disyllabic words and the fourth, seventh and ninth data points for monosyllabic words, respectively. The LDA was conducted in SPSS, using the leave-one-out cross-validation method. Before classification, univariate outliers with z-scores above 3.29 or below –3.29 and multivariate outliers detected using Mahalanobis distance with $p < .001$ were excluded (Tabachnick & Fidell, 2019).

Both the SS-ANOVAs and the LDAs across different factors compared how similar the merging tone pairs were among monosyllabic words, disyllabic non-minimal pairs and disyllabic minimal pairs. As for the LDAs for different factors, effects of syllable position (first vs. second syllable), tonal context (high, mid or low context) and word frequency (high vs. low frequency) were only examined in disyllabic non-minimal pairs, and these effects were

confirmed by Bayesian linear mixed models constructed using the brms package (Bürkner, 2017) in R.

We chose the Bayesian approach as a powerful alternative to the Frequentist approaches because Bayesian mixed effect methods provide stable estimates for groups with smaller sample sizes (i.e., only one misclassification rate for each speaker in each condition) with the help of weakly informative priors, which are likely to solve the singular fit problem we encountered in the mixed model analyses. To apply Bayesian statistics to explore complex probability distributions (in most cases), Markov Chain Monte Carlo (MCMC) is necessary. Unlike conventional methods that use fixed point estimates, Bayesian inference represents unknown parameters as probability distributions. This probabilistic representation requires a different optimization approach—one that involves sampling from the posterior distribution. MCMC is a powerful tool for estimating parameters via the posterior probability distribution. MCMC essentially performs Monte Carlo integration to estimate complex integrals using Markov chains whose equilibrium distribution approximates the target distribution. Since the brms package allows fitting linear mixed models in a lme4-like syntax within the Bayesian framework and MCMC methods, we included speakers as a random effect and the following effects as fixed effects respectively: tone pair (T2/T5, T3/T6, T4/T6), word type (monosyllabic vs. disyllabic words), minimal pair (minimal pairs vs. non-minimal pairs), syllable position, tonal context, and word frequency, as well as their interactions with different tone pairs. For example, the analytical formula for the effect of word type (disyllabic vs. monosyllabic words) is as follows: misclassification rate~word type*tone pair + (1|subject), which led to no convergence problems. As for the priors, weakly informative priors of normal (0, 10) were adopted, as recommended by previous studies in a situation where strong prior knowledge is not available (Polson & Scott, 2012; Williams et al., 2018). The Rhat for all parameters in all the Bayesian models in the following analyses was equal to 1, suggesting that the models had successfully converged.

To restate the statistical approach for clarity, one factor and its interaction with tone pairs were examined at a time within each Bayesian mixed-effects model. For each of the subsections in the Results, a different Bayesian model is reported, focusing on the individual factor and its interaction with the tone pairs. The advantage of including only one factor and its interaction with tone pairs in each model, rather than including several factors in a single model, is to reduce the complexity of the model, which can facilitate easy interpretation of the individual effects of each factor on tone merging. This approach can also minimize potential confounding and collinearity between factors, and avoid overfitting (which could occur if too many factors were included simultaneously), which would obscure the specific influence of each factor.

Regarding the analysis sequence, LDA was first performed in SPSS, with leave-one-out cross-validation (i.e., repeatedly training the model on all but one data point and then testing it on the left-out data point. This process was repeated for each data point in the dataset). The misclassification rates calculated by LDA were fitted into a Bayesian mixed-effects model using

MCMC, as described above, to examine the impact of factors like word type, minimal pair, syllable position, tonal context, and word frequency, and their interactions with tone pairs.

To complement the LDA, which only used three timepoints, we also adopted growth curve analysis (GCA) to compare the whole tone contours to present as comprehensive a picture as possible of the merging tone pairs. GCA was implemented using mixed-effects models to explore the interactions between tones, time points, and the factors (word type, minimal pair, syllable position, tonal context and word frequency) across the entire time course, offering complementary insights into how these factors may influence tone distinctions.

Unlike LDA using only three representative timepoints, i.e., 1/3, 3/4 and the offset of the tone contours, GCA incorporated all the timepoints (9 for disyllabic and 8 for monosyllabic words except for the very first and last measurement points), which could capture the dynamic trajectory of tone contours over time.

However, while GCA effectively models individual tone trajectories, it does not inherently focus on tone pair comparisons. For instance, our analysis used the contrast method within the emmeans framework to estimate differences between the target tone pairs (i.e., T2 vs. T5, T3 vs. T6, T4 vs. T6). Nonetheless, these pairwise comparisons only offer insights into specific tone pairs and do not directly evaluate overall effects across tone pairs. In contrast, LDA computes misclassification rates that aggregate differences for each tone pair, providing a more direct measure of tone pair merging. Thus, we relied on LDA to compute overall effects of different factors, but used GCA as a valuable supplement to LDA. Additionally, due to computational constraints, a simpler linear mixed-effects model was used for GCA, as Bayesian models were relatively impractical given the large dataset. Some discrepancies may also arise from differences in methodological focus and different models adopted.

For data interpretation, smaller values in GCA estimates indicate smaller differences between tone pairs, reflecting a higher degree of merging and aligning with higher misclassification rates from LDA. The F0 data (in semitone) of both monosyllabic and disyllabic words can be found online at https://osf.io/86f4m/?view_only=e22e31c0d15c4fb0950b7fa56d5ff9bf.

# 3. Results

## 3.1. Merging tone pairs (disyllabic vs. minimal pairs vs. monosyllabic)

To compare the degree of tone merging among different tone pairs, first we ran SS-ANOVAs to visualize the entire contours for the merging tone pairs using all the data points of the 17 merging speakers. **Figure 1** shows the tone contours in terms of different tone pairs (i.e., T2/T5, T3/T6 and T4/T6) and word types (i.e., disyllabic non-minimal pairs, disyllabic minimal pairs, and monosyllabic words). The non-overlapping portions mean that these portions were significantly different between the two contours. In general, the tone pairs were more similar in disyllabic words than monosyllabic words, and more similar in minimal pairs than in non-minimal pairs. The degree of merging between these pairs will be verified in the following LDAs.

LDA was used to predict the classification of possibly merging tone categories based on three data points: a third of the way along, three quarters of the way along and at the end of the tone contour. Each tone production token was correctly or incorrectly classified, and classification accuracy rates were calculated individually. Only five tones were analyzed in the following statistical test, similar to Mok et al. (2013). T1 was excluded because Cantonese T1 is stable and not merging with any other tone, and the classification accuracy rates were not significantly different between LDA results based on all six tones and only five tones without T1. A higher classification accuracy rate, equivalent to a lower misclassification rate, indicates clearer separation of different tones—that is, less merging between tones. For disyllabic words, in terms of the overall classification accuracy rate across different tones, as expected, the two reference speakers had a higher classification accuracy rate (mean: 74.3%, SD: 5.7%) than the 17 merging speakers (mean: 68.5%, SD: 9%), which is similar to the rates in monosyllabic words with the two reference speakers (mean: 91.8%, SD: 0.2%) being higher than the 17 merging speakers (mean: 81.5%, SD: 8.7%).

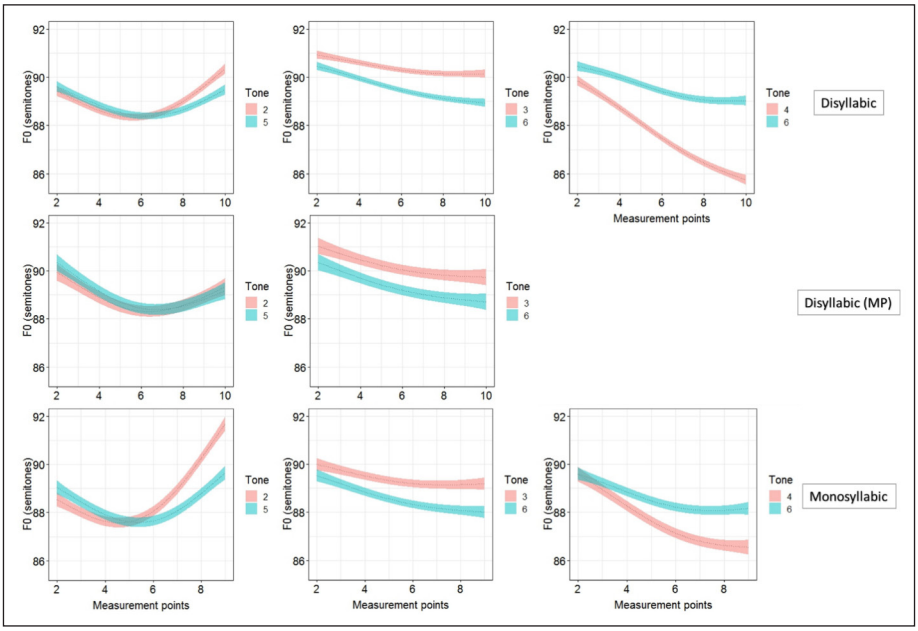



**Figure 1:** The 17 merging speakers' normalized F0 contours of the Cantonese merging tone pairs (T2/T5, T3/T6 and T4/T6 from left to right) in disyllabic non-minimal pairs, disyllabic minimal pairs and monosyllabic words (from top to bottom), calculated by SS-ANOVAs.

### 3.1.1. Disyllabic vs. monosyllabic words

**Table 3** presents the misclassification rates of the merging tone pairs of the two reference speakers (R1 and R2) and 17 merging speakers (M1∼M17), for both disyllabic and monosyllabic words. 'T2→T5' in the table means the percentage of T2 misclassified as T5. The same applies to other tone pairs. The 'overall' misclassification rate is the average of all six misclassification

percentages from 'T2→T5' to 'T6→T4'. The misclassification rates in **Table 3** were obtained following the steps of LDA in Mok et al. (2013), except that different measurement points were used in the present study, as previously discussed. This was to ensure the comparability between the disyllabic and the monosyllabic data by using measurement points a third of the way along, three quarters of the way along and at the end of the tone contour for both datasets. Thus, the misclassification rates for monosyllabic words are slightly different from those in Mok et al. (2013), which used four measurement points.

**Table 3** shows that higher misclassification rates were more frequent in disyllabic words than in monosyllabic words, echoing the patterns observed in the SS-ANOVA results above. For disyllabic words, even the two reference speakers had at least one tone pair misclassified more than 20% of the time. The rate of 20% was chosen for highlighting instead of the 10% in Mok et al. (2013) because the misclassification rates were generally much higher in disyllabic words. Different degrees of similarity were observed in the tone pairs among the speakers. Among all the potentially merging speakers, M15 produced all three tone pairs with a misclassification rate of less than 20% for both disyllabic and monosyllabic words. This speaker's misclassification rates for monosyllabic words in Mok et al. (2013) were slightly higher but still comparable to those in this study.

It is clear that none of the tone pairs was completely merged, except for T4 and T6 in monosyllabic words produced by M9. This speaker's T4 was misclassified as T6 all the time (i.e., there was a 100% misclassification rate), although her misclassification rate for T4→T6 was only 24.2% in disyllabic words.

Since there were only two reference speakers for comparison, their data were insufficient to examine the effects of various factors like word type and word frequency statistically. Therefore, the reference speakers were not included in the following linear mixed effect analyses for various effects on tone merging. Only data from the 17 merging speakers were used.

