

The Acquisition of Cantonese Vowel Length Contrast and Vowel Rounding Contrast by South Asian Students in Hong Kong

Peggy Mok¹, Chun Wai Leung¹, Chen Lan¹, Alan C. L. Yu²

¹The Chinese University of Hong Kong, Hong Kong

²University of Chicago, United States

peggymok@cuhk.edu.hk, chunwai.leung@cuhk.edu.hk, lchsapphire@gmail.com, aclyu@uchicago.edu

ABSTRACT

This study investigated the production and perception of the Cantonese vowel length contrast and vowel rounding contrast by South Asian (SA) students in Hong Kong. Twenty-six native Hong Kong Cantonese speakers and fifty-four SA speakers whose dominant language was Punjabi, Urdu, or English participated in an AX discrimination task and a picture naming task. Results suggested that SA participants had the ability to distinguish both the vowel length and vowel rounding contrasts in Cantonese. However, relative to the performance of Cantonese speakers, SA students produced Cantonese rounded vowels with less lip rounding. In addition, smaller differences were found between the long and short vowels in terms of both vowel quantity and quality. Furthermore, difficulty of a particular contrast varies according to phonetic environments. These findings provided an acquisition pattern of vowel length and vowel rounding contrasts among SA students.

Keywords: Cantonese, South Asians, language acquisition

1. INTRODUCTION

According to the 2016 by-census [5], there are 84,875 of ethnic minorities in Hong Kong who are of South Asian descent speaking wildly distinct languages, such as Urdu, Punjabi, Hindi, Gujarati, Sindhi, or Nepali, among others. Among the South Asian residents in Hong Kong who are aged 5 and above, only 40.5% of them claim to be able to use Cantonese as the usual or another spoken language. While an increasing number of these ethnic minority parents in Hong Kong are sending their children to mainstream public sector schools [4], these students often cannot rely on English as the sole medium of communication in school because of the trilingualism and biliteracy policy in Hong Kong. Moreover, with the lack of a Chinese-as-a-second-language curriculum to these non-Chinese speaking (NCS) students, they have to learn Cantonese in an “organic” way, i.e., via general exposure in the environment without explicit instructions in the grammatical and sound systems of

the language. Little is known with respect to their acquisition of Cantonese vowel system in such learning environment. Also, in recent years, even though there is increasing research on South Asian students in Hong Kong, none of them focuses on the production and perception of the vowel system in Cantonese. This study aimed to investigate the acquisition of the spoken Cantonese vowel system by school-age South Asians in Hong Kong.

1.1. Vowel system of Cantonese, Punjabi, and Urdu

Cantonese has thirteen vowel phones, which comprise seven long vowels [i: y: u: ε: œ: ɔ: a:] and six short vowels [ɪ ʊ e ø o ɐ] [6, 15, 27, 28]. Even though some linguists have analysed the vowels as three long high vowels and four pairs of mid- and low vowels contrasting in length [1, 7, 15, 17], only the /a:/ vs /ɐ/ pair contrasts in all environments (in diphthongs, before all coda stops and coda nasals), and it was thus with a high functional load [6, 27]. In such analysis, vowel quality plays an essential role in terms of Cantonese vowel contrasts, and duration differences are not recognized as a distinctive feature [25].

Punjabi has ten vowel phonemes [i u e ɪ ʊ o ə/ʌ ε ɔ a] [3, 11, 20], and its vowel quantity is considered phonetically less significant than vowel quality. Also, the ‘short’ vowels are better analysed as ‘centralised’ as opposed to ‘peripheral’. Vowel length is phonemic in Urdu, where at least /i (ɪ) u (ʊ) a (ə/ʌ)/ and /i: u: a: (ɑ:)/ are contrastive [14, 16, 19, 21]. The Cantonese vowel length contrast is hence predicted to be less foreign to Urdu speakers than to Punjabi speakers.

Regarding vowel roundedness, Cantonese has two pairs of vowels differing in lip roundedness: /i/ vs. /y/ (high front vowels) and /ɛ/ vs. /œ/ (mid front vowels). Neither Urdu nor Punjabi contrasts lip rounding for front vowels. Both Urdu and Punjabi have /i/ and /ɛ/ in their phonemic inventories, but the rounded counterparts of these two vowels (/y/ and /œ/) are absent. It would thus be interesting to see how this contrast is perceived and produced by SA speakers, and also to see if vowel height has any effect on how the rounding contrast for vowels are realized and perceived.

