Original Article

Language *and* Speech

A New System of Cantonese Tones? Tone Perception and Production in Hong Kong South Asian Cantonese

Language *and* Speech
1–25

DOI: 10.1177/00238309211046030
journals.sagepub.com/home/las
SAGE

Alan C. L. Yu
Department of Linguistics, University of Chicago, USA

Crystal W. T. Lee, Chen Lan and Peggy P. K. Mok
Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Hong Kong

Abstract
Studies in language contact have identified many instances of linguistic variation and grammatical innovations introduced by speakers from multi-ethnic urban neighborhoods. This study focuses on the variety of Cantonese spoken by South Asian youths in Hong Kong, specifically their production and perception of Hong Kong Cantonese tones. Our findings show that the South Asian Cantonese speakers have a smaller tonal inventory than the canonical six-tone system of standard Hong Kong Cantonese and their tonal discrimination abilities are also more impoverished relative to their ethnic Chinese peers. Further analysis shows a positive correlation between tonal discrimination accuracy and tonal realization distinctness among the South Asian speakers, but not among the ethnic Chinese. These findings suggest that South Asian Cantonese speakers might have developed a distinct tone system from their ethnic Chinese peers.

Keywords
Tone, Language contact, Variation, Tone perception, Tone production

1 Introduction

Studies in language contact have identified many instances of linguistic variation and grammatical innovations introduced by speakers from multi-ethnic urban neighborhoods (e.g., Multicultural London English (Torgersen et al., 2006; Kerswill et al., 2008), Straattaal (Netherlands; Appel, 1999), Rinkeby-Svenska (Sweden; Bodén, 2004; Kotsinas, 1998), Kobenhavnsk Multietnolekt (Denmark; Quist, 2005), Hood German/Kiezdeutsch (Germany; Jannedy, 2010)). This study focuses on the variety of Cantonese spoken by South Asian (SA) youths in Hong Kong. While

Corresponding author:
Peggy P. K. Mok, Department of Linguistics and Modern Languages, Chinese University of Hong Kong, Leung Kau Kui Building, Shatin, Hong Kong.
Email: peggymok@cuhk.edu.hk

Hong Kong is a predominantly Chinese society, there is a sizable population of residents from other cultural and ethnic backgrounds (~ 6.4%; Census and Statistics Department, 2012), including more than 60,000 inhabitants in Hong Kong that are of South Asian descent, many of them are born and raised in Hong Kong. Cantonese is the dominant language of Hong Kong. According to recent census information (Census and Statistics Department, 2012), a large percentage (~ 80%) of the South Asian inhabitants is illiterate in Chinese and a similarly large percentage (~ 60%) is reported to not speak Cantonese at all. Instead, in addition to English, the South Asians in Hong Kong may speak Urdu, Punjabi, Hindi, Gujarati, Sindhi, or Nepali, among others. In recent years, with the introduction of Race Discrimination Ordinance in 2008, the general public has become more aware of the sizable locally born and raised South Asian community in Hong Kong. As a result, there is an increased frequency of media coverage of South Asians in Hong Kong. While the media coverage often focuses on sensationalizing the fact that non-Chinese would learn Cantonese, the South Asian interviewees often emphasize their association of speaking Cantonese as part of their Hong Kong identity. Few commentaries cover the nature of the Cantonese spoken, even though there is strong race-based language ideology, suggesting that this segment of the Hong Kong society does not speak Cantonese properly or at all (Fleming, 2017).[1]

This study focuses specifically on the production and perception of Hong Kong Cantonese tones by a group of South Asian secondary school students. The study of tone in a contact variety of Cantonese is particularly pertinent. Traditional media portrayals of South Asians speaking Cantonese generally feature some form of tonal deviation. Unlike segmental variation, tonal deviations are hard to represent in written commentaries. Thus more detailed investigations regarding contact-induced tonal variation and change are necessary to elucidate similarity and differences among indigenous vs. contact varieties of Cantonese, and of tone languages in general (cf. sociotonetics; Stanford, 2016). The study of tonal variation in contact situation is also important for second language studies. Many studies have investigated the perception and production of tones of a second language (L2) by learners whose first language (L1) is either tonal or non-tonal (e.g., Wang & Sereno, 2006; Hao, 2012). These studies have primarily focused on the acquisition of tone in a structured instructional environment where the participants in the study were taught to produce tones in the L2. The South Asian population in Hong Kong represents a more "organic" situation in this respect in terms of L2 acquisition, particularly with respect to tone, since little is known regarding the perception and production of L2 tones by learners who learn Cantonese via general exposure in the environment without explicit instructions in the phonetics and phonology of the tone language. Specifically, most primary schools in Hong Kong use spoken Cantonese and written standard Chinese as their medium of instruction. While some secondary schools are allowed to teach in English, access to the mainstream English-medium schools in Hong Kong is very competitive. SA students often end up in Chinese-medium government(-aided) schools, which generally lack any systematic capacity to provide Chinese as a Second Language support, or, for a minority of SA students, the so-called "designated schools" which are designed specifically for ethnic minorities and teach in English (Tsung et al., 2010). As an increasing number of these ethnic minority parents in Hong Kong are sending their children to mainstream public sector schools (Carmichael, 2009), given that the primary aims of the language education in Hong Kong are to develop biliterate competency in English and Standard Written Chinese, and trilingual competency in spoken Cantonese, Putonghua, and English, SA students often cannot rely on English as the sole medium of communication in school. This sociological backdrop suggests a situation where SA students in school must acquire Cantonese via general exposure, rather than through explicit classroom instructions geared toward teaching Cantonese as a second language. The situation is further complicated by the fact that many Hong Kong-born SA moved to their parents'/grandparents' home country for a period of time before returning to Hong Kong for schooling. This unique combination of factors suggests high potential for contact-induced language variation and change within this community of Cantonese speakers.

The SA Cantonese-speaking communities in Hong Kong are also of theoretical interest to L2 research due to its internal heterogeneity. To begin with, the SA cohort in this study comprises secondary school students from different forms (from Form 1 to Form 5, similar to Grade 6 to Grade 10 in the United States), aging from 12 to 18. The sample is roughly gender-balanced (19 males and 14 females). In addition to demographic backgrounds, such as age and gender, of the South Asian speakers in Hong Kong, this study also examines the influence of a participant's dominant language on his/her acquisition of Cantonese. As noted above, South Asians in Hong Kong often speak a variety of languages at home. The majority of the SA participants in this study speaks Urdu and/or Punjabi at home. As similarities and differences between native (L1) and second (L2) languages play an important role in speech learning, the similarities and differences between the home language(s) and Cantonese might also play a role in conditioning tonal variation in SA Cantonese. The Perceptual Assimilation Model (PAM; Best, 1994), for example, argues that, if native learners assimilate two contrastive sounds in L2 into different L1 categories (the so-called two-category assimilation), the contrast will be successfully discriminated; if the two contrastive sounds are assimilated into one single L1 category (one-category assimilation), the discrimination will be inaccurate. Non-native contrasts are thus not equally difficult for listeners to perceive and the difficulty depends on how they perceive the L2 sounds in relation to L1 categories. So and Best (2010) developed the PAM for Suprasegmental (PAM-S) model, which hypothesizes that L2 tones may be assimilated into L1 tones if L1 also has tones. Assimilation is subject to the acoustic–phonetic mapping between the two systems. If L1 is a non-tone language (e.g., English, French), L2 tones can be assimilated either to L1 phrasal/sentential intonation categories, as uncategorized intonation patterns, or as non-speech melodic units, and in any case, linguistic processing of tones may not be available.

While PAM focuses on speech perception between L1 and L2, the Speech Learning Model (SLM: Flege, 1995) connects speech perception and production in L2 phonology. SLM claims that the learner's phonetic system reorganizes in response to L2 by adding new categories or modifying old ones. SLM classifies the L2 sounds, in relation to L1, as "identical," "similar," or "new." Similar sounds often result in inaccurate pronunciation, as failure to detect the difference between an L1 sound and an L2 sound leads to "equivalence classification," i.e., a single phonetic category used to process the linked L1 and L2 sounds. Then the L2 production will be influenced by features of the L1 category. By contrast, a "new" L2 sound, due to its disparate features from L1, will not be analogized to an existing L1 sound and will cause the formation of a new category for this L2 sound. However, this new category may still be dissimilar from L1 to ensure a maximal dispersion of L1 and nearby L2 sounds. The SLM has been applied to production studies of non-native sounds in diverse L1–L2 pairs (e.g., Bohn and Flege, 1996; Chang & Yao, 2011). L2 learners often produce L2 sounds with similar L1 equivalents ("similar") differently from native speakers, while they can produce L2 sounds with no L1 equivalents ("new") more accurately.