In terms of the degree of tone merging, differences between the three tone pairs were confirmed by Bayesian linear mixed models. The results were considered statistically significant if the 95% credible interval (95% CrI) did not include zero and the probability of direction (pd) was larger than 97.5%. A pd of 97.5% would correspond approximately to a two-tailed p-value of 0.05. As is illustrated in the upper panel of **Figure 2**, for disyllabic words, T4/T6 was less merged than T2/T5 ($b = 20.83$, 95% CrI = [10.67, 30.17], $pd(b > 0) = 99.98\%$) and T3/T6 ($b = 19.35$, 95% CrI = [9.32, 28.42], $pd(b > 0) = 100\%$) by the 17 merging speakers. In contrast, for monosyllabic words, the misclassification rates were not significantly different between the tone pairs (T2/T5 vs. T3/T6: $b = 3.3$, CrI = [–8.31, 15.19], $pd(b > 0) = 71.17\%$; T2/T5 vs. T4/T6: $b = –2.2$, CrI = [–13.94, 9.09], $pd(b < 0) = 64.18\%$; T3/T6 vs. T4/T6: $b = 2.37$, CrI = [–8.89, 13.69], $pd(b > 0) = 66.70\%$).

| | Disyllabic_misclassification rates (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **T2→T5** | **T5→T2** | **T3→T6** | **T6→T3** | **T4→T6** | **T6→T4** | **Overall** |
| R1 | 32.1 | 20.3 | 6.9 | 15.3 | 0.0 | 1.4 | 12.7 |
| R2 | 25.7 | 21.4 | 14.1 | 13.9 | 34.8 | 12.5 | 20.4 |
| R mean | 28.9 | 20.9 | 10.5 | 14.6 | 17.4 | 7.0 | 16.5 |
| M1 | 27.8 | 27.8 | 26.4 | 18.1 | 4.9 | 2.8 | 18.0 |
| M2 | 26.4 | 25.0 | 25.0 | 20.8 | 5.8 | 6.9 | 18.3 |
| M3 | 28.2 | 11.4 | 14.1 | 25.0 | 5.2 | 0.0 | 14.0 |
| M4 | 16.7 | 33.3 | 16.7 | 16.7 | 12.1 | 5.6 | 16.9 |
| M5 | 34.7 | 16.7 | 18.8 | 15.5 | 6.2 | 0.0 | 15.3 |
| M6 | 18.3 | 22.1 | 29.6 | 17.4 | 7.5 | 1.4 | 16.1 |
| M7 | 25.7 | 23.9 | 18.1 | 12.7 | 21.8 | 1.4 | 17.3 |
| M8 | 31.9 | 22.2 | 32.4 | 36.1 | 4.8 | 8.3 | 22.6 |
| M9 | 18.6 | 15.5 | 18.1 | 20.8 | 24.2 | 11.1 | 18.1 |
| M10 | 22.2 | 12.5 | 15.3 | 23.5 | 8.9 | 2.9 | 14.2 |
| M11 | 23.6 | 18.1 | 29.2 | 25.0 | 9.1 | 1.4 | 17.7 |
| M12 | 8.5 | 15.7 | 44.4 | 38.9 | 14.3 | 0.0 | 20.3 |
| M13 | 23.2 | 33.3 | 13.9 | 25.4 | 38.7 | 25.4 | 26.7 |
| M14 | 20.8 | 22.2 | 23.6 | 19.4 | 14.8 | 12.5 | 18.9 |
| M15 | 12.5 | 7.5 | 17.4 | 18.1 | 1.4 | 2.8 | 10.0 |
| M16 | 27.8 | 26.8 | 34.7 | 29.6 | 4.5 | 0.0 | 20.6 |
| M17 | 41.7 | 41.7 | 5.6 | 9.9 | 5.0 | 0.0 | 17.3 |
| M mean | 24.0 | 22.1 | 22.5 | 21.9 | 11.1 | 4.9 | 17.8 |

| | Monosyllabic_misclassification rates (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **T2→T5** | **T5→T2** | **T3→T6** | **T6→T3** | **T4→T6** | **T6→T4** | **Overall** |
| R1 | 2.9 | 2.8 | 5.6 | 14.3 | 6.7 | 0.0 | 5.4 |
| R2 | 8.6 | 0.0 | 8.3 | 2.8 | 2.9 | 0.0 | 3.8 |
| R mean | 5.8 | 1.4 | 7.0 | 8.6 | 4.8 | 0.0 | 4.6 |
| M1 | 11.1 | 2.8 | 13.9 | 8.3 | 5.7 | 2.8 | 7.4 |
| M2 | 37.5 | 16.7 | 19.4 | 2.8 | 62.5 | 5.6 | 24.1 |
| M3 | 3.1 | 0.0 | 21.2 | 16.7 | 0.0 | 0.0 | 6.8 |
| M4 | 15.6 | 0.0 | 24.2 | 13.9 | 8.3 | 2.8 | 10.8 |
| M5 | 11.4 | 5.6 | 11.1 | 5.6 | 8.3 | 0.0 | 7.0 |
| M6 | 16.7 | 16.7 | 22.2 | 19.4 | 2.9 | 0.0 | 13.0 |
| M7 | 17.6 | 5.7 | 13.9 | 17.6 | 23.5 | 5.9 | 14.0 |
| M8 | 14.3 | 8.3 | 33.3 | 36.1 | 0.0 | 0.0 | 15.3 |
| M9 | 14.3 | 14.3 | 17.6 | 13.9 | 100.0 | 2.8 | 27.2 |
| M10 | 2.9 | 0.0 | 11.1 | 5.7 | 0.0 | 0.0 | 3.3 |
| M11 | 5.6 | 0.0 | 2.8 | 8.6 | 43.8 | 8.6 | 11.6 |
| M12 | 0.0 | 2.9 | 27.8 | 25.7 | 15.4 | 0.0 | 12.0 |
| M13 | 29.4 | 22.2 | 2.8 | 20.0 | 44.8 | 31.4 | 25.1 |
| M14 | 22.9 | 13.9 | 8.3 | 8.3 | 37.5 | 8.3 | 16.5 |
| M15 | 5.6 | 0.0 | 16.7 | 2.8 | 8.3 | 0.0 | 5.6 |
| M16 | 25.0 | 22.2 | 27.3 | 16.7 | 0.0 | 0.0 | 15.2 |
| M17 | 25.0 | 13.9 | 2.8 | 2.8 | 22.2 | 0.0 | 11.1 |
| M mean | 15.2 | 8.5 | 16.3 | 13.2 | 22.5 | 4.0 | 13.3 |

**Table 3:** Misclassification rates (%) of the merging tone pairs of the two reference speakers (R) and 17 merging speakers (M) for disyllabic (left) and monosyllabic (right) words. Misclassification rates over 20% are shaded grey for reference.

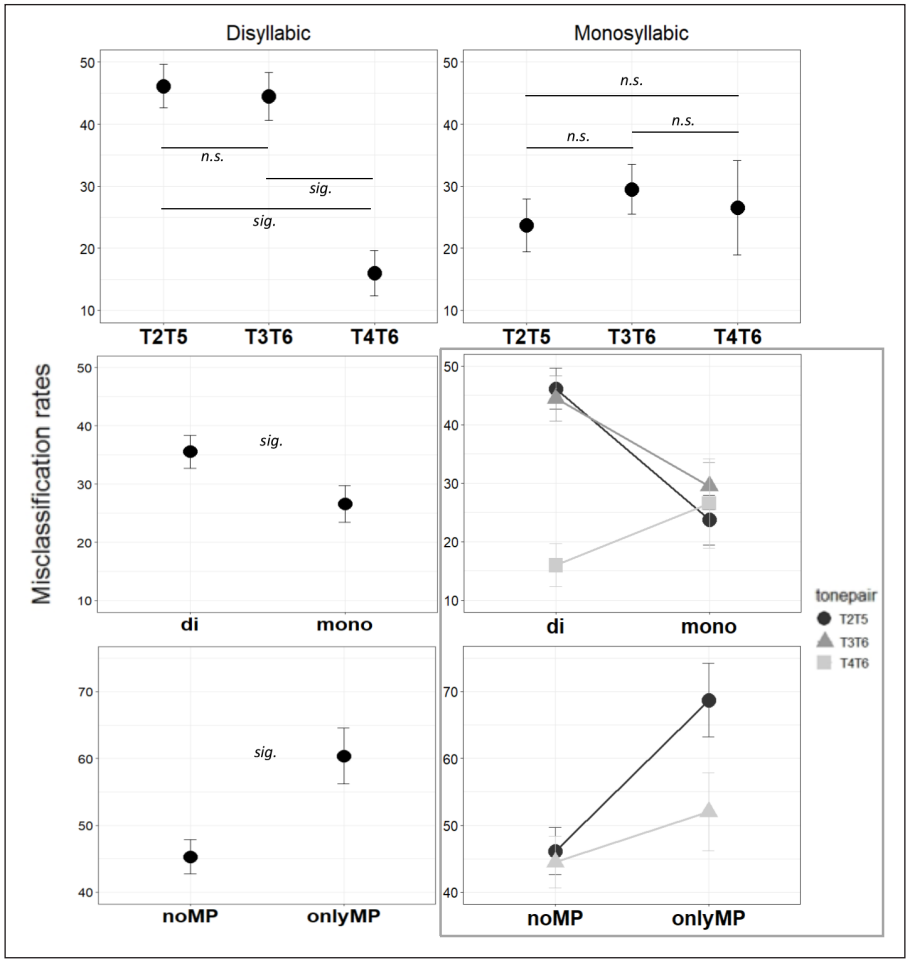



**Figure 2:** Seventeen merging speakers' misclassification rates (upper panels) in different tone pairs in disyllabic and monosyllabic words; (middle panels) influenced by word type (di = disyllabic words, mono = monosyllabic words); and (lower panels) influenced by minimal pair type (noMP = non-minimal pairs, onlyMP = minimal pairs). Interactions between the effects of word type (middle panel), minimal pair type (lower panel) and tone pair are framed in a grey box at the lower right-hand corner.

As confirmed by GCA, (see Appendix E [GCA statistics] and Appendix F [GCA graphs] for all GCA results), for both disyllabic and monosyllabic words, the degree of distinction between tone pairs increased in the following order: T2/T5 < T3/T6 < T4/T6, indicating that T2/T5 was undergoing more tone merging, followed by T3/T6, and the least merging was observed between T4 and T6, supporting the LDA and SSANOVA results in **Figure 1**.

The effects of word type (i.e., disyllabic and monosyllabic words) and its interaction with tone pairs are presented in the middle panels of **Figure 2**. The tone pairs were more similar with higher misclassification rates in disyllabic words than in monosyllabic words ($b = -7.61$, $CrI = [-14.81, -0.20]$, $pd(b < 0) = 97.85\%$). Word type also interacted with tone pairs. More specifically, the misclassification rates of T4/T6 between disyllabic and monosyllabic words did not vary as much as they did for the other two tone pairs, significantly so for T2/T5 ($b = -20.94$, $CrI = [-35.90, -5.41]$, $pd(b < 0) = 99.58\%$).

The GCA results also showed that disyllabic words exhibited greater tone merging compared to monosyllabic words in T2/T5, while for T4/T6, disyllabic words showed less tone merging than monosyllabic words (as also shown in **Figure 2**). The smaller difference between disyllabic and monosyllabic words for the T3/T6 tone pair was not significant.

The results showed that the tone pairs were more similar in disyllabic words, significantly so for T2/T5. This was expected since the tone pairs in disyllabic words are influenced by tonal coarticulation. The lower misclassification rate of T4/T6 in disyllabic words and its slight variation between different word types compared with the other two tone pairs are probably due to the fact that T4 is creakier than other tones, thus leading to a better separation between T4 and T6 with this extra cue, even in disyllabic words. This tone pair is also progressing much more slowly in terms of merging than the other two tone pairs, as mentioned in the Introduction.

As for the relationship between the degree of tone merging in monosyllabic and disyllabic words, the positive correlations between monosyllabic and disyllabic misclassification rates in the upper panels of **Figure 3** indicate that individuals who produced more similar tones in monosyllabic words also merged tones more frequently in disyllabic words. The positive correlation was found across all tone pairs, as well as within each tone pair. Such positive correlations also indicate that tonal coarticulation was not the only reason that the tone pairs were more similar in disyllabic than in monosyllabic words.

### 3.1.2. Disyllabic non-minimal vs. minimal pairs

The disyllabic materials can be further divided into the non-minimal pair and the minimal pair sets (differing only in tones). The minimal pair set consisted of the T2/T5 and T3/T6 pairs only. **Table 4** presents the misclassification rates of T2/T5 and T3/T6 for disyllabic non-minimal pairs and minimal pairs respectively, showing that the speakers (including the two reference speakers) had at least one tone pair misclassified more than 20% of the time. The target tones in both the non-minimal and the minimal pair sets were not merged completely (i.e., 100%) but to a substantial degree (over 30%~40% for non-minimal pairs by 7 merging speakers, over 40%~50% for minimal pairs by 10 merging speakers).