The similarities and differences between native language (L1) and the target language (L2) play an important role in speech learning. Theories on speech acquisition such as Perceptual Assimilation Model (PAM) [2] and Speech Learning Model (SLM) [9] 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.

The PAM intends to see whether there is category assimilation between native and non-native sounds. This theory discussed that if naïve learners assimilate two contrastive sounds in L2 into different L1 categories, the contrast will be successfully discriminated. On the other hand, if the two contrastive sounds are assimilated into one single L1 category, the discrimination will be inaccurate. Non-native contrasts are not equally difficult for listeners to perceive.

The SLM connects speech perception and speech production in L2 phonology. This model argues that the more an L2 sound is perceived differently from the nearest L1 sound, the more likely a new category will be developed, making it easier for the learners to perceive and produce the sound accurately.

Based on these models, we predicted that the difficulty of Cantonese vowel length contrast may differ depending on the phonetic environments, as this contrast is absent in diphthongs in both Punjabi and Urdu. In addition, differences can be expected for Urdu and Punjabi speakers because of the presence vs. partial-presence of vowel length contrast in their L1s. For the rounding contrast, although either language does not have this contrast, since both languages have the unrounded vowels /i/ and /ɛ/, they may be able to perceive the differences between the unrounded and rounded counterparts, but have difficulties in production.

2. METHODS

2.1. Participants

54 SA participants (29 males, 25 females) and 26 Cantonese participants (11 males, 15 females) participated in this study. Among the 54 SA participants, aged between 12 and 18, 15 participants were dominant in Punjabi, 18 in Urdu, 11 in English and 12 were dominant in more than one language (e.g. Punjabi and English). The Cantonese participants, aged between 15 and 18, were either Form 5 students in a government-aided secondary school in Hong Kong, or first-year undergraduate students at a university in Hong Kong. All participants reported having no hearing or speech problems.

2.2. Materials

The production experiment was a picture-naming task which intended to elicit a set of Cantonese monosyllabic real words. 78 pictures, each accompanied by its name in Chinese character(s) and the English gloss (aiming to facilitate production), were used. 14 of the 78 pictures were chosen to investigate the participants' ability to produce vowel length contrast (/a:/ vs. /ə/), and 22 were intended to examine the vowel roundedness contrast (/i/ vs. /y/ and /ɛ/ vs. /œ/).

The perception experiment made use of an AX discrimination paradigm, with all stimuli produced by a phonetically trained female Cantonese researcher. There were 100 AA (same) pairs and 100 (different) AB pairs of monosyllabic words. Each AA pair was composed of two monosyllables which were identical but came from two different recordings (e.g. [sa:n]₁ vs [sa:n]₂). The AB pairs consisted of two monosyllables which were tonally identical but differed in one segment (i.e., a minimal pair for the targeted contrast; e.g. [jɛŋ] vs [jœŋ]). In total, 20 trials featured vowel length contrast and 40 featured vowel roundedness contrast. Filler items concerning consonantal contrasts were included.

2.3. Procedure

All participants were paid to attend a perception experiment and a production experiment, as well as to fill in a language background questionnaire. Half of the participants did the production experiment first, while the other half did the perception one first. The experiment sessions involving secondary school students were conducted in a quiet classroom in their respective schools while those involving university students were conducted in a sound-proof booth.

For the production experiment, a digital audio recorder was placed approximately 20 cm away from the participant to record their speech. The pictures were shown one by one in a randomised order to each participant. They were asked to naturally name the item in each picture in isolation twice, with the monosyllabic character. No time restriction was imposed. If a participant was able to name the picture, they produced the target monosyllabic word by themselves (attempted). Hints were provided by the experimenter when needed. If a participant really could not produce the item, the participant would repeat the target monosyllabic word at least twice after the experimenter (shadowed). The two clearest repetitions of each item were chosen for analysis.