Both PAM and SLM suggest that the hardest elements in L2 phonology for learners are not the ones that are very different from their L1, but rather those bearing similar features as their L1. To this end, it is important to note that Cantonese has six lexical tones (T): T1 [55] high-level, T2 [25] high-rising, T3 [33] mid-level, T4 [21] low-falling, T5 [23] low-rising, and T6 [22] low-level (Bauer & Benedict, 1997). An example of the canonical tonal realization of these tones can be found on the left panel of Figure 1. Recent studies have found that some younger Cantonese speakers may be undergoing tonal changes (e.g., Mok et al., 2013). Such younger speakers tend to merge T2 and T5, as well as T3 and T6, respectively, in their production but the process is still very much at its early stages. English, a Germanic language, and Urdu, an Indo-Aryan language,

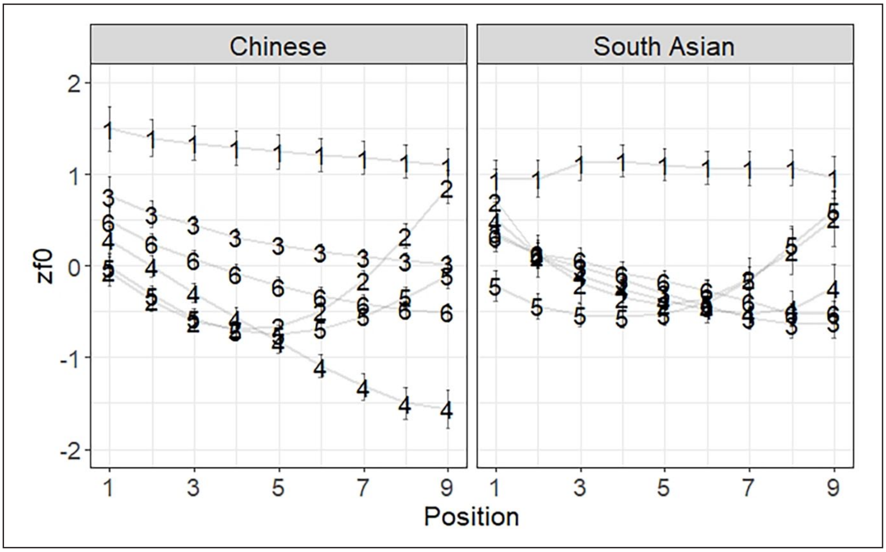

Figure 1. Average z-transformed f0 of the six Cantonese tones across nine measurement positions produced by Chinese (left) and South Asian (right) Cantonese speakers. Error bars present the 95% confidence intervals.

are both non-tonal, but have lexical stress and phrase-/sentence-level intonation. Punjabi, also an Indo-Aryan language, has been described as having three phonemically distinct tones: high-tone (rising-falling), low-tone (falling) and mid-tone (mid-level) (Gill, 1960). Punjabi possesses tonal and non-tonal words (Lata, 2011). In tonal words, there are two tones, i.e., high vs. low tones. Non-tonal words carry the mid-tone which is predicted by rules of redundancy (Singh et al., 2015). Furthermore, it has been argued that younger Punjabi speakers decreasingly rely on pitch contours to distinguish word meanings and tend to replace rising and level pitch by falling pitch, possibly due to the ease of articulation (Rafi, 2010).

We hypothesize that the complex Cantonese tone system may pose general difficulties for speakers with different self-assessed language dominance backgrounds. While Urdu does not have lexical tone, the frequent use of a rising (LH) contour in Urdu, especially on non-final content words (Hussain, 1997; Jabeen, 2010), might prompt Urdu-speaking learners to assimilate the two rising tones in Cantonese into single category. English-dominant speakers might have similar difficulties with the two rising tones given their similarity to the question intonation in English. Indeed, Kan and Schmid (2019) examined the perception of two Cantonese tone pairs, high-rising T2 vs. low-rising T5 (the so-called "similar" pairs) and the high-level T1 vs. low-falling T4 (the "dissimilar" pairs), and found that Cantonese-English bilingual children in the US generally scored lower than monolingual controls in Hong Kong, but performed particularly worse with the rising tone pairs. They attributed the poor discrimination performance with the "similar" pairs to the rising tones being mapped onto the same question intonation in English, while the "dissimilar" tone pairs were mapped onto distinct intonational categories in English (i.e., T1 to flat pitch, while T4 to the falling declarative statement intonation).

On the other hand, given that Punjabi is a tone language, Punjabi-speaking Cantonese learners may have greater difficulties over Cantonese lexical tones than Urdu-speaking learners. In particular, Punjabi-speaking Cantonese learners might find the two rising tones in Cantonese difficult to distinguish as both might be assimilated as the Punjabi high tone, which has a rising component.

They might also have difficulties distinguishing the lower register tones (T3, T4, T6) since those tones might be assimilated to the mid-tone category in Punjabi.

To be sure, experience with L1 tones does not necessarily hinder the acquisition of L2 tones. In some instances, studies have found that experience with L1 tones may in fact facilitate the acquisition of L2 tones. Francis et al. (2008), for example, found in their study of perceptual learning of Cantonese by Mandarin and English speakers that Mandarin speakers were more successful at learning some Cantonese tones than others, suggesting that influence of tonal knowledge from the L1 is not necessarily so clear cut and might ultimately be specific to the nature of the tonal mappings between languages. Mandarin listeners, for example, found the 23 (low-rising) and 25 (high-rising) tones more confusable than English listeners, while English listeners were more likely to misidentify the 23 (low-rising) tone as a 33 (mid-level) tone. Wayland and Guion (2004) likewise reported that native Mandarin listeners are better at learning Thai tones than native English listeners. By comparing the acquisition of Cantonese tones by SA learners from different dominant language backgrounds, our study should provide a fruitful avenue for investigating the effects of prior experience in L1 tones on the acquisition of L2 ones.

In a recent study of 21 L1-Urdu children and 20 L1-Cantonese children, all preschoolers between 4 and 6 years of age, Yao et al. (2019) found that L1-Urdu participants' production of T1 was on par with that of L1-Cantonese controls, but their acquisition of mid-low pitch range (T3, T4, and T6) showed low accuracy and high degree of confusion. Interpreting from PAM's perspective, they attributed the great confusion between the two rising tones, T2 and T5, to both tones being assimilated to the Urdu LH intonation category via Single Category assimilation. The confusion between T3, T4, and T6, which are all in the mid-low range, were seen as being influenced from being Uncategorized/Non-Assimilable, lack of salient acoustic distinction, and ongoing tone mergers. Our study, which focuses on SA youths rather than preschoolers, should offer useful, albeit indirect, developmental evidence, examining whether patterns observed in preschoolers extend to teenagers and young adults. This study also included SA participants whose dominant languages extend beyond Urdu, affording us the opportunity to examine the impact that the linguistic diversity within the SA community has on their Cantonese development.

Another factor that may influence language acquisition in general is individual variability that might stem from social, biological, as well as neuro-cognitive factors (Yu & Zellou, 2019). To this end, this study emphasizes the potential influence of individual variability in speech perception and production on tonal variation and change. Specifically, this study also investigates the perception–production link within speaker. Previous studies on the perception–production link have found variable results. Studies using paradigms such as altered auditory feedback (e.g., Houde & Jordon, 2002; Xu et al., 2004; Shiller et al., 2009; Katseff, 2011) and phonetic imitation (e.g., Nielsen, 2011; Babel, 2012; Yu et al., 2013) have generally found speakers to be quite adept at adjusting their production patterns in the face of adjusted perceptual feedback or altered perceptual experiences, suggesting a close link between speech perception and production. Studies in second language acquisition, however, have found that speakers' production in L2 often lags behind the individual's perceptual ability in that language. A particularly fruitful avenue to explore the perception–production link lies in the correlation between the distinctness of an individual's production of a contrast and how well the individual discriminates that contrast (Newman et al., 2001; Perkell et al., 2004a, 2004b). Perkell et al. (2004b), for example, show that while there is generally a substantial contact of the underside of the tongue tip with the lower alveolar ridge during the production of /s/ but not /ʃ/, the degree of acoustic contrast between /s/ and /ʃ/ among a gender-balanced cohort of 20 native speakers of American English is related to differences in their use of contact contrastively and in their discriminative performance. The most distinct sibilant productions were from participants who used contact in producing /s/ but not /ʃ/ and who had high discrimination scores, while the participants who did not use contact differentially to produce the sibilants would produce the least distinct sibilant contrasts and would also discriminate synthetic sibilants less well. Concerning the study of tone, while there are many studies examining the production and perception of tones in different languages, no study to date has explored how contrast distinctness might play a role in the perception and production of tone. The lacunae may partly stem from difficulties in quantifying tonal distinctness. One of the aims of this study is to advance one such way to measure tonal distinctness and to explore how tonal contrast distinctness relates to tonal discrimination.

In what follows, this study begins with a description of the cohorts studied as well as the methodological specifics. Results are presented in Section 3, demonstrating that the SA cohort exhibits a smaller tonal inventory than the Chinese cohort and the tonal discrimination abilities of the SA cohort are also more impoverished relative to the Chinese speakers. Further analysis shows a positive correlation between tonal discrimination accuracy and tonal realization distinctness among the SA participants, but not among the Chinese cohort. We argue in Section 4 that SA Cantonese exhibit a distinct tone system from the Chinese speakers and explore the potential sources of this new system.

2 Method

2.1 Participants

Forty-eight SA and thirty-three Hong Kong Chinese participants took part in both the production and perception experiments. As some of the SA participants' first language is a language other than Urdu, Punjabi, or Hong Kong Cantonese, their data were excluded in the following analysis. Participants who were not born in Hong Kong or whose first languages included Mandarin were also excluded from the analysis. While all SA participants, aged 12 to 18, are speakers of Urdu and/or Punjabi, 10 self-reported to be Punjabi-dominant, 13 Urdu-dominant, and 7 English-dominant; 3 claimed to be fluent in multiple (non-Cantonese) languages and were not included in the following analysis in order to ensure the dominant language-based subgroups within the SA cohort were as balanced as possible. None reported Cantonese as their dominant language. The Chinese participants, aged from 15 to 18, were either secondary students or first-year undergraduate students at a university in Hong Kong. In the end, the final data set included production and perceptual responses from 29 SA participants (19 males, 10 females) and 22 Chinese (CC) participants (8 males, 14 females). One Chinese female participant only took part in the production study but not in the perceptual one. All participants, who were paid a nominal fee to take part in the experiment, reported no hearing or speech problems.

2.2 Materials

Production data was elicited via a picture-naming task, which included 84 pictures. While lexical frequency was not formally controlled for, many of the items were taken from the Hong Kong Cantonese Articulation Test (HKCAT; Cheung et al., 2006), which is designed to be used with young children; additional common words were added to ensure the tonal contrasts were balanced. Each picture was accompanied by the corresponding Chinese character and English gloss to facilitate production. The tonal production stimuli were part of a larger study examining SA Cantonese. Of the 84 pictures, 12 were intended to elicit the tone contrasts. The target words were the six tonal variants of the syllables [ji] and [si]; all the resultant combinations are real monosyllabic words in Cantonese. The other pictures were designed to elicit segmental contrasts.