As confirmed by Bayesian mixed models, the lower panels in **Figure 2** show that the tones were more similar in minimal pairs ($b = 12.03$, CrI = [3.46, 20.35], $pd(b > 0) = 99.50\%$). As for its interaction with tone pairs, T2/T5 was influenced differently from T3/T6 in terms of minimal pair types. If the disyllabic words were not further divided into different minimal pair types, T2/T5 was not misclassified more often than T3/T6 (in **Figure 2**, upper panels), but the lower right-hand panel in **Figure 2** illustrates that the misclassification rate of T2/T5 (68.7%) was higher than that of T3/T6 (52%) in minimal pair production ($b = 12.9$, CrI = [0.43, 25.7], $pd(b > 0) = 97.62\%$). The SS-ANOVA results in **Figure 1** (left column) showed that T2/T5 were completely merged in minimal pairs, while only the tone offsets (measuring points 8 to 10) were different in disyllabic non-minimal pairs. The two tones were significantly different earlier (from point 6 onwards) in monosyllabic words. The GCA results also showed a marginally larger difference between minimal vs. non-minimal pairs for T2/T5 ($p = 0.054$), while the smaller difference for T3/T6 was not significant.

The reason for the tone pairs being more similar in minimal pairs may be related to the fact that in the production experiment, the minimal pairs were presented to the speakers randomly rather than sequentially, so the speakers might not be aware of the existence of minimal pairs. They therefore might not try to highlight the contrast between different tones. In addition, since the segments were identical in the minimal pairs, the tone contours were not influenced by different consonants and vowels. As a result, the subtle difference between the tones would be further reduced, making the tones in minimal pairs appear less distinct. However, it is unclear why T2/T5 minimal pairs were more similar than T3/T6 minimal pairs. It may be because the merging of T2/T5 is more advanced than T3/T6, as demonstrated in some previous studies.

Finally, as illustrated in the lower panels of **Figure 3**, the positive correlation between misclassification rates in non-minimal pairs and minimal pairs indicate that individuals who merged the tones more in non-minimal pairs also did so more often in minimal pairs. This correlation was maintained across both tone pairs of T2/T5 and T3/T6, as well as within each tone pair.

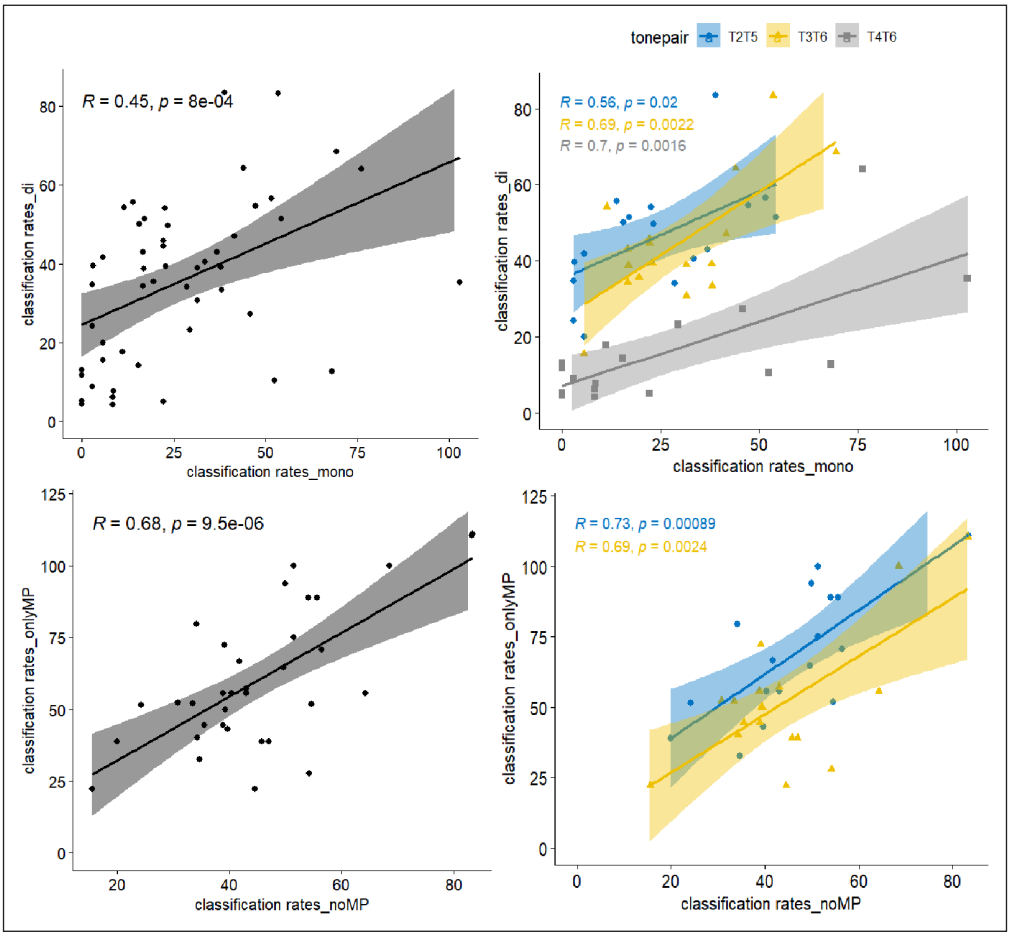



**Figure 3:** Correlations between monosyllabic and disyllabic misclassification rates (upper panels), as well as between non-minimal pair and minimal pair misclassification rates (lower panels) across (left) and within (right) tone pairs by the 17 merging speakers.

## 3.2. Effects of syllable position

The misclassification rates of target tones appearing in the first and second syllables of the disyllabic words are presented in Appendix B. The effect of syllable position and its interaction with tone pairs was further examined and the results are presented in **Figure 4**. Using Bayesian mixed models, the main effect of syllable position was significant (b = 7.95, CrI = [0.18, 15.60], pd(b > 0) = 97.70%), with the tone pairs being more similar in the second syllables than in the first syllables, i.e., there was stronger carryover coarticulation than anticipatory coarticulation. However, the insignificant interaction effect showed that the effect of syllable position was the same across tone pairs. Since the tone pairs in the first syllables of disyllabic words were similar to those in monosyllabic words (b = –0.74, CrI = [–7.33, 5.95], pd(b > 0) = 59.13%), i.e., they were not more merged than monosyllables, the overall larger degree of tone merging in the disyllabic words should be attributed to the tone pairs being more similar in the second syllables. The GCA results also confirmed that the T2/T5 and T3/T6 pairs had a larger degree of tone merging in the second syllable compared to the first, while the difference for the T4/T6 pair was not significant.

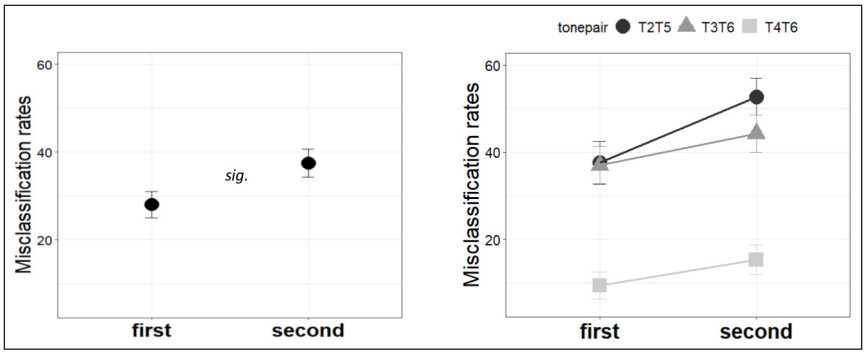



**Figure 4:** Effects of syllable position and the syllable position × tone pair interaction as produced by the 17 merging speakers.

## 3.3. Effects of tonal context

To examine the effect of tonal context on tone merging, the condition can be divided into the target tones appearing on the first or the second syllables, so the tonal contexts (high, mid and low; see **Table 1** for the grouping) will follow the first syllables or precede the second syllables respectively. The misclassification rates of the target tones appearing on the first and the second syllables in different tonal contexts are presented in Appendix C (target tones on the first syllable) and Appendix D (target tones on the second syllable) respectively.

First, for the tonal contexts following the target tones as the first syllables (i.e., anticipatory coarticulation), as illustrated in the upper panels of **Figure 5**, the misclassification rate was not

| | Disyllabic (noMP)_misclassification rates (%) | | | | |
|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | Overall |
| R1 | 32.1 | 20.3 | 6.9 | 15.3 | 18.7 |
| R2 | 25.7 | 21.4 | 14.1 | 13.9 | 18.8 |
| R mean | 28.9 | 20.9 | 10.5 | 14.6 | 18.7 |
| M1 | 27.8 | 27.8 | 26.4 | 18.1 | 25.0 |
| M2 | 26.4 | 25.0 | 25.0 | 20.8 | 24.3 |
| M3 | 28.2 | 11.4 | 14.1 | 25.0 | 19.7 |
| M4 | 16.7 | 33.3 | 16.7 | 16.7 | 20.9 |
| M5 | 34.7 | 16.7 | 18.8 | 15.5 | 21.4 |
| M6 | 18.3 | 22.1 | 29.6 | 17.4 | 21.9 |
| M7 | 25.7 | 23.9 | 18.1 | 12.7 | 20.1 |
| M8 | 31.9 | 22.2 | 32.4 | 36.1 | 30.7 |
| M9 | 18.6 | 15.5 | 18.1 | 20.8 | 18.3 |
| M10 | 22.2 | 12.5 | 15.3 | 23.5 | 18.4 |
| M11 | 23.6 | 18.1 | 29.2 | 25.0 | 24.0 |
| M12 | 8.5 | 15.7 | 44.4 | 38.9 | 26.9 |
| M13 | 23.2 | 33.3 | 13.9 | 25.4 | 24.0 |
| M14 | 20.8 | 22.2 | 23.6 | 19.4 | 21.5 |
| M15 | 12.5 | 7.5 | 17.4 | 18.1 | 13.9 |
| M16 | 27.8 | 26.8 | 34.7 | 29.6 | 29.7 |
| M17 | 41.7 | 41.7 | 5.6 | 9.9 | 24.7 |
| M mean | 24.0 | 22.1 | 22.5 | 21.9 | 22.7 |

| | Disyllabic (onlyMP)_misclassification rates (%) | | | | |
|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | Overall |
| R1 | 33.3 | 20.0 | 11.1 | 11.1 | 18.9 |
| R2 | 5.6 | 23.5 | 11.1 | 11.1 | 12.8 |
| R mean | 19.5 | 21.8 | 11.1 | 11.1 | 15.9 |
| M1 | 44.4 | 44.4 | 11.1 | 11.1 | 27.8 |
| M2 | 52.9 | 22.2 | 22.2 | 16.7 | 28.5 |
| M3 | 11.8 | 31.3 | 38.9 | 33.3 | 28.8 |
| M4 | 58.8 | 35.0 | 35.3 | 16.7 | 36.5 |
| M5 | 55.6 | 44.4 | 16.7 | 23.5 | 35.1 |
| M6 | 27.8 | 27.8 | 27.8 | 11.1 | 23.6 |
| M7 | 33.3 | 31.3 | 11.1 | 41.2 | 29.2 |
| M8 | 50.0 | 38.9 | 55.6 | 44.4 | 47.2 |
| M9 | 33.3 | 46.2 | 27.8 | 16.7 | 31.0 |
| M10 | 15.8 | 16.7 | 16.7 | 38.9 | 22.0 |
| M11 | 22.2 | 44.4 | 11.1 | 16.7 | 23.6 |
| M12 | 22.2 | 29.4 | 88.2 | 22.2 | 40.5 |
| M13 | 35.3 | 35.3 | 16.7 | 33.3 | 30.2 |
| M14 | 22.2 | 33.3 | 27.8 | 29.4 | 28.2 |
| M15 | 16.7 | 22.2 | 16.7 | 27.8 | 20.9 |
| M16 | 26.7 | 25.0 | 33.3 | 22.2 | 26.8 |
| M17 | 61.0 | 50.0 | 11.1 | 11.1 | 33.3 |
| M mean | 34.7 | 34.0 | 27.5 | 24.5 | 30.2 |

**Table 4:** Misclassification rates (%) of the merging tone pairs of the two reference speakers (R) and 17 merging speakers (M) for disyllabic non-minimal pairs (left) and minimal pairs (right). Misclassification rates over 20% are shaded grey for reference.

significantly different across tonal contexts, and the insignificant effect of tonal context did not change within different tone pairs, although the patterns were similar to the significant tonal context effect preceding the target tones as the second syllables (i.e., carryover coarticulation, see below).