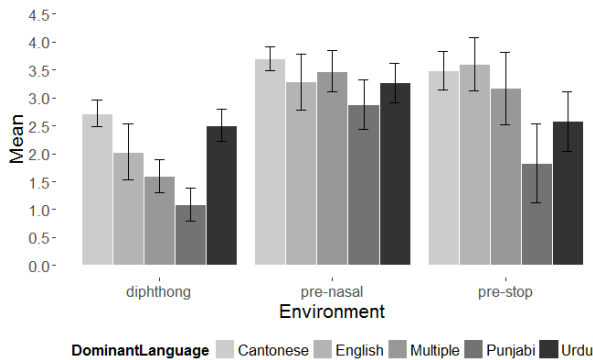
For the perception experiment, a laptop with E-Prime 2.0 was used to present the AX pairs of monosyllabic stimuli with the inter-stimulus interval being 500 ms. A fixation point appeared on the screen

prior to the playing of each pair, and the point disappeared when the sounds were being played. The participants listened to the word pairs in a randomized order, and were then instructed to use the index and middle fingers of their dominant hand to give response by pressing the “same” and “different” keys labelled on the keyboard according to the contrast. They were reminded to respond as quickly and accurately as possible within 5000 ms. Ten practice trials were given prior to the actual experiment, and feedback was given in the practice session as well.

3. RESULTS

The acoustic analyses were done in Praat using FormantPro [26]. The formant frequencies of the relevant tokens were measured at 10 equidistant points along the vowel interval. The raw data in Hertz were manually checked and any anomaly was corrected. They were then converted to Bark units using Traunmüller’s [24] formula. For the perception analyses, the accuracy on the vowel length and vowel rounding contrast was calculated, and the d-prime (d') scores, which take bias into account [18], were computed. Separate linear mixed effects analyses were conducted on the d'-scores using the lmerTest package in R (R Core Team, 2016).

Figure 1: The d'-scores on vowel length contrast of Cantonese and SA participants in three phonetic environments.



3.1. Vowel length contrast

Fig. 1 shows the mean accuracy in d'-scores of Cantonese and SA participants in three phonetic environments. The linear mixed effect model (details not given here due to page limit) reveals that the phonetic environments had a significant effect on the participants’ discrimination performance. The SA participants performed generally worse than Cantonese participants, and significantly so in diphthong environment ($F(1,76) = 7.195, p = .009$). Among the language dominance groups, the Punjabi-dominant participants performed significantly worse

than the Cantonese participants ($\beta = -1.32, t = -2.66, p < 0.01$), particularly in the diphthong environment.

The production analyses of Cantonese vowel length contrast involve both vowel quantity and vowel quality. Table 1 summarizes the differences between the mean value of [a:] and [ɐ] produced by the Cantonese and the SA participants in different environments. There was a larger durational difference between the Cantonese [a:] and [ɐ] than between the SA (attempted) [a:] and [ɐ]. Moreover, results show that the duration of [a:] and [ɐ] produced by the Cantonese participants were significantly different in all phonologically contrastive environments, namely in diphthong ($F(1,205) = 18.149, p < .001$), pre-nasal ($F(1,311) = 546.681, p < .001$), and pre-stop ($F(1,206) = 151.296, p < .001$). However, the durational difference produced by the SA participants, both attempted and shadowed, were significantly different in the pre-nasal ($p < .001$) and pre-stop ($p < .001$) environments only, but were not significantly different in the diphthong environment.

With respect to vowel quality, the formant data were analysed using Gu’s [11] Smoothing Spline ANOVA (SSANOVA). Simply put, any non-overlapping area represents significant difference between the formant measurements. Because of the page limit, only one figure of typical F1 and F2 trajectories was displayed here (Fig. 2). Data for the vowels produced in the diphthong [a:i ɐi] environments and in pre-stops and pre-nasals were not shown in Figure 2. For Cantonese participants, the formant patterns of [a:] vs. [ɐ] in diphthongs were quite different in both F1 and F2. The SA participants had similarly different patterns for attempted [a:i ɐi], and to a lesser extent, shadowed [a:i ɐi]. Nevertheless, the large Cantonese differences in F1 and F2 for [a:u ɐu] were not seen in the SA production. For pre-stop and pre-nasal environments, the SA participants could produce the difference in F1 between [a:] and [ɐ], but not in F2.