The perception task consisted of an AX discrimination task with 150 AA pairs and 150 AB pairs. Five syllables, namely [fa:n], [fan], [si], [jan], and [ji], each carrying six tones in Cantonese, were chosen. Therefore, there were 30 target monosyllables. They were paired up with monosyllables having the same segment and same tone to form the AA pairs, and with monosyllables with the same segment and different tones (e.g., T1/T2, T2/T5) to form the AB pairs. Since there were 15 possible tone combinations to form the AB pairs, and the order of the AB pairs was counterbalanced, there were 150 AB pairs altogether (15 tone combinations × 2 orders × 5 syllables). The AA pair of each syllable appeared five times in order to balance the number of the AB pairs (6 tone combinations × 5 syllables × 5 repetitions = 150). The stimuli were produced by a phonetically trained female native speaker of Hong Kong Cantonese of ethnic Chinese background. The two stimuli in the AA pairs were not from the same recording. All pairs were randomized in the perception task.

2.3 Procedure

With the exception of one of female Chinese participants, who did not complete the perceptual task, all subjects participated in both the production and perception tasks, and finished a language background questionnaire in one sitting. Half of the participants did the production task first, while the remaining half did the perception task first. The experiments were conducted in quiet classrooms in the respective schools (for the secondary school students) or a recording booth (for the university students). In the production experiment, all subjects were instructed to say the monosyllabic words in the pictures naturally. They were given plenty of time to pronounce the words themselves. When needed, the experimenters would provide hints if they had difficulties in recalling the pronunciation. If a student knew the word, they pronounced the target monosyllabic word by themselves at least two times (i.e., the self-attempted tokens); in the end, three repetitions were elicited for each item. If the participants did not know the word, they were prompted to repeat after the experimenters at least twice (i.e., the shadowed tokens). Only self-initiated production responses were included in the analysis and the two repetitions with the best quality were chosen for analysis. Their speech was recorded using a digital audio recorder placed approximately 20 cm away from them. The final dataset in the production analysis consists of productions from 30 SA speakers (18 females; 7 English-dominant, 10 Punjabi-dominant, and 13 Urdu-dominant) and 22 Chinese speakers. F0 were measured only on the rhyme of the target syllables. The tracking is performed by taking measures of F0 at nine equidistant points using ProsodyPro (Xu, 2013). The F0 measurements were *z*-transformed by participant prior to further analysis.

In the AX discrimination task, each participant listened to a randomized list of 300 monosyllabic word pairs and was asked to judge if the tones of the two monosyllables were the same. The stimuli were presented using E-Prime 2.0 on a laptop computer via headphones. They were instructed to use the index and the middle fingers of their dominant hand to press the keys for the "same" and "different" responses on the keyboard. Also, they were required to respond as quickly (no more than 5000 ms) and as accurately as possible. Participants were given a short practice session prior to the actual task. The task was divided into six blocks, with short breaks scheduled between blocks. Prior to each trial, a fixation point appeared on the screen. Participants were given feedback only in the practice session. The inter-stimulus interval was 500 ms. Reaction time was measured from the offset of the second monosyllable.

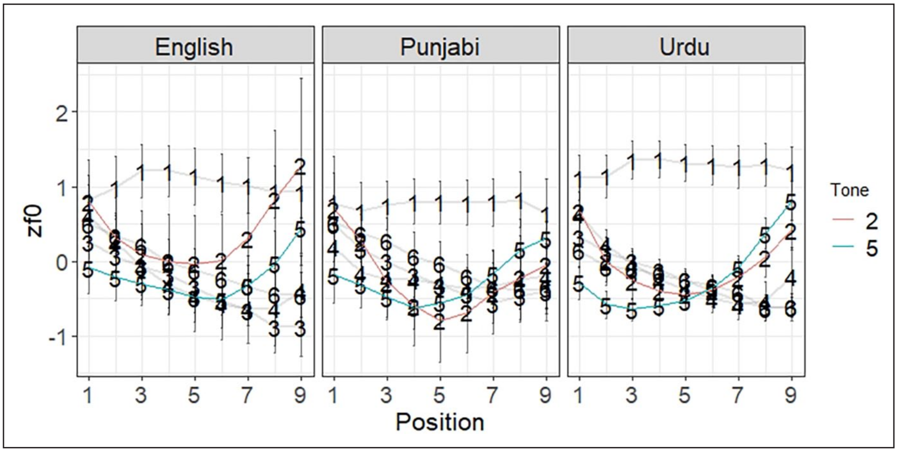

Figure 2. Average z-transformed F0 of the six Cantonese tones across nine measurement positions produced by South Asian Cantonese speakers with different dominant language backgrounds. Error bars present the 95% confidence intervals. T2 and T5 are highlighted in red and blue respectively for ease of reference.

3 Results

Figure 1 summarizes the tonal production results. Qualitatively, there are fewer tonal distinctions in the SA productions (right panel) compared with the production of the CC cohort (left panel). To begin with, the SA tone space appears to be much smaller than that of the CC cohort. That is, the range of normalized F0 used by the SA cohort is smaller than that of the CC cohort. Also, certain tonal distinctions appear to be neutralized. The contrast between T3 and T6, and possibly their contrasts with T4, appear to have been neutralized in the SA productions. The realization of the two rising tones (T2 and T5) in the SA production also differs from that of the CC cohort. The SA T2 and T5 are mainly differentiated at the tonal onset, rather than at the offset, as in the case of the CC cohort. Figure 2 subdivided the tonal productions of the SA cohort by the speakers' dominant languages. Visual inspection suggests that SA speakers who self-reported English as their dominant language have more distinct T2 and T5 than speakers whose dominant language is Punjabi or Urdu. The next section tests these observations through a series of regression analyses.

3.1 Tonal pattern analysis

In order to reduce the complexity of the tonal comparisons across cohorts, the (normalized) F0 trajectories were parameterized using discrete cosine transform (DCT) over the sonorous portion of the target syllable (Watson & Harrington, 1999; Harrington, 2010). The DCT allows the reduction of the complexity of the F0 trajectory into a triplet of coefficients that are proportional to the mean, linear slope, and curvature of F0 respectively. The F0 values were averaged by tone and position for each participant prior to the DCT analysis. Since exploratory analysis found no significant difference in DCT values between [ji] and [si] syllables, the difference in syllable onset was collapsed, resulting in a set of three DCT coefficients for each tone for each participant. The duration measurements were normalized by *z*-scoring within-speaker prior to further analysis. The normalized duration values and individual DCT coefficients were modeled in terms of linear regressions. Besides Tone (six levels), other predictors tested included AGE, GENDER, DOMINANT LANGUAGE (English, Punjabi, and Urdu), and years of Cantonese EXPOSURE. As noted in the Introduction, the linguistic background of the participants, particularly the tonal nature of their self-assessed dominant language, could influence their learning of Cantonese. A participant's exposure to Cantonese input is another potential confound that could influence language acquisition outcomes. As noted above, while all participants included in this analysis were born in Hong Kong, some of them moved to their parents' home country for a period of time before returning to Hong Kong for schooling. Thus, a participant's years of Cantonese exposure might differ from the participant's chronological age. However, because the two are highly correlated, we tested the two predictors independently, but not in the same model. DOMINANT LANGUAGE was contrast-coded in the following way: the first contrasts between the CC and SA cohorts; the second contrasts between the Punjabi-dominant participants and other SA participants (i.e., the English- and Urdu-dominant SA participants); the third contrasts between the Urdu- and English-dominant participants. All other categorical variables were sum-coded, while the continuous variables were centered and scaled. Model comparisons revealed no significant effects of GENDER, AGE, or EXPOSURE as main predictors nor in terms of their individual interaction with Tone and were therefore dropped in the final analysis. TONE, DOMINANT LANGUAGE, and their interactions were significant predictors in the models for DCT1 (F0 mean) and DCT2 (F0 slope) and duration. TONE is the only significant predictor in the DCT3 (f0 curvature) model. A summary of the regression models for all three DCT coefficients and the duration measurements is shown in Table 1. In order to examine the differences in tonal realization across language cohorts, post-hoc pairwise comparisons, using the emmeans package in R (Lenth et al., 2021), were conducted. We focused on differences in DCT coefficients and duration between tones within each language, rather than across languages, in order to discern how many tonal distinctions are supported within each language dominance subgroup. The associated *p*-value are corrected for multiple comparisons using the Tukey HSD method.

Before diving into the results of the tonal pattern analysis, it is worth recalling that, per the discussion in the Introduction, we hypothesize that the SA participants would have the most difficulty in differentiating T2 and T5, the two rising tones and the lower register tones (T3, T4, T6) due to their similarity to existing tonal categories or general lack of salient acoustic distinction.

Figure 3 shows the model predictions for all three DCT coefficients and the normalized duration measure for different tones and, with the exception of DCT3, in different dominant languages. Focusing first on the results of DCT1, which correlates with mean F0, visual inspection suggests that the CC cohort shows roughly four clusters of DCT1 values, while the SA cohorts show only two. These observations are echoed in the pairwise comparison results. Table 1 summarizes the model predicted differences between tones and their significance levels. For the CC cohort, T1 has a significantly higher DCT1 than all other tones, while T4 is the lowest. The DCT1 of T2, the high rising tone, is not distinct from those of T5 and T6, but is lower than that of T3. The DCT1 of T3 is also higher than T4 and T5, but not T6. T5 is also not distinct from T6.

The DCT1 values of the different SA cohorts pattern quite differently from those of the CC cohort, but are remarkably similar across the different SA subgroups (see Table 2). Specifically, in all SA cohorts, T1 differs from all other tones, but the non-high tones do not significantly differ from each other in terms of DCT1. The only exception to this overall pattern is found among the English-dominant SA participants; while their T1 is higher than all other tones in terms of DCT1, the T1/T2 difference in DCT1 is not significant.

DCT2 correlates negatively with slope. The higher the DCT2 value, the more negative the slope of the tonal contour. Qualitatively, Figure 3 suggests that there are roughly three or four clusters of DCT2 values among the CC cohort, and possibly also three clusters for the Urdu-dominant cohort.