The GCA results revealed subtler differences among the tone pairs. As with the LDA results, there were no effects of tonal context for the T2/T5 pair, while T3/T6 were more similar in the mid context than the high and low contexts, and the T4/T6 were more similar in the low context compared to the mid context. The discrepancy between the LDA and GCA results can be explained by the fact that the LDA data focused on the later parts of the tone contours, while the GCA involved the whole contours. It can be seen in Appendix F that both the T3/T6 and T4/T6 tone pairs were more similar at the onset of the tones in the respective tonal contexts.

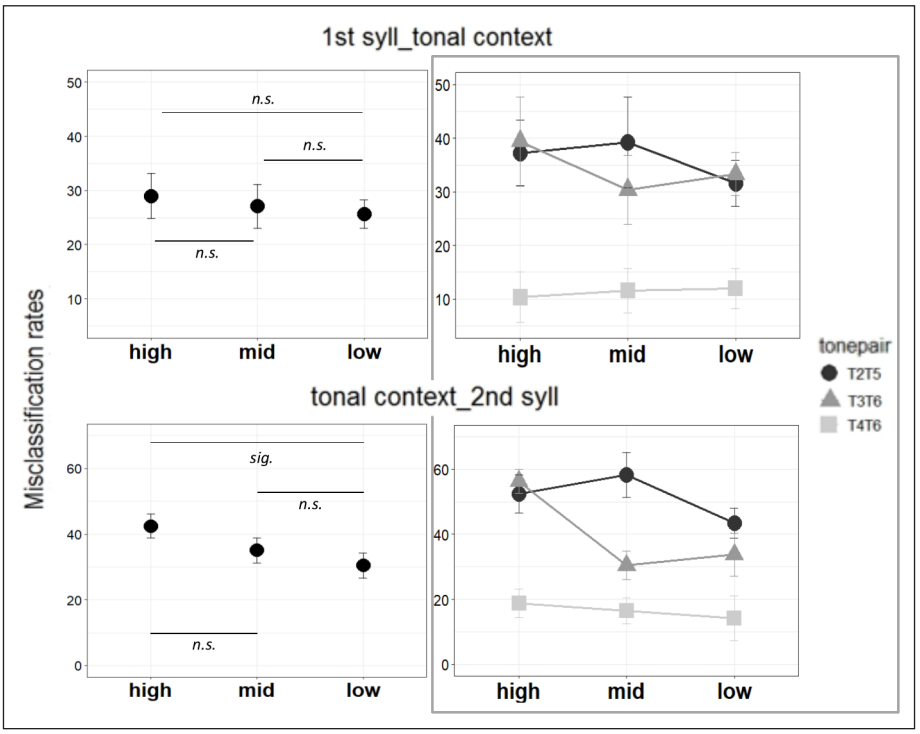



**Figure 5:** Effects of tonal context following target tones on the first syllables (upper), tonal context preceding target tones on the second syllables (lower), and tonal context × tone pair interactions (highlighted in a grey box), as produced by the 17 merging speakers.

Second, for the tonal contexts preceding the target tones on the second syllables (carryover coarticulation), the lower panel in **Figure 5** shows that the misclassification rate was significantly higher in the high tonal context than in the low tonal context (b = 9.07,

CrI = [0.36, 17.72], pd(b > 0) = 97.88%). Moreover, the effect of tonal context varied with different tone pairs. The overall less accurate classification in the high tonal context may be related to the significant comparisons of T3/T6 between high and mid as well as between high and low contexts (i.e., a higher misclassification rate in a high tonal context than in mid or low contexts). This is confirmed by post-hoc comparisons for the Bayesian mixed model (b = 25.95, CrI = [11.98, 39.63], pd(b > 0) = 100%; b = –20.99, CrI = [–34.29, –7.59], pd(b < 0) = 99.92%). Cantonese T3 [33] and T6 [22] begin with relatively low tone values, so the preceding high tonal context, which is much more distinct from the low tones, is likely to make the small difference between the starting points of T3 and T6 even smaller, thus leading to less separation after a high tonal context for the target tones of T3/T6 appearing in the second syllables of disyllabic words. It is also related to the finding that the tone pairs appearing as the second syllables were more similar due to stronger carryover coarticulation.

For the other tone pairs, T2 [25] and T5 [23], as well as T4 [21] and T6 [22], although they also begin with a low tone value, the LDA results showed that the preceding high tonal context did not influence them differently. Nevertheless, the GCA results indicated that all three tone pairs were more similar in a higher tonal context (T2/T5 being more similar in the high vs. the low contexts; T3/T6 being more similar in the high vs. both the mid and low contexts; T4/T6 being more similar in both the high and mid vs. the low contexts). This can be explained again by the fact that GCA included the whole contours, so the earlier carryover coarticulatory effect on the tone contours can be better captured.

## 3.4. Effects of word frequency

The misclassification rates of high- and low-token frequency disyllabic words are presented in **Table 5**. As is shown in the upper panel of **Figure 6**, there was no effect of word frequency (b = –1.99, CrI = [–10.57, 6.31], pd(b < 0) = 68.42%) and interaction between word frequency and tone pairs was not significant. There was also no significant interaction between the merging direction (e.g., T2→T5 compared with T5→T2) and the tone pairs.

The GCA results indicated that word frequency was significant for T2/T5, while there was no significant word frequency effect for the T3/T6 and T4/T6 tone pairs. The GCA contour data (Appendix F) indicates that while there was not much difference for T5 in high vs. low frequency words, the whole T2 contour was shifted downwards in low frequency words, closer to the pitch range of T5. Thus, T2/T5 were more similar in low frequency than high frequency words.

| | High frequency_misclassification rates (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 | Overall |
| M1 | 38.9 | 38.9 | 22.2 | 19.4 | 5.9 | 0.0 | 20.9 |
| M2 | 25.0 | 19.4 | 16.7 | 19.4 | 0.0 | 11.1 | 15.3 |
| M3 | 31.4 | 13.9 | 11.4 | 36.1 | 6.9 | 0.0 | 16.6 |
| M4 | 16.7 | 34.3 | 13.9 | 16.7 | 21.2 | 11.1 | 19.0 |
| M5 | 38.9 | 19.4 | 11.1 | 14.3 | 3.0 | 0.0 | 14.5 |
| M6 | 27.8 | 24.2 | 19.4 | 17.6 | 8.8 | 5.9 | 17.3 |
| M7 | 25.0 | 36.1 | 16.7 | 11.4 | 11.1 | 8.6 | 18.2 |
| M8 | 27.8 | 25.0 | 31.4 | 36.1 | 6.1 | 13.9 | 23.4 |
| M9 | 22.9 | 31.4 | 16.7 | 25.0 | 19.4 | 8.3 | 20.6 |
| M10 | 27.8 | 16.7 | 13.9 | 37.1 | 6.9 | 5.7 | 18.0 |
| M11 | 22.2 | 19.4 | 22.2 | 25.0 | 6.3 | 0.0 | 15.9 |
| M12 | 19.4 | 25.0 | 41.7 | 38.9 | 9.4 | 0.0 | 22.4 |
| M13 | 32.4 | 38.9 | 11.1 | 22.9 | 36.7 | 25.7 | 28.0 |
| M14 | 38.9 | 22.2 | 19.4 | 13.9 | 11.1 | 13.9 | 19.9 |
| M15 | 19.4 | 13.9 | 17.6 | 16.7 | 0.0 | 2.8 | 11.7 |
| M16 | 30.6 | 38.9 | 33.3 | 30.6 | 5.7 | 0.0 | 23.2 |
| M17 | 30.6 | 52.8 | 11.1 | 11.4 | 3.2 | 0.0 | 18.2 |
| M mean | 28.0 | 27.7 | 19.4 | 23.1 | 9.5 | 6.3 | 19.0 |

| | Low frequency_misclassification rates (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 | Overall |
| M1 | 22.2 | 13.9 | 36.1 | 19.4 | 7.4 | 5.6 | 17.4 |
| M2 | 27.8 | 27.8 | 36.1 | 19.4 | 8.8 | 2.8 | 20.5 |
| M3 | 22.2 | 8.8 | 13.9 | 16.7 | 6.9 | 0.0 | 11.4 |
| M4 | 25.0 | 26.5 | 25.0 | 11.1 | 3.0 | 2.8 | 15.6 |
| M5 | 27.8 | 19.4 | 21.2 | 19.4 | 9.4 | 0.0 | 16.2 |
| M6 | 14.3 | 20.0 | 34.3 | 20.0 | 0.0 | 0.0 | 14.8 |
| M7 | 32.4 | 28.6 | 22.2 | 13.9 | 32.1 | 0.0 | 21.5 |
| M8 | 27.8 | 16.7 | 33.3 | 36.1 | 0.0 | 2.8 | 19.5 |
| M9 | 14.3 | 8.3 | 13.9 | 13.9 | 29.0 | 19.4 | 16.5 |
| M10 | 19.4 | 8.3 | 16.7 | 18.2 | 14.8 | 0.0 | 12.9 |
| M11 | 25.0 | 22.2 | 36.1 | 33.3 | 11.8 | 0.0 | 21.4 |
| M12 | 8.6 | 20.6 | 38.9 | 63.9 | 3.2 | 0.0 | 22.5 |
| M13 | 20.0 | 30.6 | 16.7 | 30.6 | 34.4 | 33.3 | 27.6 |
| M14 | 13.9 | 25.0 | 27.8 | 27.8 | 11.1 | 5.6 | 18.5 |
| M15 | 8.3 | 9.7 | 17.1 | 19.4 | 8.8 | 2.8 | 11.0 |
| M16 | 27.8 | 34.3 | 33.3 | 31.4 | 3.2 | 0.0 | 21.7 |
| M17 | 36.1 | 30.6 | 2.8 | 8.3 | 3.4 | 0.0 | 13.5 |
| M mean | 21.9 | 20.7 | 25.0 | 23.7 | 11.0 | 4.4 | 17.8 |

**Table 5:** Misclassification rates (%) of the merging tone pairs with high-frequency (left) and low-frequency (right) words as produced by the 17 merging speakers. Misclassification rates over 20% are shaded grey for reference.

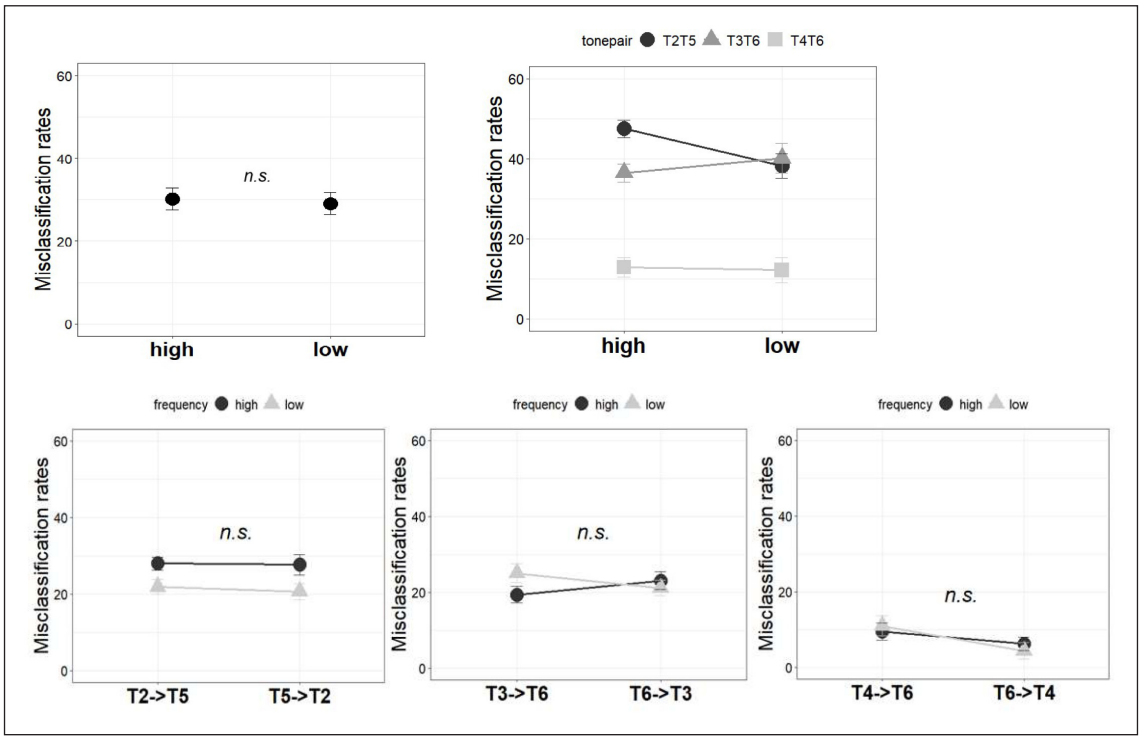



**Figure 6:** Effects of word frequency and word frequency × tone pair interactions (upper) and word frequency × merging direction interactions (lower) for different tone pairs, as produced by the 17 merging speakers.

## 4. Discussion

The results above illustrate that tone merging in Hong Kong Cantonese is subject to powerful phonetic influences. As predicted, the merging tone pairs were more similar in disyllabic words than in monosyllabic words due to tonal coarticulation, and they were the most similar in minimal pairs with identical segments (with the rising tone pair T2/T5 completely merged in minimal pairs, as illustrated in **Figure 1**). The higher similarity in disyllabic words and minimal pairs was found in both the reference and merging speakers (see **Tables 3** and **4**). Tonal coarticulation was asymmetric in that carryover coarticulation (i.e., the target tone being the second syllable) was stronger than anticipatory coarticulation. A high tonal context rendered the tone pairs, especially the level tone pair T3/T6, to be more similar than in other tonal contexts, with possible reasons discussed in Section 3.3 above. Finally, the effects of word token frequency did not influence tone merging in general.

As shown in previous studies using monosyllabic words (Mok et al., 2013; Zhang 2019), tone merging in Hong Kong Cantonese is at an incipient stage. The inclusion of disyllabic data in the current study provides a wider perspective for us to consider the tone-merging phenomenon

more comprehensively. An important question to ask is are the tone pairs more merged or are they just more coarticulated in disyllabic words? This is not a trivial question, as it touches on the core nature of the phenomenon of reduced tonal contrast. In other words, is the reduced tonal contrast due to representational change (merging), or is it a phonetic phenomenon (coarticulation)? Monosyllabic data only allow the merging perspective, while disyllabic data can be interpreted in both ways. There is evidence in the data to show that the observed reduced tonal contrast is not just a phonetic phenomenon. The positive correlations between monosyllabic and disyllabic misclassification rates across the three tone pairs and within each tone pair (see **Figure 3**) demonstrate that individuals who produced more similar tones in monosyllabic words also merged tones more frequently in disyllabic words. Moreover, the interaction between tonal context and tone pair (Section 3.3) illustrates that while anticipatory coarticulation was weak, tonal context significantly interacted with the tone pairs in stronger carryover coarticulation so tone merging and tonal coarticulation can have different effects. Thus, the representational tone-merging phenomenon is real (as already shown in previous studies using monosyllabic words with no coarticulation). Phonetic tonal coarticulation simply renders the tone pairs to be even more similar in disyllabic words. Carryover coarticulation preceded by a high tonal context is the most vulnerable condition for reduced tonal contrast.

As tone merging and tonal coarticulation coexist in the synchronic data, a logical question to ask is which phenomenon comes first? It is reasonable to assume that coarticulation precedes merging as coarticulation is also commonly found in languages not undergoing any sound change. Furthermore, Yang and Xu (2019) found strong cross-linguistic tendencies in tone change directionality among 45 diverse languages, which they argued had an articulatory basis in tonal coarticulation and truncation in connected speech. In fact, coarticulation has been proposed to be an important source of sound change. Both Ohala (1981, 1983, 1993) and Beddor (2009, 2012) argued that sound change occurs because listeners do not parse the coarticulatory effect with the source that gave rise to it, although the suggested perceptual mechanisms behind this differ in their two accounts. Ohala and Beddor mainly discussed coarticulatory effects on segmental sound changes, while our data illustrate that the same principles can be found in suprasegmental sound changes as well. Li et al.'s (2020) data independently corroborate our findings by demonstrating how carryover coarticulation could reduce the acoustic difference between the offsets of T2 and T5. They also argued that the overall lowering of T2 in connected speech could reflect an ongoing sound change in Cantonese. Their data from three age groups reflect the reduced contrast between the offsets of T2 and T5 from senior to middle-age to young speakers (Li & Guan, 2019), mirroring the patterns caused by carryover coarticulation. Our data additionally show how the other two tone pairs, especially the T3/T6 pair, became more similar due to a high tonal context in carryover coarticulation. Thus, if only monosyllabic and disyllabic words are considered, we can reasonably speculate that the possible origins of tone

merging in Cantonese might have started with coarticulation in the second syllables of disyllabic words instead of in monosyllabic words. Future studies can include words of various length to further investigate how tonal coarticulation and tone merging may interact in different syllable positions.

There are other reasons why the second syllables are more conducive to sound change in addition to coarticulation, or conversely, why the first syllables are more resistant to change. The first syllable of a disyllabic word, i.e., the word-initial syllable, enjoys some privilege over the second/word-final syllable in several ways. First, the word-initial syllable is crucial in lexical retrieval and word recognition (Browman, 1978; Marslen-Wilson & Welsh, 1978; Meyer, 1990), thus its psycholinguistic prominence over the second syllable makes it more likely for them to be pronounced more distinctly. Common sound change patterns and phonological processes found in word-final position (e.g., the loss of final nasal/stop consonants, neutralization of voicing contrasts) also support the point that the word-final syllable is a weaker position conducive to change.

Second, domain-initial strengthening is a robust phonetic phenomenon that can be observed at different prosodic levels (Cho, 2004; Cho & Keating, 2001). Its effects extend beyond the first segment (Byrd, 2000; Fougeron & Keating, 1997) and can also be found with lexical tones, for example, in Taiwan Southern Min (Pan, 2007) and Thai (Silpachai, 2024). Thus, domain-initial strengthening can affect the whole syllable. Disyllabic words in Cantonese often coincide with the prosodic unit of a phonological word (Wong, 2006a; Wong et al., 2005) which is further down the prosodic hierarchy. Thus, it is conceivable that the word-initial syllables are stronger and more resistant to change and coarticulation due to domain-initial strengthening.

The disyllabic data not only confirm that the tone merging phenomenon is real and might have started with the second syllables of disyllable words, they also illustrate that tone merging is dynamic and variegated, and is not a strictly categorical phenomenon, at least at the incipient stage. Synchronic variation caused by individual differences is an important source of sound change. Yu and Zellou (2019) discussed in detail how individuals may differ in various dimensions and how such differences may contribute to sound change. Nevertheless, in addition to the between-speaker differences which Yu and Zellou expounded on, variation can be found within speakers as well, as illustrated by our data. In contrast to the monosyllabic data, the differences between the reference and merging speakers are considerably blurred in the disyllabic data (**Table 3**), in that the two reference speakers also had noticeable misclassification rates closer to some of the merging speakers. This is noteworthy as the two reference speakers were professional speakers (one being a speech therapist cum researcher, one being a phonetician) who clearly distinguished the six Cantonese tones in their speech when producing monosyllabic words. The effects of tonal context seem to be more pronounced on the reference speakers, probably because their tones were better differentiated in monosyllabic words to begin with (their misclassification

rates were much lower in monosyllabic words), so tonal context could exert a stronger influence on them. Still, when comparing minimal pairs versus non-minimal pairs (**Table 4**), the two reference speakers could distinguish the tone pairs better than the merging speakers, albeit with a reduced difference. It was not possible to do any statistical comparisons to confirm the above observations with only two reference speakers, but the data do suggest that although they were similarly affected by tonal context, the reference speakers still generally outperformed the merging speakers in the condition in which the tone pairs were the most difficult to distinguish (i.e., minimal pairs, the right panel in **Table 4**). If the two chosen professional speakers also demonstrated such within speaker variation in an experimental setting, it is conceivable that the within-speaker variation for other speakers would be larger, particularly in natural conversation, and that the difference between merging and non-merging speakers would likely be on a moving continuum depending on the types of speech data being considered, e.g., experimental vs. natural settings, monosyllabic vs. disyllabic words. Currently, there is very little work on within-speaker variation and sound change, but it is a promising direction for future research. Thus, the recent emphasis on individual variation should be expanded to examine how within-speaker variation may interact with between-speaker individual differences and contribute to sound change as well.

The coarticulatory basis and individual variation discussed above clearly demonstrate the dynamic and continuous nature of tone merging, which is a real phenomenon. One interesting question to ask is when and how the dynamic and continuous phonetic phenomenon might become categorical and phonological. Also, for how long can the continuous and categorical patterns coexist? Obviously, having only production data cannot answer the questions satisfactorily, as perception is an important, if not the most important, part of sound change (Beddor, 2009, 2023; Ohala, 1981, 1993). Would similar within-speaker variation also be found in perceptual patterns? Would the production-perception link be on a moving continuum according to different types of speech data as well? Answers to these interesting questions can only be found by further studies including both production and perception data in rich phonetic contexts.

The co-existence of coarticulation and merging, and the within- and between-speaker variation shown in our data, can be understood using the hybrid exemplar-model of speech production as described in Pierrehumbert (2002). Since tone merging is still in an incipient stage in Hong Kong, it is reasonable to expect that the phonological representations of the tones are not altered yet (or at least not completely altered), while the observed variations can be explained by the phonetic implementations influenced by different factors affecting the weights and activations of the stored exemplars, which are subject to incremental updating. It can be assumed that the exemplars with reduced phonetic contrast are located at the more peripheral regions between the category labels. As the long-term representations of words (or higher phonological abstractions [e.g., tones], which are incorporated in the hybrid model)

include probability distributions over phonetic outcomes, with increased exposure to more similar exemplars encountered in both production and perception due to the variations intrinsic to connected speech (contextual coarticulation), it is conceivable that the mapping between the category labels and the more frequent or activated exemplars can gradually shift towards the originally peripheral regions, resulting in a representational change over time (merging).

One expected outcome of the above hybrid exemplar model is a strong frequency effect on phonetic reduction, with high frequency words being more lenited or less distinct than low frequency words. However, no token frequency effect was observed in the LDA results, while T2 was shifted down in low frequency words in the GCA results. It may look contradictory to the model prediction at first, but Pierrehumbert (2002) noted that for low frequency words, the proportion of exposures which occurred in the context of an experiment would be higher than for more common words (in our case, equal number of high vs. low frequency words). Thus, the proportional effect of phonetic reduction due to contextual variation would be higher for low frequency words than for high frequency words, which was indeed the case in the CGA results. Another possible reason for the lack of token frequency effect (also reported in Mok et al. [2013]) is that the frequency count was based on a written corpus, which may not reflect spoken token frequency, as well as some of the target words, which are mostly used in written contexts, or that the token frequency differences between the high vs. low words need to be enlarged for the frequency effect to surface due to the written nature of the corpus.

Finally, our disyllabic data also illustrate that the progression of merging is different among the three tone pairs. The T2/T5 pair is the most advanced in merging (with a complete overlap in minimal pairs), followed by the T3/T6 pair. The T4/T6 pair is the slowest. Such patterns within Hong Kong Cantonese are also mirrored by the tone-merging patterns in the three varieties of Zhuhai, Macao and Hong Kong Cantonese, discussed in the Introduction and reported in Zhang (2019), who did not find any speakers merging the T4/T6 pair in any of the three Cantonese varieties. The slower progression of the T4/T6 pair can be explained by the additional cue of creaky voice in T4 for better differentiation (Fung & Wong, 2023; Yu & Lam, 2014). Interestingly, our disyllabic data also show a larger difference between the T4/T6 pair than the other two tone pairs (see **Figure 1** and **Table 3**). The faster rate of T2/T5 might be related to the difference in type frequency (not token frequency, as discussed above). According to Fok-Chan (1974), T2 is the most frequent tone in Cantonese, while Leung et al. (2004) also showed that T2 is among the more frequent tones. Both studies show that T3 and T6 had comparable type frequency. Being more frequent in speech, it is easy for the high rising offset of T2 to undershoot in conversational speech and become closer to T5 (see **Figure 1**). Li et al. (2020) showed exactly such a pattern. Our GCA data also found that T2 was shifted down in low token frequency words. **Tables 3** and **4** in the current study also show that even the two reference speakers were more affected for the T2/T5 pair than the T3/T6 pair for monosyllabic versus disyllabic words and minimal pairs

versus non-minimal pairs. Whether or not other factors cause different rates of merging among the three tone pairs awaits further investigation.