Table 1. Summary of regression models for duration and DCT coefficients 1 (F0 mean), 2 (F0 slope), & 3 (F0 curvature). The intercept refers to T1 produced by the CC cohort. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

PREDICTOR	Duration (SE)	*t*-value	DCT1 (SE)	*t*-value	DCT2 (SE)	*t*-value	DCT3 (SE)	*t*-value
Intercept	–0.02 (0.04)	–0.52	0.04 (0.04)	1.15	0.16 (0.03)	6.14***	0.19 (0.01)	13.94***
Contrast $1_{CC/SA}$	0.03 (0.07)	0.38	–0.06 (0.07)	–0.83	0.11 (0.05)	2.34*		
Contrast $2_{Punjabi/OtherSA}$	–0.04 (0.10)	–0.38	–0.11 (0.10)	–1.05	0.01 (0.07)	0.13		
Contrast $3_{Urdu/English}$	0.01 (0.05)	0.23	–0.02 (0.06)	–0.37	–0.04 (0.04)	–0.93		
T1	0.51 (0.08)	6.58***	1.56 (0.08)	19.71 ***	–0.17 (0.06)	–2.96**	–0.25 (0.03)	–8.41***
T1:Contrast1	–0.64 (0.14)	–4.44***	0.36 (0.15)	2.40 *	0.12 (0.11)	1.14		
T1:Contrast2	0.00 (0.20)	–0.02	–0.39 (0.21)	–1.88	–0.10 (0.15)	–0.70		
T1:Contrast3	–0.27 (0.12)	–2.25*	0.18 (0.12)	1.44	0.00 (0.09)	–0.03		
T2	–0.39 (0.09)	–4.36***	–0.07 (0.09)	–0.74	–0.29 (0.07)	–4.44***	0.33 (0.03)	10.31***
T2:Contrast1	0.48 (0.16)	3.09**	–0.30 (0.16)	–1.87	–0.48 (0.12)	–4.18***		
T2:Contrast2	0.01 (0.25)	0.04	–0.41 (0.26)	–1.59	0.48 (0.19)	2.57*		
T2:Contrast3	0.02 (0.13)	0.13	–0.25 (0.14)	–1.87	0.14 (0.10)	1.44		
T3	–0.07 (0.08)	–0.84	–0.24 (0.08)	–2.96 **	0.28 (0.06)	4.77***	–0.11 (0.03)	–3.70***
T3:Contrast1	0.53 (0.15)	3.63***	0.89 (0.15)	5.92 ***	–0.25 (0.11)	–2.33*		
T3:Contrast2	–0.62 (0.21)	–2.88**	0.22 (0.22)	1.00	–0.08 (0.16)	–0.52		
T3:Contrast3	–0.01 (0.12)	–0.10	0.10 (0.12)	0.83	–0.03 (0.09)	–0.34		
T4	–0.33 (0.08)	–4.15***	–0.59 (0.08)	–7.34 ***	0.33 (0.06)	5.60***	–0.01 (0.03)	–0.19
T4:Contrast1	–1.10 (0.15)	–7.53***	–0.69 (0.15)	–4.58 ***	0.43 (0.11)	3.98***		
T4:Contrast2	0.67 (0.21)	3.26**	0.21 (0.21)	1.02	–0.55 (0.15)	–3.60***		
T4:Contrast3	0.10 (0.12)	0.83	0.12 (0.13)	0.92	0.05 (0.09)	0.57		
T5	0.04 (0.08)	0.53	–0.48 (0.08)	–5.88 ***	–0.39 (0.06)	–6.65***	0.17 (0.03)	5.62***
T5:Contrast1	0.27 (0.15)	1.82	–0.25 (0.15)	–1.63	0.23 (0.11)	2.16*		
T5:Contrast2	0.22 (0.21)	1.04	0.01 (0.22)	0.07	0.21 (0.16)	1.29		
T5:Contrast3	0.29 (0.12)	2.45*	0.03 (0.12)	0.23	–0.14 (0.09)	–1.61		

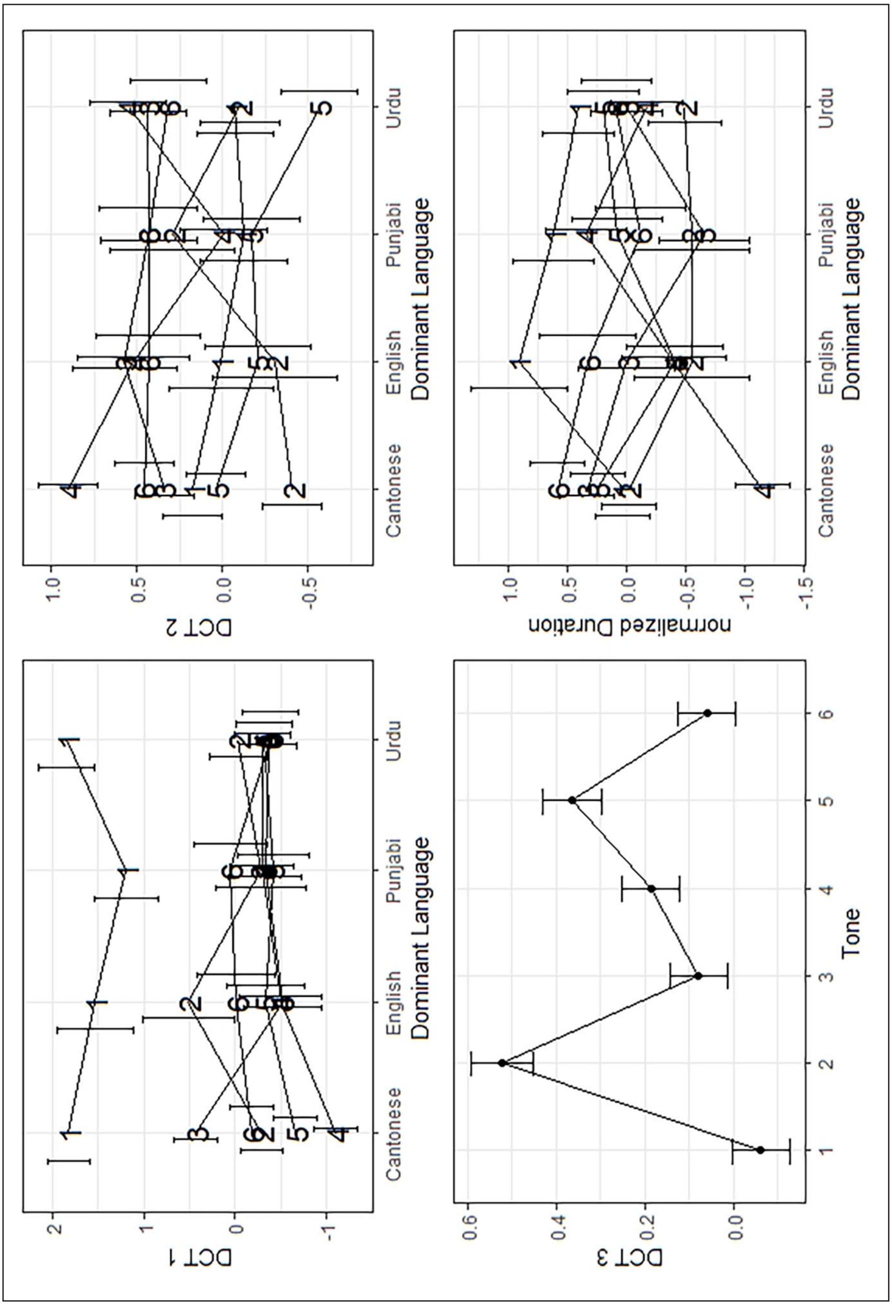

Figure 3. Model predictions of the first three DCT coefficients for the six Cantonese tones by the dominant language of the participants. Error bars indicate standard errors.

Table 2. Summary of estimated marginal mean differences in DCT1 between tones within each dominant language cohort. * $= p < 0.05$, ** $= p < 0.01$, *** $= p < 0.001$. The p-values are corrected for multiple comparisons using the Tukey HSD method.

Dominant Language	Tone	T2	T3	T4	T5	T6
Cantonese	T1	2.12***	1.4***	2.93***	2.48***	2.01***
	T2		–0.72**	0.82***	0.37	–0.1
	T3			1.54***	1.09***	0.61
	T4				–0.45	–0.92***
	T5					–0.47
English	T1	1.02	2.06***	2.04***	1.87***	1.55***
	T2		1.03	1.01	0.85	0.52
	T3			–0.02	–0.19	–0.51
	T4				–0.16	–0.49
	T5					–0.32
Punjabi	T1	1.48***	1.52***	1.49***	1.61***	1.15**
	T2		0.05	0.01	0.14	–0.33
	T3			–0.04	0.09	–0.38
	T4				0.13	–0.34
	T5					–0.47
Urdu	T1	1.88***	2.21***	2.15***	2.17***	2.23***
	T2		0.33	0.27	0.28	0.35
	T3			–0.06	–0.04	0.03
	T4				0.02	0.08
	T5					0.07

DCT2 values do not appear to be very different across tones for the English- or Punjabi-dominant participants. The pairwise comparison results echo these observations. As summarized in Table 3, for the CC cohort, T4, the low falling tone, has a higher DCT2 than all tones except F6, the low level tone. T2 has lower DCT2 except for T5, the other rising tone in the system. The level tones (T1, T3, T6) and T5, the low- to mid-rising tone, do not differ from each other in terms of DCT2. There are considerably less significant DCT2 differences among tone pairs within the SA cohort. To begin with, no significant DCT2 difference between tones was observed within the Punjabi- or English-dominant subgroup. For the Urdu participants, T5 has a significantly lower DCT2 value than T3, T4, and T6, suggesting that the slope of T5, the low-to-mid-rising tone, is more positive than the three lower register tones. T4, the low-falling tone, is higher in DCT2 than T1 and T2, indicating that T4 has a more negative slope than the higher register tones.