In conclusion, the disyllabic data in the current study provides an opportunity to investigate the ongoing tone merging in Cantonese more comprehensively and also offers a new perspective on suprasegmental data, which allows us to revisit some important issues in the sound change literature, such as the sources of sound change and individual variation. As mentioned above, only disyllabic production data were included. While perception and the perception-production link are important in understanding sound change, more focused research with both production and perception data are needed to answer the questions of why and how sound change happens.

## Appendices

| High Frequency (σ1 Target) | | | | | High Frequency (σ2 Target) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Target words** | **Jyutping** | **Gloss** | **σ1 tone** | **σ2 tone** | **Target words** | **Jyutping** | **Gloss** | **σ1 tone** | **σ2 tone** |
| 之間 | zi1 gaan1 | between | 1 | 1 | 應該 | jing1 goi1 | should | 1 | 1 |
| 因此 | jan1 ci2 | therefore | 1 | 2 | 許多 | heoi2 do1 | many | 2 | 1 |
| 經濟 | ging1 zai3 | economy | 1 | 3 | 對於 | deoi3 jyu1 | for | 3 | 1 |
| 他們 | taa1 mun4 | they | 1 | 4 | 由於 | jau4 jyu1 | due to | 4 | 1 |
| 參與 | caam1 jyu5 | participate | 1 | 5 | 已經 | ji5 ging1 | already | 5 | 1 |
| 因爲 | jan1 wai6 | because | 1 | 6 | 第三 | dai6 saam1 | third | 6 | 1 |

**Appendix A:** Sample word list for target Tone 1.

Since there were only two reference speakers for comparison, their data were insufficient to examine the effects of various factors like syllable position (Appendix B), and tonal context (Appendices C and D) on tone merging. Therefore, only data from the 17 merging speakers were included in the following misclassification rates tables.

| | First syllable_misclassification rates (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **T2→T5** | **T5→T2** | **T3→T6** | **T6→T3** | **T4→T6** | **T6→T4** | **Over-all** |
| M1 | 19.4 | 8.3 | 22.2 | 8.3 | 2.8 | 0.0 | 10.2 |
| M2 | 30.6 | 16.7 | 13.9 | 11.1 | 5.6 | 0.0 | 13.0 |
| M3 | 8.3 | 14.7 | 8.3 | 16.7 | 6.7 | 0.0 | 9.1 |
| M4 | 22.2 | 34.3 | 16.7 | 16.7 | 2.8 | 2.8 | 15.9 |
| M5 | 30.6 | 13.9 | 18.2 | 11.1 | 0.0 | 0.0 | 12.3 |
| M6 | 16.7 | 11.8 | 22.2 | 14.7 | 2.9 | 0.0 | 11.4 |
| M7 | 16.7 | 5.7 | 11.1 | 8.3 | 14.3 | 2.8 | 9.8 |
| M8 | 13.9 | 11.1 | 33.3 | 38.9 | 0.0 | 2.8 | 16.7 |
| M9 | 20.0 | 13.9 | 25.0 | 22.2 | 20.0 | 11.1 | 18.7 |
| M10 | 11.1 | 11.1 | 13.9 | 11.8 | 2.9 | 0.0 | 8.5 |
| M11 | 11.1 | 2.8 | 25.0 | 13.9 | 5.6 | 0.0 | 9.7 |
| M12 | 11.4 | 11.4 | 36.1 | 36.1 | 8.8 | 0.0 | 17.3 |
| M13 | 38.9 | 25.0 | 11.1 | 28.6 | 25.8 | 22.9 | 25.4 |
| M14 | 30.6 | 30.6 | 25.0 | 11.1 | 8.3 | 2.8 | 18.1 |
| M15 | 8.3 | 0.0 | 14.3 | 11.1 | 2.9 | 5.6 | 7.0 |
| M16 | 30.6 | 33.3 | 38.9 | 25.0 | 0.0 | 0.0 | 21.3 |
| M17 | 44.4 | 30.6 | 5.6 | 2.9 | 0.0 | 0.0 | 13.9 |
| M mean | 21.5 | 16.2 | 20.0 | 17.0 | 6.4 | 3.0 | 14.0 |

| | Second syllable_misclassification rates (%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **T2→T5** | **T5→T2** | **T3→T6** | **T6→T3** | **T4→T6** | **T6→T4** | **Over-all** |
| M1 | 30.6 | 30.6 | 25.0 | 13.9 | 4.0 | 0.0 | 17.4 |
| M2 | 36.1 | 22.2 | 36.1 | 33.3 | 6.1 | 11.1 | 24.2 |
| M3 | 20.0 | 19.4 | 20.0 | 13.9 | 3.6 | 0.0 | 12.8 |
| M4 | 25.0 | 38.2 | 19.4 | 13.9 | 6.7 | 2.8 | 17.7 |
| M5 | 41.7 | 30.6 | 13.9 | 14.3 | 3.4 | 0.0 | 17.3 |
| M6 | 25.7 | 44.1 | 40.0 | 20.0 | 12.5 | 2.9 | 24.2 |
| M7 | 32.4 | 27.8 | 16.7 | 20.0 | 15.0 | 0.0 | 18.7 |
| M8 | 22.2 | 22.2 | 20.0 | 30.6 | 21.4 | 0.0 | 19.4 |
| M9 | 17.1 | 14.3 | 13.9 | 13.9 | 15.6 | 11.1 | 14.3 |
| M10 | 36.1 | 16.7 | 13.9 | 5.9 | 13.6 | 0.0 | 14.4 |
| M11 | 19.4 | 33.3 | 30.6 | 33.3 | 16.7 | 0.0 | 22.2 |
| M12 | 16.7 | 28.6 | 47.2 | 33.3 | 10.3 | 0.0 | 22.7 |
| M13 | 24.2 | 27.8 | 16.7 | 30.6 | 38.7 | 22.2 | 26.7 |
| M14 | 27.8 | 13.9 | 13.9 | 36.1 | 8.3 | 8.3 | 18.1 |
| M15 | 5.6 | 12.9 | 17.6 | 16.7 | 0.0 | 0.0 | 8.8 |
| M16 | 19.4 | 20.0 | 33.3 | 25.7 | 6.5 | 0.0 | 17.5 |
| M17 | 44.4 | 50.0 | 5.6 | 13.9 | 20.0 | 0.0 | 22.3 |
| M mean | 26.1 | 26.6 | 22.6 | 21.7 | 11.9 | 3.4 | 18.7 |

**Appendix B:** Misclassification rates (%) of the merging tone pairs of the 17 merging speakers with the target tone on the first (left) and second (right) syllables. Misclassification rates over 20% are shaded grey.

| | 1st_high_misclassification rates (%) | | | | | |
|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 |
| M1 | 16.7 | 16.7 | 16.7 | 0.0 | 0.0 | 0.0 |
| M2 | 33.3 | 0.0 | 16.7 | 0.0 | 16.7 | 0.0 |
| M3 | 16.7 | 0.0 | 0.0 | 16.7 | 16.7 | 0.0 |
| M4 | 33.3 | 16.7 | 33.3 | 50.0 | 0.0 | 0.0 |
| M5 | 16.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| M6 | 16.7 | 16.7 | 33.3 | 33.3 | 0.0 | 0.0 |
| M7 | 33.3 | 0.0 | 0.0 | 16.7 | 33.3 | 0.0 |
| M8 | 33.3 | 0.0 | 0.0 | 0.0 | 0.0 | 16.7 |
| M9 | 33.3 | 50.0 | 0.0 | 16.7 | 75.0 | 0.0 |
| M10 | 0.0 | 0.0 | 16.7 | 20.0 | 0.0 | 0.0 |
| M11 | 33.3 | 16.7 | 16.7 | 16.7 | 0.0 | 0.0 |
| M12 | 50.0 | 16.7 | 33.3 | 66.7 | 0.0 | 0.0 |
| M13 | 50.0 | 16.7 | 33.3 | 33.3 | 0.0 | 16.7 |
| M14 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| M15 | 0.0 | 0.0 | 33.3 | 33.3 | 0.0 | 0.0 |
| M16 | 33.3 | 33.3 | 50.0 | 50.0 | 0.0 | 0.0 |
| M17 | 33.3 | 16.7 | 33.3 | 0.0 | 0.0 | 0.0 |
| M mean | 25.5 | 11.8 | 18.6 | 20.8 | 8.3 | 2.0 |

| | 1st_mid_misclassification rates (%) | | | | | |
|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 |
| M1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| M2 | 50.0 | 16.7 | 16.7 | 0.0 | 16.7 | 0.0 |
| M3 | 16.7 | 0.0 | 33.3 | 50.0 | 0.0 | 0.0 |
| M4 | 0.0 | 0.0 | 0.0 | 16.7 | 16.7 | 0.0 |
| M5 | 33.3 | 0.0 | 16.7 | 16.7 | 0.0 | 0.0 |
| M6 | 16.7 | 16.7 | 0.0 | 0.0 | 0.0 | 0.0 |
| M7 | 0.0 | 0.0 | 16.7 | 33.3 | 33.3 | 0.0 |
| M8 | 16.7 | 16.7 | 33.3 | 33.3 | 0.0 | 0.0 |
| M9 | 0.0 | 16.7 | 33.3 | 16.7 | 40.0 | 0.0 |
| M10 | 33.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| M11 | 16.7 | 50.0 | 50.0 | 0.0 | 0.0 | 0.0 |
| M12 | 0.0 | 0.0 | 50.0 | 16.7 | 0.0 | 0.0 |
| M13 | 33.3 | 33.3 | 0.0 | 0.0 | 16.7 | 0.0 |
| M14 | 66.7 | 33.3 | 0.0 | 16.7 | 40.0 | 16.7 |
| M15 | 16.7 | 0.0 | 16.7 | 0.0 | 0.0 | 0.0 |
| M16 | 50.0 | 33.3 | 33.3 | 0.0 | 16.7 | 0.0 |
| M17 | 50.0 | 50.0 | 16.7 | 0.0 | 0.0 | 0.0 |
| M mean | 23.5 | 15.7 | 18.6 | 11.8 | 10.6 | 1.0 |

| | 1st_low_misclassification rates (%) | | | | | |
|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 |
| M1 | 16.7 | 8.3 | 25.0 | 12.5 | 4.2 | 0.0 |
| M2 | 33.3 | 16.7 | 12.5 | 8.3 | 0.0 | 0.0 |
| M3 | 8.3 | 12.5 | 16.7 | 16.7 | 0.0 | 0.0 |
| M4 | 20.8 | 30.4 | 12.5 | 4.2 | 16.7 | 4.2 |
| M5 | 16.7 | 4.2 | 9.5 | 12.5 | 4.2 | 0.0 |
| M6 | 16.7 | 13.6 | 16.7 | 13.0 | 4.2 | 0.0 |
| M7 | 29.2 | 8.7 | 8.3 | 4.2 | 13.0 | 8.3 |
| M8 | 4.2 | 4.2 | 33.3 | 33.3 | 0.0 | 0.0 |
| M9 | 13.0 | 20.8 | 20.8 | 25.0 | 28.6 | 4.2 |
| M10 | 12.5 | 12.5 | 12.5 | 8.7 | 8.7 | 0.0 |
| M11 | 8.3 | 4.2 | 20.8 | 8.3 | 8.3 | 0.0 |
| M12 | 4.3 | 8.7 | 25.0 | 37.5 | 18.2 | 0.0 |
| M13 | 25.0 | 20.8 | 0.0 | 26.1 | 28.6 | 30.4 |
| M14 | 25.0 | 33.3 | 25.0 | 8.3 | 5.0 | 8.3 |
| M15 | 0.0 | 0.0 | 17.4 | 8.3 | 4.2 | 4.2 |
| M16 | 20.8 | 29.2 | 37.5 | 25.0 | 0.0 | 0.0 |
| M17 | 29.2 | 25.0 | 8.3 | 12.5 | 0.0 | 0.0 |
| M mean | 16.7 | 14.9 | 17.8 | 15.6 | 8.5 | 3.5 |