The DCT3 values differ across tones. As summarized in Table 4, the DCT3 value of T1 differs significantly from those of all other tones, indicating that T1 has a flatter contour than the rising (T2 and T5) and falling tones (T4). T2 is also higher in DCT3 than all other tones, including T5, suggesting that T2 is generally more concave-upward than the other tones. Both rising tones, T2 and T5, differ significantly from T3, T4, T6, the lower register tones. The latter three tones do not differ from each other in terms of DCT3.

In terms of duration, as summarized in Table 5, the CC cohort shows a significantly shorter T4 than all other tones. The SA cohort in general exhibits quite distinct duration profiles from the CC cohort. The English-dominant SA participants showed longer T1 compared to T2, T3, and T5 ($p < 0.01$). The Punjabi-dominant SA participants have longer T1 compared to T2 ($p < 0.05$)

Table 3. Summary of estimated marginal mean differences in DCT2 between tones within each dominant language cohort. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. The p-values are corrected for multiple comparisons using the Tukey HSD method.

Language	Tone	T2	T3	T4	T5	T6
Cantonese	T1	0.58**	–0.17	–0.73***	0.14	–0.29
	T2		–0.75***	–1.31***	–0.44	–0.87***
	T3			–0.56**	0.3	–0.12
	T4				0.86***	0.44
	T5					–0.42
English	T1	0.31	–0.56	–0.51	0.21	–0.42
	T2		–0.88	–0.83	–0.1	–0.74
	T3			0.05	0.78	0.14
	T4				0.72	0.09
	T5					–0.64
Punjabi	T1	–0.42	–0.56	–0.12	0.04	–0.56
	T2		–0.14	0.29	0.46	–0.14
	T3			0.43	0.6	0
	T4				0.17	–0.43
	T5					–0.6
Urdu	T1	0.03	–0.51	–0.62*	0.49	–0.39
	T2		–0.54	–0.65*	0.47	–0.42
	T3			–0.11	1***	0.12
	T4				1.12***	0.23
	T5					–0.88***

Table 4. Summary of estimated marginal mean differences in DCT3 between tones. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. The p-values are corrected for multiple comparisons using the Tukey HSD method.

Tone	T2	T3	T4	T5	T6
T1	–0.58***	–0.14*	–0.25***	–0.42***	–0.12
T2		0.44***	0.34***	0.16*	0.46***
T3			–0.11	–0.28***	0.02
T4				–0.18**	0.13
T5					0.3***

and T3 ($p < 0.001$) and a shorter T3 compared to T4 ($p < 0.05$). The Urdu-dominant SA participants have longer T1 than T2 ($p < 0.05$); the rest of the tones are otherwise similar in duration.

To summarize, the regression results of the three DCT and duration dimensions suggest that, unlike the CC cohort, the SA cohort in general does not differentiate between T2 and T5, the two rising tones, nor do they differentiate between T3, T4, and T6, the three lower register tones. These findings are consistent to the hypotheses laid out above and point to a reduced tone system among the SA cohort, relative to the CC counterpart. Among the SA participants, there is remarkably little difference in their tonal realizations across language dominance subgroups, with the only exception of the steeper slope of T5 among the Urdu-dominant participants.

Table 5. Summary of estimated marginal mean differences in duration between tones. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. The p-values are corrected for multiple comparisons using the Tukey HSD method.

Language	Tone	T2	T3	T4	T5	T6
Cantonese	T1	0.05	–0.3	1.18***	–0.21	–0.55
	T2		–0.36	1.12***	–0.27	–0.61
	T3			1.48***	0.09	–0.25
	T4				–1.39***	–1.73***
	T5					–0.34
English	T1	1.46**	0.92	1.31**	1.32**	0.59
	T2		–0.55	–0.15	–0.14	–0.88
	T3			0.4	0.41	–0.33
	T4				0.01	–0.72
	T5					–0.74
Punjabi	T1	1.17*	1.27***	0.27	0.54	0.73
	T2		0.11	–0.9	–0.63	–0.44
	T3			–1*	–0.73	–0.54
	T4				0.27	0.46
	T5					0.19
Urdu	T1	0.9*	0.41	0.57	0.21	0.32
	T2		–0.49	–0.32	–0.69	–0.58
	T3			0.17	–0.2	–0.09
	T4				–0.37	–0.26
	T5					0.11

3.2 Tonal distance analysis

To further examine how the tones of the SA cohort differ from those of the CC cohort, the tonal distance between the two cohorts was calculated. Taking the (normalized) duration and DCT coefficients of the mean F0 trajectories of each tone of the CC speakers as the centroids (the mean of each tone in a four-dimensional space), we calculated the Euclidean distance from each tone production of each SA speaker to the respective tonal centroid. For example, the first three DCT coefficients for Tone 1 for the CC speakers are 1.79, 0.16, and 0.024, respectively, and the normalized duration is 0.02. The three DCT coefficients for the same tone for SA participant #1 were 1.44, −0.74, −0.85, and the normalized duration is 0.99. The Euclidean distance between the two points in the four-dimensional space is 1.63.

The distance-from-centroid measurements were modeled in terms of linear regression. Besides Tone (six levels), other predictors tested included participants' AGE, GENDER, DOMINANT LANGUAGE (English, Punjabi, and Urdu), and years of Cantonese EXPOSURE. TONE, GENDER, DOMINANT LANGUAGE were all treatment-coded with T1, Female, and Punjabi set as the reference levels respectively. All other variables were treated as continuous variables and were *z*-scored and centered. The final model includes TONE and years of Cantonese EXPOSURE. The inclusion of the other predictors, including their two-way interactions with TONE, did not significantly improve model likelihood. Table 6 summarizes the regression model for the centroid distance analysis. The tonal distance between T1 produced by the SA cohort relative to the CC cohort is 1.22. The tonal distances from the CC centroid of T2, T3, and T5 do not differ from T1 significantly. However, the SA to CC distance for T4 is significantly larger than

Table 6. Summary of the regression model for centroid distance. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

	Coef (SE)	t value
INTERCEPT	1.22 (0.09)	13.28***
T2	0.05 (0.14)	0.38
T3	0.03 (0.13)	0.19
T4	0.48 (0.13)	3.70***
T5	–0.22 (0.13)	–1.68
T6	–0.40 (0.13)	–3.03**
EXPOSURE	–0.19 (0.04)	–4.79***

that of T1 ($\beta = 0.48, t = 3.70, p < 0.001$), while the tonal distance of T6 is significantly smaller ($\beta = -0.40, t = -3.03, p < 0.001$). Overall, with the exception of T6, the tones of the SA speakers are relatively distinct from those of the CC cohort. Years of Cantonese exposure also significantly affects the tonal distance from centroid. The more years a SA participant is exposed to Cantonese, the smaller the distance between his/her tones and the CC centroids ($\beta = -0.19$, $t = -4.79, p < 0.001$).

3.3 Discrimination accuracy analysis

In order to take account of response bias, a discrimination accuracy score (P) of each tone pair for each participant was calculated based on the following formula (Xu et al., 2006):

$$P = P(''S''\,|\,S)\cdot P(S) + P(''D''\,|\,D)\cdot P(D)$$

where $P(''S''\,|\,S)$ refers to the percentage of Same responses of all the same pairs, whereas $P(''D''\,|\,D)$ refers to the percentage of Different response among all the different pairs. $P(S)$ and $P(D)$ refer to the probabilities of the Same (AA, BB) and Different (AB, BA) trials in a unit. The score ranges from 0 to 1, indicating 0% to 100% correct discrimination respectively.

Discrimination accuracy scores for the stimulus pairs were modeled using a series of linear mixed-effects regressions fitted in R, using the lmer() function from the lme4 package (Bates et al., 2011). Besides the two main predictors, the tone Pair Type (15 levels) and DOMINANT LANGUAGE (4 levels), we also tested the inclusion of AGE, GENDER, and years of Cantonese EXPOSURE, and their interaction with two main variables. However, the inclusion of AGE, GENDER, and EXPOSURE did not significantly improve model likelihood and were therefore not included in the final analysis. The model also included by-subject random intercepts. DOMINANT LANGUAGE was contrast-coded in the following way: the first contrasts between the CC and SA cohorts; the second contrasts between the Punjabi-dominant participants and other SA participants (i.e., the English- and Urdu-dominant SA participants); the third contrasts between Urdu- and English-dominant participants. PAIR TYPE was sum-coded. AGE was centered and z-scored.

The regression model for discrimination accuracy is summarized in Table 7. Language dominance affects tonal discrimination accuracy. As illustrated in Figure 4, the CC cohort has better discrimination accuracy than every SA cohort with different dominant languages ($\beta = 0.22$, $t = 8.1, p < 0.001$). The Punjabi-dominant participants also showed a significantly lower discrimination accuracy than the other SA participants ($\beta = -0.12, t = -3.27, p < 0.01$).

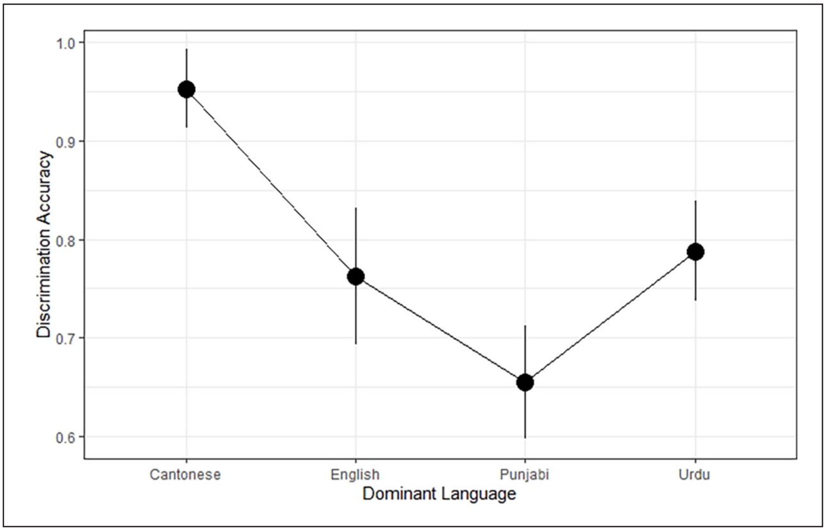

Figure 4. Model predictions of discrimination accuracy by language dominance. Error bars indicate 95% confidence intervals.