**Appendix C:** Misclassification rates (%) of the merging tone pairs of the 17 merging speakers in different tonal contexts (high, mid and low from left to right) following the first syllables with target tones. Misclassification rates over 20% are shaded grey.

| | high_2nd_misclassification rates (%) | | | | | |
|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 |
| M1 | 8.3 | 8.3 | 58.3 | 25.0 | 0.0 | 0.0 |
| M2 | 41.7 | 50.0 | 33.3 | 33.3 | 30.0 | 8.3 |
| M3 | 16.7 | 0.0 | 45.5 | 33.3 | 0.0 | 0.0 |
| M4 | 8.3 | 20.0 | 25.0 | 16.7 | 0.0 | 0.0 |
| M5 | 25.0 | 25.0 | 8.3 | 27.3 | 10.0 | 0.0 |
| M6 | 54.5 | 36.4 | 25.0 | 33.3 | 33.3 | 8.3 |
| M7 | 25.0 | 25.0 | 33.3 | 16.7 | 0.0 | 0.0 |
| M8 | 33.3 | 16.7 | 36.4 | 41.7 | 14.3 | 0.0 |
| M9 | 16.7 | 27.3 | 25.0 | 33.3 | 9.1 | 8.3 |
| M10 | 41.7 | 16.7 | 16.7 | 18.2 | 28.6 | 0.0 |
| M11 | 33.3 | 50.0 | 33.3 | 33.3 | 27.3 | 0.0 |
| M12 | 16.7 | 27.3 | 33.3 | 25.0 | 10.0 | 0.0 |
| M13 | 33.3 | 41.7 | 25.0 | 16.7 | 27.3 | 25.0 |
| M14 | 25.0 | 25.0 | 33.0 | 25.0 | 18.2 | 33.3 |
| M15 | 8.3 | 14.3 | 18.2 | 25.0 | 0.0 | 0.0 |
| M16 | 8.3 | 36.4 | 16.7 | 36.4 | 9.1 | 0.0 |
| M17 | 25.0 | 50.0 | 8.3 | 41.7 | 18.2 | 0.0 |
| M mean | 24.8 | 27.7 | 27.9 | 28.3 | 13.8 | 4.9 |

| | mid_2nd_misclassification rates (%) | | | | | |
|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 |
| M1 | 16.7 | 16.7 | 8.3 | 25.0 | 0.0 | 0.0 |
| M2 | 50.0 | 33.3 | 33.3 | 33.3 | 0.0 | 0.0 |
| M3 | 45.5 | 33.3 | 16.7 | 8.3 | 0.0 | 0.0 |
| M4 | 25.0 | 58.3 | 0.0 | 8.3 | 20.0 | 0.0 |
| M5 | 50.0 | 25.0 | 16.7 | 8.3 | 0.0 | 0.0 |
| M6 | 25.0 | 33.3 | 16.7 | 16.7 | 9.1 | 0.0 |
| M7 | 36.4 | 25.0 | 25.0 | 18.2 | 28.6 | 0.0 |
| M8 | 58.3 | 33.3 | 16.7 | 33.3 | 25.0 | 0.0 |
| M9 | 25.0 | 8.3 | 8.3 | 0.0 | 30.0 | 16.7 |
| M10 | 41.7 | 16.7 | 0.0 | 0.0 | 37.5 | 0.0 |
| M11 | 25.0 | 8.3 | 8.3 | 8.3 | 9.1 | 0.0 |
| M12 | 33.3 | 41.7 | 25.0 | 25.0 | 10.0 | 0.0 |
| M13 | 9.1 | 16.7 | 16.7 | 25.0 | 8.3 | 33.3 |
| M14 | 33.3 | 8.3 | 8.3 | 8.3 | 11.1 | 0.0 |
| M15 | 8.3 | 0.0 | 8.3 | 16.7 | 8.3 | 0.0 |
| M16 | 16.7 | 16.7 | 25.0 | 25.0 | 0.0 | 0.0 |
| M17 | 58.3 | 58.3 | 8.3 | 16.7 | 33.3 | 0.0 |
| M mean | 32.8 | 25.5 | 14.2 | 16.3 | 13.5 | 2.9 |

| | low_2nd_misclassification rates (%) | | | | | |
|---|---|---|---|---|---|---|
| | T2→T5 | T5→T2 | T3→T6 | T6→T3 | T4→T6 | T6→T4 |
| M1 | 16.7 | 16.7 | 8.3 | 8.3 | 0.0 | 0.0 |
| M2 | 16.7 | 25.0 | 33.3 | 25.0 | 0.0 | 0.0 |
| M3 | 16.7 | 8.3 | 0.0 | 16.7 | 22.2 | 0.0 |
| M4 | 25.0 | 33.3 | 8.3 | 8.3 | 0.0 | 0.0 |
| M5 | 25.0 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| M6 | 33.3 | 45.5 | 54.5 | 18.2 | 0.0 | 0.0 |
| M7 | 9.1 | 8.3 | 0.0 | 8.3 | 0.0 | 0.0 |
| M8 | 33.3 | 25.0 | 16.7 | 25.0 | 33.3 | 25.0 |
| M9 | 36.4 | 33.3 | 8.3 | 16.7 | 27.3 | 8.3 |
| M10 | 25.0 | 16.7 | 16.7 | 9.1 | 0.0 | 0.0 |
| M11 | 0.0 | 25.0 | 50.0 | 50.0 | 0.0 | 0.0 |
| M12 | 33.3 | 16.7 | 41.7 | 25.0 | 11.1 | 0.0 |
| M13 | 30.0 | 16.7 | 8.3 | 16.7 | 87.5 | 16.7 |
| M14 | 16.7 | 16.7 | 16.7 | 25.0 | 9.1 | 0.0 |
| M15 | 16.7 | 8.3 | 9.1 | 0.0 | 0.0 | 0.0 |
| M16 | 8.3 | 8.3 | 25.0 | 16.7 | 0.0 | 0.0 |
| M17 | 41.7 | 25.0 | 8.3 | 0.0 | 0.0 | 0.0 |
| M mean | 22.6 | 20.8 | 18.0 | 15.8 | 11.2 | 2.9 |

**Appendix D:** Misclassification rates (%) of the merging tone pairs of the 17 merging speakers in different tonal contexts (high, mid and low from left to right) preceding the second syllables with target tones. Misclassification rates over 20% are shaded grey.

| Condition = disyllabic words | | | | |
|---|---|---|---|---|
| **Tone pairs** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2T5 vs T3T6** | –0.732 | 0.0407 | –17.968 | <.0001 |
| **T2T5 vs T4T6** | –1.825 | 0.0407 | –44.825 | <.0001 |
| **T3T6 vs T4T6** | –1.093 | 0.0498 | –21.935 | <.0001 |

| Condition = monosyllabic words | | | | |
|---|---|---|---|---|
| **Tone pairs** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2T5 vs T3T6** | –0.48 | 0.0308 | –15.583 | <.0001 |
| **T2T5 vs T4T6** | –1.122 | 0.0314 | –35.698 | <.0001 |
| **T3T6 vs T4T6** | –0.642 | 0.0382 | –16.818 | <.0001 |

| Condition = disyllabic vs. monosyllabic words | | | | |
|---|---|---|---|---|
| **Tone pairs** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2 vs T5** | –0.4043 | 0.0467 | –8.662 | <.0001 |
| **T3 vs T6** | –0.0463 | 0.0466 | –0.994 | 0.3201 |
| **T4 vs T6** | 0.312 | 0.0483 | 6.463 | <.0001 |

| Condition = minimal pairs vs. non-minimal pairs | | | | |
|---|---|---|---|---|
| **Tone pairs** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2 vs T5** | –0.1214 | 0.062 | –1.957 | 0.0504 |
| **T3 vs T6** | –0.0101 | 0.0623 | –0.163 | 0.8707 |

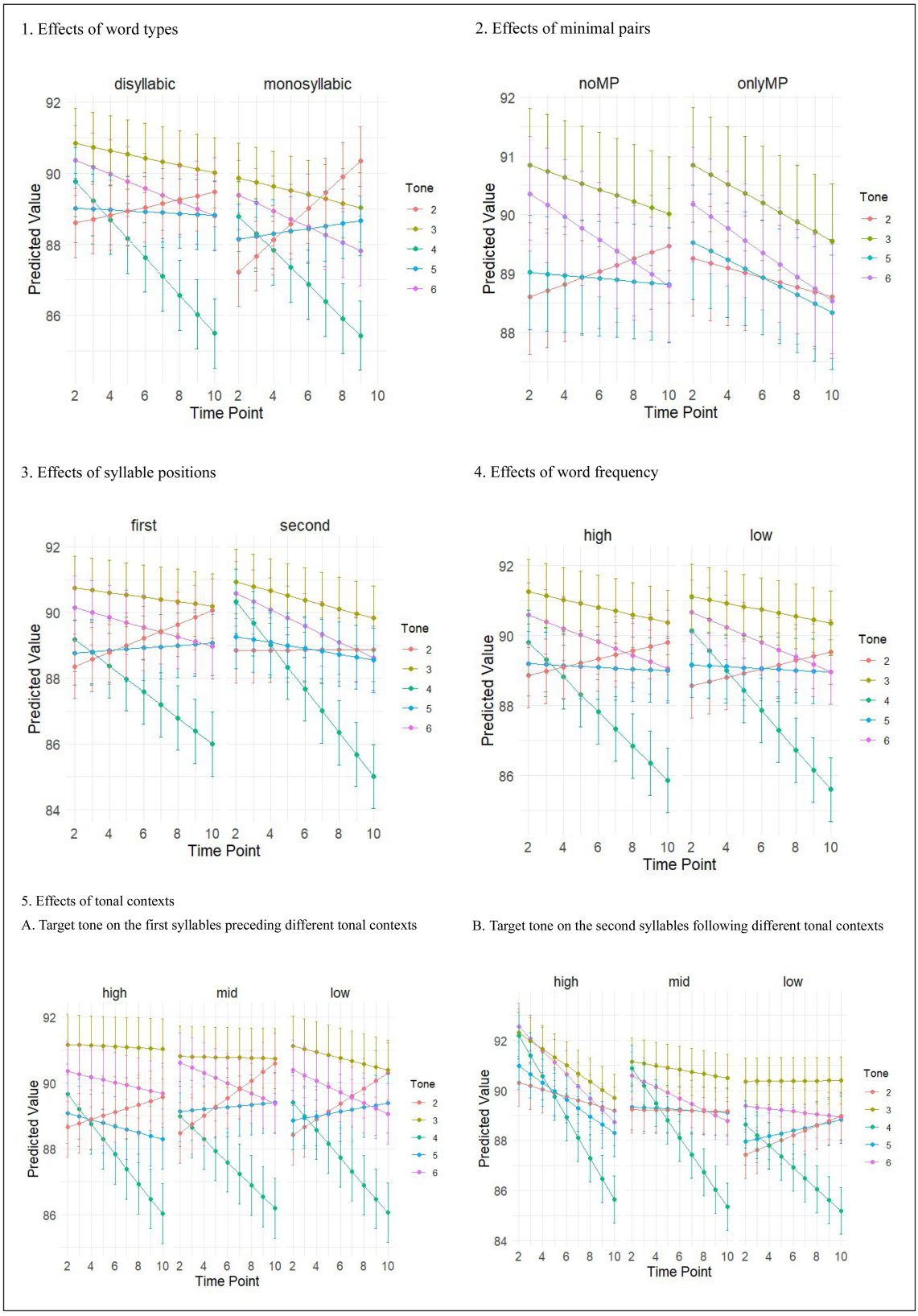
| Condition = target in 1st syllables vs. 2nd syllables | | | | |
|---|---|---|---|---|
| **Tone pairs** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2 vs T5** | 0.3372 | 0.0573 | 5.888 | <.0001 |
| **T3 vs T6** | 0.1264 | 0.0572 | 2.209 | 0.0272 |
| **T4 vs T6** | 0.0511 | 0.0572 | 0.893 | 0.3719 |