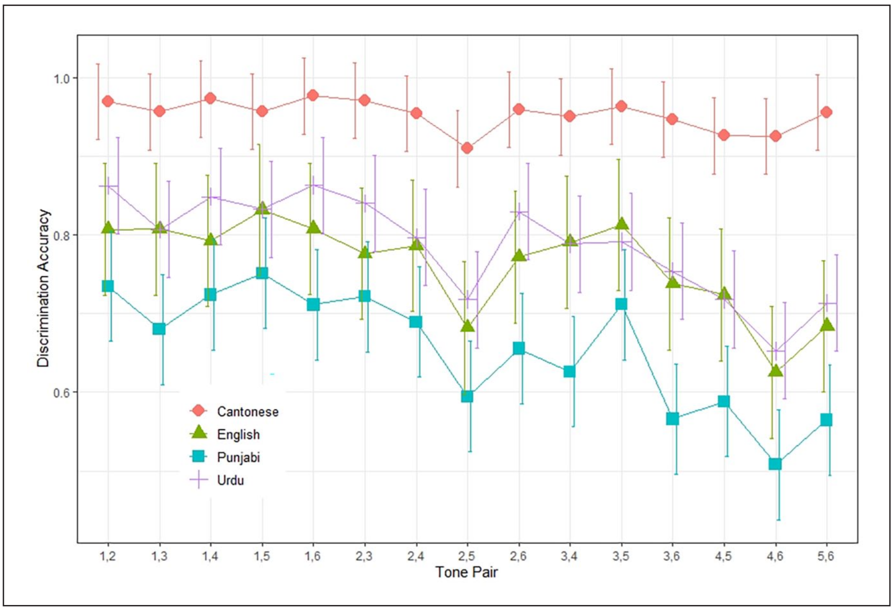

Figure 5. Model predictions of discrimination accuracy score for each tone pair between cohorts with different dominant languages. Error bars indicate 95% confidence intervals.

The model also showed significant effects of TONE PAIR. In general, discrimination is poorest for tone pairs T2/T5 ($\beta = -0.06, t = -6.62, p < 0.001$), T3/T6 ($\beta = -0.04, t = -4.01, p < 0.001$) and T4/T6 ($\beta = -0.11, t = -11.63, p < 0.001$). Crucially, there are significant interactions between TONE PAIR and DOMINANT LANGUAGE. Model predictions of discrimination accuracy score for each tone pair between cohorts with different dominant languages are illustrated in Figure 5. The

Table 7. Summary of the regression model for discrimination accuracy. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Predictor	Coef (SE)	t–value	Predictor	Coef (SE)	t–value
INTERCEPT	0.79 (0.01)	55.90***			
T1/T2	0.05 (0.01)	5.61***	T1/T2: Contrast 1	–0.05 (0.02)	–2.72**
T1/T3	0.02 (0.01)	2.40*	T1/T3: Contrast 1	–0.03 (0.02)	–1.43
T1/T4	0.04 (0.01)	4.68***	T1/T4: Contrast 1	–0.03 (0.02)	–1.82
T1/T5	0.05 (0.01)	5.58***	T1/T5: Contrast 1	–0.07 (0.02)	–3.61***
T1/T6	0.05 (0.01)	5.23***	T1/T6: Contrast 1	–0.04 (0.02)	–1.92
T2/T3	0.04 (0.01)	3.91***	T2/T3: Contrast 1	–0.03 (0.02)	–1.46
T2/T4	0.02 (0.01)	1.78	T2/T4: Contrast 1	–0.02 (0.02)	–1.13
T2/T5	–0.06 (0.01)	–6.62***	T2/T5: Contrast 1	0.03 (0.02)	1.49
T2/T6	0.01 (0.01)	1.52	T2/T6: Contrast 1	–0.01 (0.02)	–0.60
T3/T4	0.00 (0.01)	–0.08	T3/T4: Contrast 1	0.00 (0.02)	–0.15
T3/T5	0.03 (0.01)	3.14**	T3/T5: Contrast 1	–0.03 (0.02)	–1.44
T3/T6	–0.04 (0.01)	–4.01***	T3/T6: Contrast 1	0.04 (0.02)	2.34*
T4/T5	–0.05 (0.01)	–5.26***	T4/T5: Contrast 1	0.03 (0.02)	1.73
T4/T6	–0.11 (0.01)	–11.63***	T4/T6: Contrast 1	0.11 (0.02)	6.13***
Contrast $1_{CC/SA}$	0.22 (0.03)	8.10***			
Contrast $2_{Punjabi/OtherSA}$	–0.12 (0.04)	–3.27**			
Contrast $3_{Urdu/English}$	0.01 (0.02)	0.58			
T1/T2: Contrast 2	0.02 (0.02)	0.78	T1/T2: Contrast 3	0.02 (0.01)	1.02
T1/T3: Contrast 2	–0.01 (0.02)	–0.29	T1/T3: Contrast 3	–0.01 (0.01)	–0.86
T1/T4: Contrast 2	0.02 (0.02)	0.94	T1/T4: Contrast 3	0.02 (0.01)	1.04
T1/T5: Contrast 2	0.04 (0.02)	1.57	T1/T5: Contrast 3	–0.01 (0.01)	–0.83
T1/T6: Contrast 2	0.00 (0.02)	–0.17	T1/T6: Contrast 3	0.01 (0.01)	1.00
T2/T3: Contrast 2	0.03 (0.02)	1.35	T2/T3: Contrast 3	0.02 (0.01)	1.31
T2/T4: Contrast 2	0.02 (0.02)	0.72	T2/T4: Contrast 3	–0.01 (0.01)	–0.49
T2/T5: Contrast 2	0.01 (0.02)	0.58	T2/T5: Contrast 3	0.01 (0.01)	0.36
T2/T6: Contrast 2	–0.03 (0.02)	–1.02	T2/T6: Contrast 3	0.02 (0.01)	1.09
T3/T4: Contrast 2	–0.04 (0.02)	–1.73	T3/T4: Contrast 3	–0.01 (0.01)	–0.93
T3/T5: Contrast 2	0.03 (0.02)	1.18	T3/T5: Contrast 3	–0.02 (0.01)	–1.56
T3/T6: Contrast 2	–0.06 (0.02)	–2.40*	T3/T6: Contrast 3	0.00 (0.01)	–0.31
T4/T5: Contrast 2	–0.01 (0.02)	–0.51	T4/T5: Contrast 3	–0.02 (0.01)	–1.05
T4/T6: Contrast 2	–0.01 (0.02)	–0.44	T4/T6: Contrast 3	0.00 (0.01)	0.06

discrimination accuracy differences between the CC and SA participants are significantly reduced for tone pairs T1/T2 ($\beta = -0.05, t = -2.72, p < 0.01$) and T1/T5 ($\beta = -0.07, t = -3.61, p$ 0.001), but widened for T3/T6 ($\beta = 0.04, t = 2.34, p < 0.05$) and T4/T6 ($\beta = 0.11, t = 6.13, p < 0.001$). Discrimination accuracy across tone pairs is similar across the different SA subgroups, other than a significantly larger T3/T6 discrimination accuracy difference between the Punjabi-dominant SA participants and the other SA participants.

3.4 Interim summary

Overall, our findings thus far show that the SA cohort exhibits different tonal perception and production patterns relative to the CC cohort. In particular, the SA cohort appears to have neutralized the six-way tonal contrast to a three-way contrast in production. That is, the two rising tones (T2 and T5) are neutralized and so are the contrasts between the lower register tones (T3, T4, T6). The patterns of tonal merger in production mirror the tonal discrimination results. The SA cohort exhibits significantly lower degree of discrimination accuracy, with the Punjabi-dominant SA participants having the most difficulties. Overall, tonal discrimination is poorest between the rising tones (i.e., T2/T5) and between the lower register tones (i.e., between T3/T6, T4/T5, T4/T6, T5/T6). The SA cohort has particular difficulties discriminating the lower register tones. The patterns of tonal discrimination difficulties are similar across SA participants. The fact that the Punjabi-dominant SA participants having more difficulties with tonal discrimination, especially between T3 and T6 (i.e., the mid- and low-level tones), than their fellow SA participants is consistent with the predictions of PAM and SLM. That is, the fact that Punjabi has contrastive tones that share similar contour (i.e., the "rising" high tone) and pitch height (the mid-tone) might have hindered their discrimination accuracy of Cantonese tones.

The fact that discriminability differs between tone pairs also raises questions about the cause of such a difference. Previous studies suggest that certain tonal differences are intrinsically perceptually harder to discriminate than others. This interpretation is certainly supported by the fact that the tones that are harder to discriminate share similar characteristics (e.g., all low register tones, all rising tones with similar tonal onsets, etc.). Given that the SA cohorts show difficulties in producing tonal distinctions, might their poor tonal discrimination be related to their reduced tonal distinctness in production? To this end, it is worth noting that our study on tonal production also reveals significant individual variability in tonal realization. In light of the significant individual variability in tonal discrimination and tonal production, another possible venue to explore with regard to differences in tonal discriminability might come from the nature of the perception–production link. This is the topic of the next section.

3.5 The perception–production link

As noted in the Introduction, various studies have found that an individual's ability to discriminate a contrast is mirrored by his/her ability to produce the contrast distinctly. In this section, the perceptual discrimination results reviewed above are examined in concert with the participants' own production results.

Discrimination accuracy scores reviewed above were used as a measure of perceptual distinctness. In order to match the nature of the perceptual data, acoustic distance between tones within each participant's tone space was calculated. The distinctness between tones was calculated using the Euclidean distance between tones in a four-dimensional space, similar to the tonal distance analysis above. The first three dimensions were defined by the three DCT coefficients. The fourth dimension was based on the averaged difference in normalized duration between tones. That is, the acoustic duration of each token was *z*-scored within-speaker and averaged by tone. The differences between tones were calculated using the averaged *z*-scored duration for each tone for each speaker. Unlike the tonal distance calculated above, which was defined as the distance between each participant's tone and the tonal centroid defined by the production results of the CC speakers (i.e., the tonal distance calculated is between-speakers), the tonal distance here is defined within-speaker, between pairs of tones (i.e., between a speaker's own T1 and T2, or T2 and T4, etc.), analogous to the setup of the tonal pairs in the discrimination task. To examine the relationship between perceptual distinctness and acoustic tonal distance, a mixed-effects model was constructed where the dependent variable is the discrimination accuracy while the independent variables were the production-based tonal DISTANCE between tones and DOMINANT LANGUAGE. By-subject random intercepts were included, as well as by-subject random slopes for TONAL DISTANCE. TONAL DISTANCE was centered and *z*-scored, while DOMINANT LANGUAGE was treatment-coded with the CC cohort as the baseline. The model formula in lme4 format

Table 8. Summary of the linear regression model for the perception-production link. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Predictor	Coef (SE)	*t*-value
INTERCEPT	0.95 (0.02)	45.82***
TONAL DISTANCE	0.01 (0.01)	1.43
English	–0.19 (0.04)	–4.58***
Punjabi	–0.30 (0.04)	–7.74***
Urdu	–0.16 (0.03)	–4.84***
TONAL DISTANCE:English	0.01 (0.01)	0.74
TONAL DISTANCE:Punjabi	0.03 (0.01)	3.10**
TONAL DISTANCE:Urdu	0.02 (0.01)	2.74**

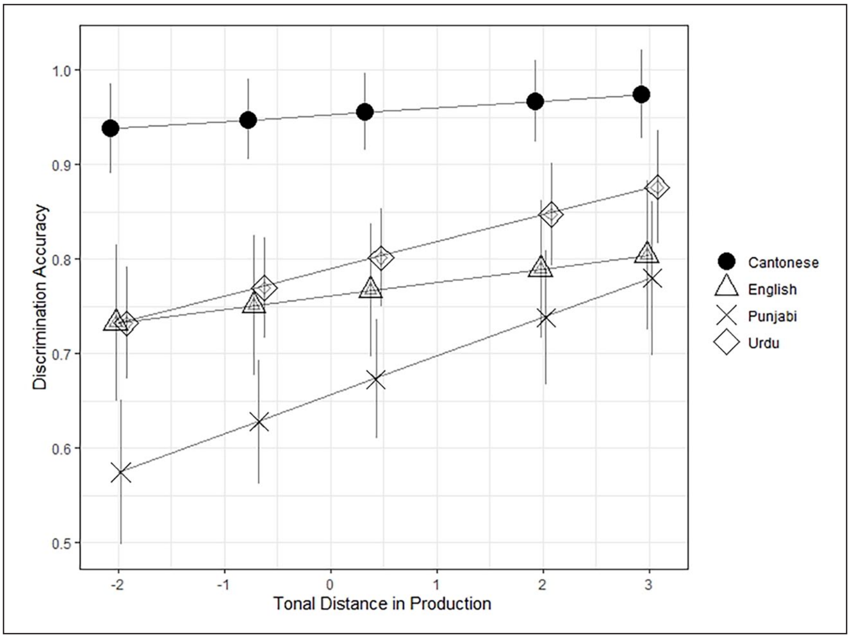

Figure 6. Model predictions of the relationships between perceptual distinctness, as measured by the discrimination accuracy score for each tone pair, and acoustic tonal distance, as measured by the averaged Euclidean distance between tones in a tone space defined by three DCT coefficients and normalized duration, observed between the cohorts with different dominant languages. Error bars indicate 95% confidence intervals.

is Discrimination accuracy ~ TONAL DISTANCE * DOMINANT LANGUAGE + (1+ TONAL DISTANCE | Subject).

The perception–production regression model is summarized in Table 8. Recall that the intercept refers to the mean discrimination accuracy of the CC cohort. TONAL DISTANCE is not a significant predictor, suggesting that there is not a significant relation between discrimination accuracy and acoustic tonal distance within the CC cohort. Similar to the results above, the different SA cohorts exhibit significantly lower discrimination accuracy relative to the CC cohort. Of particular interest are the significant interactions between SA participants from different dominant language backgrounds and TONAL DISTANCE. Figure 6 shows the relationship between perceptual distinctness and acoustic tonal distance as modulated by the participants' dominant language. With the exception of the English-dominant SA participants, the rest of the SA cohort exhibits significantly more positive correlations between discrimination accuracy and acoustic tonal distance relative to the CC cohort, suggesting that the greater the acoustic distance between two tones, the better the discrimination accuracy.

4 Discussion and conclusion

This study investigated the tonal patterns among a group of South Asian Cantonese youths in Hong Kong from both production and perceptual perspectives. Relative to the CC cohort, the SA cohort exhibits a reduced tonal inventory and a lowered level of overall tonal discrimination accuracy. The SA cohort has difficulties in producing and perceiving contrasts between the two rising tones (T2 vs. T5) and between lower register tones (i.e., T3, T4, T6). These findings are consistent with the production data from L1 Urdu preschoolers reported in Yao et al. (2019). With respect to L2 tonal acquisition, our findings show that, while SA participants have general difficulties with tonal perception and production, the Punjabi-dominant speakers have the most difficulties with tonal discrimination, particularly between T3 and T6, the lower register level tones. This pattern is consistent with the idea that the tonal system of Punjabi might have influenced the acquisition of tones in Cantonese. Specifically, Punjabi-dominant SA learners might have assimilated the two rising tones in Cantonese to the Punjabi high tone, which has a rising component, while the lower register tones to the Punjabi mid-tone. As noted earlier, Yao et al. (2019) attribute the production difficulties among L1 Urdu preschoolers with the two rising tones in Cantonese to the assimilation of these tones to the Urdu LH intonation category, while the confusion between the lower register tones to influence from being Uncategorized/Non-Assimilable (from PAM-S's perspective), a lack of salient acoustic distinction, and ongoing tone mergers in Cantonese. Our findings suggest that similarity between the L1 intonation system (e.g., Urdu, English) and the L2 tone system might not exert as strong an effect on L2 tonal acquisition as the similarity between the L1 tone (e.g., Punjabi) and the L2 tone systems.

Our findings are consistent with the idea that the SA participants have a tonal system with fewer tonal categories than those observed among the CC speakers, suggesting that a new sub-variety of Hong Kong Cantonese might be emerging where the tone system is simplified. To be sure, it is as yet unclear what the source of this new system is. There are several potential candidates. To begin with, it is worth noting that, despite the internal linguistic diversity of the SA community, the nature of the tone system reduction is remarkably consistent across SA subgroups. That is, they all converge to merging the two rising tones and the lower register tones. This consistency might stem from a general lack of salient acoustic distinction between the rising tones and between the lower register tones in Cantonese (Fok, 1974; Gandour, 1981); even the CC cohort exhibits some difficulties with these tones in the discrimination task. While this general psychophysical explanation for tonal merger might explain why certain tones are likely to merge, it remains unclear if the SA Cantonese speakers innovated this reduced tone system *per se*. Our SA participants might have acquired a variety of Cantonese with a tonal inventory already reduced. To this end, it is worth noting that, with the exception of one SA participant who reported her mother as a native speaker of both English and Cantonese, none of the parents of the other SA participants speak Cantonese as their dominant language, suggesting that the variety of Cantonese they are exposed to might come from their parents or caretakers, who are themselves L2 speakers of Cantonese with varying levels of competence. It should be noted that, as one reviewer pointed out, the tonal inventory of the SA participants might reflect one typical of late L2 Cantonese learners and a more "native-like" tonal inventory may be achieved given enough exposure. That is, the tone system of the SA participants may eventually be indistinguishable from the CC cohort's system given more Cantonese exposure. On the other hand, if the reduced tone system is the result of an emergence of a distinct variety of Cantonese, then the reduced tone system observed in our SA cohort should remain stable over time. The "late L2 Cantonese learner" interpretation is partly supported by the fact that years of Cantonese exposure has a positive effect on discrimination performance. However, the lack of an analogous Cantonese exposure effect on tonal production complicates this late L2 learner interpretation. Future investigation comparing the SA's tonal inventory to those of CC and late L2 Cantonese learners who are not of SA background could help shed more light on whether a genuine new subvariety of Cantonese tone has emerged that is not attributable to imperfect L2 learning.

The fact that the SA cohort exhibits instances of tonal mergers also brings to mind the above-mentioned ongoing tonal mergers that are happening in Cantonese among ethnic Chinese speakers (e.g., Mok et al., 2013). In particular, the two rising tones T2/T5, the two level tones (T3/T6), and the low-falling and low-level tones (T4/T6) are all reported to be undergoing mergers, although the changes appear to be in the early stage of development. The parallelism between the tone mergers found in the SA Cantonese speakers and tonal mergers observed among some Chinese Cantonese speakers begs the question: Are the two groups undergoing the same sound change in progress or is this the result of parallel, but independent, development? As Mok et al. (2013) noted, the ongoing tone mergers in Hong Kong Cantonese is still at the beginning stages and there are still six tone categories in production even among young speakers who exhibit signs of mergers between T2 and T5, between T3 and T6, and between T4 and T6. The tonal mergers happening in SA Cantonese are also different in that the SA cohort has a merger between T3 and T4, which is not observed in previous reports of tone mergers in Cantonese. Moreover, the fact that tonal merger is more pervasive among the SA participants and the tonal mergers observed among the SA cohort are much more complete might reflect the fact that tone mergers are more advanced among the SA population than among the Chinese Cantonese speakers in Hong Kong.