| Condition = 1st syll_tonal context | | | | | |
|---|---|---|---|---|---|
| **Tone pairs** | **Tonal contexts** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2 vs T5** | **high – low** | 0.1925 | 0.101 | 1.903 | 0.1377 |
| | **high – mid** | 0.1603 | 0.128 | 1.253 | 0.422 |
| | **low – mid** | –0.0322 | 0.101 | –0.319 | 0.9455 |
| **T3 vs T6** | **high – low** | 0.0534 | 0.101 | 0.528 | 0.8575 |
| | **high – mid** | 0.3024 | 0.128 | 2.361 | 0.0478 |
| | **low – mid** | 0.249 | 0.101 | 2.456 | 0.0374 |
| **T4 vs T6** | **high – low** | 0.185 | 0.101 | 1.829 | 0.16 |
| | **high – mid** | –0.2318 | 0.128 | –1.81 | 0.1663 |
| | **low – mid** | –0.4168 | 0.101 | –4.114 | 0.0001 |

| Condition = tonal context_2nd syll | | | | | |
|---|---|---|---|---|---|
| **Tone pairs** | **Tonal contexts** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2 vs T5** | **high – low** | –0.0727 | 0.0962 | 3.248 | 0.0033 |
| | **high – mid** | 0.153 | 0.0961 | 1.592 | 0.249 |
| | **low – mid** | –0.1594 | 0.0959 | –1.661 | 0.2204 |
| **T3 vs T6** | **high – low** | –0.8611 | 0.0959 | –8.98 | <.0001 |
| | **high – mid** | –0.7807 | 0.0958 | –8.146 | <.0001 |
| | **low – mid** | 0.0804 | 0.0958 | 0.839 | 0.6787 |
| **T4 vs T6** | **high – low** | –0.5245 | 0.0959 | –5.47 | <.0001 |
| | **high – mid** | 0.1419 | 0.0959 | 1.48 | 0.3005 |
| | **low – mid** | 0.6664 | 0.0959 | 6.95 | <.0001 |

| **Condition = high vs. low frequency words** | | | | |
|---|---|---|---|---|
| **Tone pairs** | **Estimate** | **SE** | **z.ration** | **p-value** |
| **T2 vs T5** | 0.2424 | 0.0575 | 4.994 | <.0001 |
| **T3 vs T6** | 0.0537 | 0.0575 | 0.934 | 0.3505 |
| **T4 vs T6** | –0.0649 | 0.0575 | –1.128 | 0.2592 |

**Appendix E:** GCA statistics: Differences between the merging tone pairs of the 17 merging speakers influenced by different factors. In Condition = disyllabic/monosyllabic words, we compared the differences among the three tone pairs. In other conditions, we examined the effects of word type (disyllabic vs. monosyllabic words), minimal pair (minimal pairs vs. non-minimal pairs), syllable position (first vs. second syllables), tonal context (following first syllables and preceding second syllables respectively), and word frequency (high vs. low frequency words). For example, in Condition = disyllabic vs. monosyllabic words, significant T2 vs T5 (Estimate = –0.4043, SE = 0.0467, $p < 0.0001$) means that this tone pair was more similar in the disyllabic words condition.




**Appendix F:** GCA graphs plotted using raw data points and model-fitted predictions. The solid lines represent the model-predicted values based on fixed effects from the linear mixed effects model, while the error bars indicate the 95% confidence intervals of these predictions. The graphs show tone contours in the merging tone pairs of the 17 merging speakers influenced by difference factors.

## Acknowledgements

The authors would like to thank Dr. Peggy Wong for her invaluable effort in the initial stage of the project and Summer Mut for her hard work in processing the acoustic data. They are also grateful for the helpful comments provided by the reviewers, which improved the manuscript in many ways. All remaining errors are our own.

## Competing interests

The authors have no competing interests to declare.

## References

Bauer, R., Cheung, K. H., & Cheung, P. M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change, 15*, 211–225. https://doi.org/10.1017/S0954394503152039

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Mouton de Gruyter. https://doi.org/10.1515/9783110823707

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*, 785–821. https://doi.org/10.1353/lan.0.0165

Beddor, P. S. (2012). Perception grammars and sound change. In M. J. Solé & D. Recasens (Eds.), *Initiation of sound change: Perception, production, and social factors* (pp. 37–55). John Benjamin. https://doi.org/10.1075/cilt.323.06bed

Beddor, P. S. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change. *Journal of Phonetics, 97*, 101228. https://doi.org/10.1016/j.wocn.2023.101228

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Browman, C. P. (1978). Tip of the tongue and slip of the ear. Implications for language processing. *UCLA Working Papers in Phonetics, 42*.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica, 57*, 3–16. https://doi.org/10.1159/000028456

Chan, S. D., & Tang, Z. X. (1990). Quantitative analysis of lexical distribution in different Chinese communities in the 1990's. *Yuyan Wenzi Yingyong (Applied Linguistics)*, *3*, 10–18.

Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique, 45*, 24–27.

Chao, Y. R. (1947). *Cantonese primer*. Greenwood Press. https://doi.org/10.4159/harvard.9780674732438

Chen, S., Wiltshire, C., & Li, B. (2018). An updated typology of tonal coarticulation properties. *Taiwan Journal of Linguistics, 16*, 79–114. https://doi.org/10.6519/TJL.2018.16(2).3

Chen, Y., & Li, Q. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics, 54*, 123–150. https://doi.org/10.1016/j.wocn.2015.10.002

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics, 32*, 141–176. https://doi.org/10.1016/S0095-4470(03)00043-3

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190. https://doi.org/10.1006/jpho.2001.0131

Fok-Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong University Press.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 106*, 3728–3740. https://doi.org/10.1121/1.418332

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *Journal of the Acoustical Society of America, 146*, EL424–EL430. https://doi.org/10.1121/1.5133661

Fung, S. Y., & Wong, E. Y. C. (2023). Separated and reunified: An apparent time investigation of the voice quality differences between Hong Kong Cantonese and Guangzhou Cantonese. *PLoS ONE, 18*, e0293058. https://doi.org/10.1371/journal.pone.0293058

Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics, 22*, 477–492. https://doi.org/10.1016/S0095-4470(19)30296-7

Gu, W., & Lee, T. (2007). Effects of tonal context and focus on Cantonese F0. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1033–1036, Saarbrücken.

Han, M. S., & Kim, K. O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics, 2*, 223–232. https://doi.org/10.1016/S0095-4470(19)31272-0

Kapatsinski, V. (2023). Understanding the roles of type and token frequency in usage-based linguistics. In M. Díaz-Campos & S. Balasch (Eds.), *The handbook of usage-based linguistics* (pp. 91–106).Wiley-Blackwell. https://doi.org/10.1002/9781119839859.ch5

Kej, J., Smyth, V., So, L. K. H., Lau, C. C., & Capell, K. (2002). Assessing the accuracy of production of Cantonese lexical tones: A comparison between perceptual judgement and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing, 7*, 25–38. https://doi.org/10.1179/136132802805576535

Laniran, Y. O. (1992). *Intonation in tone languages: The phonetic implementation of tones in Yoruba.* [Doctoral dissertation, Cornell University].

Law, S. P., Fung, R. S. Y., & Kung, C. (2013). An ERP study of good production vis-à-vis poor perception of tones in Cantonese: Implications for top-down speech processing. *PLoS ONE, 9*, e54396. https://doi.org/10.1371/journal.pone.0054396

Leung, M. T., Law, S. P., & Fung, R. S. Y. (2004). Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments & Computers, 36*, 500–505. https://doi.org/10.3758/BF03195596

Li, B., & Guan, Y. (2019). Generational differences in production of a tonal contrast in Hong Kong Cantonese. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)*, 186–190, Melbourne.

Li, B., Guan, Y., & Chen, S. (2020). Carryover effects on tones in Hong Kong Cantonese. In *Proceedings of Speech Prosody 2020*, 489–493, Tokyo. https://doi.org/10.21437/SpeechProsody.2020-100

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Meyer, A. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524–545. https://doi.org/10.1016/0749-596X(90)90050-A

Mok, P., Zuo, D., & Wong, P. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change, 25*, 341–370. https://doi.org/10.1017/S0954394513000161

Ohala, J. J. (1981). The listener as a source of sound change. In C. Masek, R. Hendrick & M. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago Linguistic Society, The University of Chicago.

Ohala, J. J. (1983). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). Longman.

Ohala, J. J. (1993). Coarticulation and phonology. *Language and Speech, 36*, 155–170. https://doi.org/10.1177/002383099303600303

Ou, J., & Law, S. P. (2016). Individual differences in processing pitch contour and rise time in adults: A behavioral and electrophysiological study of Cantonese tone merging. *Journal of the Acoustical Society of America, 139*, 3226–3237. https://doi.org/10.1121/1.4954252

Pan, H. H. (2007). Initial strengthening of lexical tones in Taiwanese Min. In C. Gussenhoven & T. Riad (Eds.), *Experimental studies in word and sentence prosody (Volume 2)* (pp. 271–292). De Gruyter Mouton. https://doi.org/10.1515/9783110207576.2.271

Peng, S. H. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics, 25*, 371–400. https://doi.org/10.1006/jpho.1997.0047

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101–140). Mouton de Gruyter. https://doi.org/10.1515/9783110197105.1.101

Polson, N. G., & Scott, J. G. (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis, 7*, 887–902. https://doi.org/10.1214/12-BA730

Rose, P. (2004). The acoustics and probabilistic phonology of short-stopped syllable tones in Hong Kong Cantonese. In *Proceedings of The 10th Australian International Conference on Speech Science & Technology*, 445–450, Sydney.

Sarmah, P., Dihingia, L., & Lalhminghlui, W. (2015). Contextual variation of tones in Mizo. In *Proceedings of Interspeech 2015*, Dresden. https://doi.org/10.21437/Interspeech.2015-25

Silpachai, A. (2024). The boundary-induced modulation of obstruents and tones in Thai. *Journal of Phonetics, 102*, 101291. https://doi.org/10.1016/j.wocn.2023.101291

Sun, X., & Huang, T. (2015). Gradience in contextual tonal realization processes: An instrumental study of Nanjing Chinese. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using Multivariate Statistics (7th edition)*. Pearson.

Williams, D. R., Rast, P., & Bürkner, P. (2018). *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/7tbrm

Wong, P., & Chan, H. Y. (2018). Acoustic characteristics of highly distinguishable Cantonese entering and non-entering tones. *Journal of the Acoustical Society of America, 143*, 765–779. https://doi.org/10.1121/1.5021251

Wong, P. C. M. (1999). The effect of downdrift in the production and perception of Cantonese level tones. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, 2395–2397, Berkeley.

Wong, W. Y. P. (2006a). *Syllable Fusion in Hong Kong Cantonese Connected Speech.* [Doctoral dissertation, The Ohio State University].

Wong, W. Y. P., Chan, M. K. M., & Beckman, M. E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In S. A. Jun (Ed.), *Prosodic models and transcription: towards prosodic typology* (pp. 271–300). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0010

Wong, Y. W. (2006b). Contextual tonal variations and pitch targets in Cantonese. In *Proceedings of Speech Prosody 2006*, Dresden. https://doi.org/10.21437/SpeechProsody.2006-77

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61–83. https://doi.org/10.1006/jpho.1996.0034

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

Yang, C., & Xu, Y. (2019). Cross-linguistic trends in tone change: A review of tone change studies in East and Southeast Asia. *Diachronica, 36*, 417–459. https://doi.org/10.1075/dia.18002.yan

Yiu, C. Y. (2009). A preliminary study on the change of rising tones in Hong Kong Cantonese: An experimental study (in Chinese). *Language and Linguistics, 10*, 269–291.

Yu, A. C. L. (2020). The phonetics of sound change. In R. D. Janda, B. D. Joseph & B. S. Vance (Eds.), *The handbook of historical linguistics* (Vol. II, pp. 293–313). https://doi.org/10.1002/9781118732168.ch14

Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics, 5*, 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815

Yu, K., & Lam, H. W. (2014). The role of creaky voice in Cantonese tone perception. *Journal of the Acoustical Society of America, 136*, 1320–1333. https://doi.org/10.1121/1.4887462

Yu, Y. (2011). Contextual tonal variations in Vientiane Lao. *Journal of Mekong Societies, 7*, 1–16.

Zhang, J. (2019). Tone mergers in Cantonese: Evidence from Hong Kong, Macao, and Zhuhai. *Asia-Pacific Language Variation, 5*, 28–49. https://doi.org/10.1075/aplv.18007.zha