As alluded earlier, the tone pairs that are undergoing mergers were difficult for both CC and SA cohorts, even though only the SA cohorts exhibit robust tonal mergers in production. This state of affairs further points to parallel development as an explanation. The exact mechanisms behind these parallel developments are not clear, however. To the extent that tonal mergers among the ethnic Chinese Cantonese speakers is an endogenous process (i.e., not the results of language contact, say, between recent immigrants from China who speak other varieties of Chinese), tonal mergers might stem from individual variability in general neural mechanisms in tonal encoding (e.g., Ou & Law, 2016). Tonal mergers among CC speakers might also have resulted from contact between different varieties of Cantonese or Chinese languages. The contact-induced tone mergers among non-native speakers of Cantonese who are ethnic Chinese are likely to be different from those experienced by the SA speakers, however. To begin with, the similarities and differences between the tone systems of Cantonese and other Chinese varieties are not likely to be parallel to those between Cantonese and the languages spoken by speakers in the SA community. Moreover, there might also be differences in patterns of socialization. The close-knit nature of the SA communities, as well as (if not because of) the religious, cultural, and linguistic differences between the SA communities and the Chinese population (Erni & Leung, 2014), might encourage the more rapid development of an innovative tonal system among SA Cantonese speakers, while the ethnic Chinese non-native speakers of Cantonese might be in closer contact with other ethnic Chinese who are native speakers of Cantonese, resulting in a weaker momentum in the development of a novel variety. More community-based sociophonetic research is needed to ascertain the scenarios laid out here, however.

Our results also reveal significant correlation between perceptual discrimination accuracy and tonal production distinctness among the SA participants, suggesting speakers who produced more distinct tones are more accurate in discriminating them. It is not clear if individuals who exhibit more distinct tonal production might exhibit a less reduced tonal inventory. A quick examination of the perception–production correlation did not reveal any significant tone-pair-specific correlation between perceptual discrimination and tonal production distinctness. Future investigations with an expanded cohort might clarify this possibility.

The fact that the perception–production link is observed within the SA cohort (with the notable exception of the English-dominant SA participants) but not among the CC cohort is intriguing. As reviewed in the Introduction, past studies have found that perceptual distinctness is correlated to distinctness in production, even among native speakers of the language. Our findings suggest that, while the CC cohort exhibits a wide range of tonal distinctness profile acoustically, they are nonetheless highly accurate at tonal discrimination. This dissociation between perceptual distinctness and tonal distinctness in production might stem from the fact that the AX discrimination task is not a sufficiently sensitive task to discern individual variation in perceptual discrimination among the native speakers (although it is sufficient for detecting variation within the SA cohort). The task sensitivity for establishing a perception–production link is not unprecedented. Zellou (2017), for example, found a significant relationship between individuals' production of anticipatory nasal coarticulation on vowels and their patterns of perceptual compensation in a discrimination task, but the speakers' responses in a second task, making explicit judgments about nasality in context, did not correlate with their productions. To this end, it is worth nothing that Ou and Law (2016), for example, examined the processing and production of Cantonese tones, found that perceptual acuity as reflected in event-related potentials is associated with rise time of sound amplitude envelope. Moreover, individual differences in efficiency of tone discrimination in response latency and magnitude of neural responses to rise time were correlated with acoustic measures of F0 offset and rise time differences in productions of the two rising tones. Further research focusing on the neurophysiological response difference between SA and CC participants might reveal additional information regarding the varieties of Cantonese spoken by these subcommunities in Hong Kong.

In sum, this study provided detailed perceptual and acoustic evidence for a reduced tone system among the SA participants with different language dominance profile, relative to their CC counterparts. Further sociolinguistic research is needed to ascertain whether a new variety of Cantonese is emerging among the SA community.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We are grateful to the students and staff at Delia Memorial School (Glee path) and FDBWA Szeto Ho Secondary School for their participation in this study and the Standing Committee on Language Education and Research (SCOLAR) Research and Development Projects 2015–16 (Project Ref. No. LF3-EL/164/12) for funding this project. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission.

ORCID iDs

Alan C. L. Yu https://orcid.org/0000-0002-1953-5502

Peggy P. K. Mok https://orcid.org/0000-0002-9284-6083

Note

1. In government office discourse, ethnic minorities are simply referred as Non-Chinese Speaking.

References

Appel, R. (1999). Straattaal. De mengtaal van jongeren in Amsterdam [Strattaal. The mixed language of adolescents in Amsterdam]. *Toegepaste Taalwetenschap in Artikelen*, *62*(2):39–55.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4*, R package version 0.999375-38.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Trends in Linguistics: Studies and Monographs 102. Mouton de Gruyter.

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Bodén, P. (2004). A new variety of Swedish? In *Proceedings of the 10th Australian International Conference on Speech Science & Technology* (pp. 475–480).

Bohn, O. S., & Flege, J. E. (1996). Perception and production of a new vowel category by adult second language learners. In *Second-Language Speech: Structure and Process* (pp. 53–73). Mouton de Gruyter.

Carmichael, S. (2009). *Language rights in education: A study of Hong Kong's linguistic minorities*, volume 19 of *Occasional Paper*. Centre for Comparative and Public Law, the University of Hong Kong.

Census and Statistics Department (2012). *2011 population census. Summary results*. http://www.census2011.gov.hk/pdf/summaryresults.pdf

Chang, C. B., & Yao, Y. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *Journal of the Acoustical Society of America*, *129*(6), 3964–3980.

Cheung, P., Ng, A., & To, C. K. S. (2006). *Hong Kong Cantonese Articulation Test*. City University of Hong Kong.

Erni, J. N., & Leung, L. Y.-M. (2014). *Understanding South Asian Minorities in Hong Kong*. Hong Kong University Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issue in Cross-Language Research* (pp. 233–277). New York Press.

Fleming, K. (2017). Hong Kong's language ecology and the racialized linguistic order. *Language Ecology*, *1*(1), 25–43.

Fok, C. Y-Y. (1974). A perceptual study of tones in Cantonese. *Centre of Asian Studies Occasional Papers and Monographs* (p. 18). Centre of Asian Studies, University of Hong Kong.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*, 268–294.

Gandour, J. (1981). Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics*, *9*(1), 20–36.

Gill, S. (1960). Punjabi tonemics. *Anthropological Linguistics*, *2*(6), 11–18.

Hao, Y. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*, 269–279.

Harrington, J. (2010). *Phonetic analysis of speech corpora*. Wiley-Blackwell.

Houde, J. F., & Jordon, M. I. (2002). Sensorimotor adaptation of speech i: Compensation and adaptation. *Journal of Speech, Language and Hearing Research*, *45*, 295–310.

Hussain, S. (1997). *Phonetic correlates of lexical stress in Urdu* [PhD thesis, Northwestern University].

Jabeen, F. (2010). *An acoustic study of the influence of Urdu on the intonation patterns of English in Pakistan* [MA thesis, Government College University Faisalabad].

Jannedy, S. (2010). The usage and distribution of so in spontaneous Berlin Kiezdeutsch. *ZAS Papers in Linguistics*, *52*, 43–61.

Kan, R. T. Y., & Schmid, M. S. (2019). Development of tonal discrimination in young heritage speakers of Cantonese. *Journal of Phonetics*, *73*, 240–254.

Katseff, S. (2011). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308.

Kerswill, P., Torgersen, E. N., & Fox, S. (2008). Reversing "drift": Innovation and diffusion in the London diphthong system. *Language Variation and Change*, *20*, 451–491.

Kotsinas, U.-B. (1998). Language contact in rinkeby, an immigrant suburb. In J. Androutsopoulos & A. Scholz (Eds.), *Jugendsprache* (pp. 125–148). Lang.

Lata, S. (2011). *Challenges for design of pronunciation lexicon specification (PLS) for Punjabi language*. Department of Information Technology, Govt of India.

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *emmeans*, R package version 1.*6*.1.

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.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America*, *109*, 1181–1196.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142.

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*(6), 3226. https://doi.org/10.1121/1.4954252

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, *116*(4), 2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., & Guenther, F. H. (2004b). The distinctness of speakers' /s/–/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language and Hearing Research*, *47*, 1259–1269.

Quist, P. (2005). New speech varieties among immigrant youth in Copenhagen – a case study. In V. Hinnenkamp & K. Meng (Eds.), *Sprachgrenzen überspringen. Sprachliche Hybridität und polykulturelles Selbstverständnis* (pp. 145–161). Tübingen.

Rafi, M. S. (2010). Semantic variations of Punjabi toneme. *Language in India*, *10*, 56–65.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America*, *125*(2), 1103–1113.

Singh, A., Pandey, D., & Agrawal, S. S. (2015). Analysis of punjabi tonemes. In *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference, Mar 11–13, New Delhi, India* (pp. 1694–1697).

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273-293.

Stanford, J. (2016). Sociotonetics using connected speech: A study of Sui tone variation in free-speech style. *Asia-Pacific Language Variation*, *2*(1), 48–81.

Torgersen, E., Kerswill, P., & Fox, S. (2006). Ethnicity as a source of changes in the London vowel system. In F. Hinskens (Ed.), *Language variation- European perspectives. Selected papers from the Third International Conference on Language Variation in Europe (ICLaVE3)* (pp. 249–263). Benjamins.

Tsung, L., Zhang, Q., & Cruickshank, K. (2010). Access to majority language and educational outcomes: South Asian background students in postcolonial Hong Kong. *Diaspora, Indigenous, and Minority Education*, *4*(1), 17–32. doi: 10.1080/15595690903442256

Wang, Y., & Sereno, J. A. (2006). L2 acquisition and processing of Mandarin Chinese tones. In P. Li, L. Bates, & O. Tzeng (Eds.), *Handbook of Chinese Psycholinguistics* (pp. 250–257). Cambridge University Press.

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, *106*, 458–468.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712

Xu, Y. (2013). ProsodyPro a tool for large-scale systematic prosody analysis. In B. Bigi & D. Hirst (Eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7–10). Laboratoire Parole et Langage.

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of the Acoustical Society of America*, *116*, 1168–1178.

Xu, Y., Gandour, J. T. Larson, & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*, 1063–1074.

Yao, Y., Chan, A., Fung, R., Wu, W. L., Leung, N., Lee, S., & Luo, J. (2019). Cantonese tone production in pre-school Urdu-Cantonese bilingual minority children. *International Journal of Bilingualism*, 1–19. doi: 10.1177/1367006919884659

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *PLOS ONE*, *8*(9), e74746.

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

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*(1):13–29.