Table 1: The durational differences (in ms) between the mean values of [a:] and [ɐ] by Cantonese and SA speakers (attempted and shadowed).

| | Diphthong | Pre-nasal | Pre-stop |
|----------------|-----------|-----------|----------|
| Cantonese | 0.04 | 0.16 | 0.09 |
| SA (attempted) | 0.01 | 0.11 | 0.05 |
| SA (shadowed) | 0.05 | 0.16 | 0.07 |

3.2. Vowel rounding contrast

Lip-rounding results in lowered F2 and F3 [8, 10, 13, 22, 23] and is therefore useful in distinguishing between rounded and unrounded sounds. Because of the page limit, only the attempted data produced by

the Cantonese and SA participants were displayed in Fig. 3. Results show that the rounded [y] and [œ] had lower F3 values than the unrounded [i] and [ɛ] respectively across the board, indicating that the rounded vowels lip-rounding produced by the SA participants was generally smaller than the Cantonese participants. For instance, [ɛ] and [œ] produced by the Cantonese participants were significantly different in their F3 ($F(1,502) = 35.542, p < .001$), whereas for speakers of Urdu, both attempted and shadowed, no significant differences were found. In addition, the F3 difference between the high vowel pair [i y] produced by the SA participants is generally closer to the Cantonese reference than that for the mid vowel pair [ɛ œ]. For example, [i] produced by Punjabi-dominant speakers was not significantly different from the Cantonese [i] in F1 ($p = 1.000$), F2 ($p = .993$) and F3 ($p = .127$); nor was [y] significantly different from the Cantonese [y] in F1 ($p = .987$), F2 ($p = .955$) and F3 ($p = .258$). However, their [œ] and the Cantonese [œ] were significantly different in F1 ($p < .001$).

The linear mixed effect model (details not given here due to page limit) indicated that the phonetic environment had a significant effect on the performance of the participants in the discrimination task. Results illustrated that the SA participants performed significantly worse than the Cantonese participants when the vowels were in the pre-nasal ($F(1,76) = 6.250, p = .015$) and pre-stop ($F(1,76) = 7.443, p = .008$) environments, but not when the vowels appeared word-finally. In terms of language dominance, the Punjabi participants did significantly worse than the Cantonese speakers among the SA participants ($\beta = -0.99, t = -2.50, p < 0.05$).

Figure 2: F1 and F2 trajectories for [a:] and [ɐ] produced by the Cantonese (above) and SA (attempted, bottom left; shadowed, bottom right) speakers in the diphthong [a:u ɐu] environment.

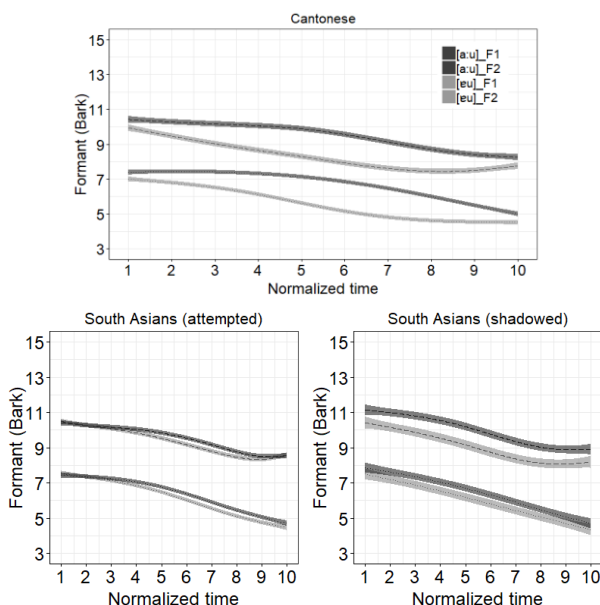
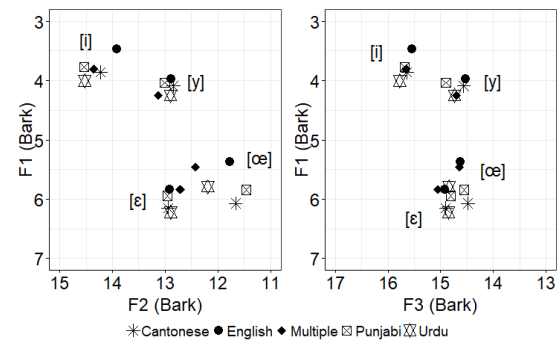


Figure 3: Relative positions of [i, y, ɛ, œ] produced by the Cantonese and SA (attempted) participants.



4. DISCUSSION

This study examined the acquisition pattern of Cantonese vowel length contrast and vowel rounding contrast by school-age SAs residing in Hong Kong. Overall, our research findings suggest that these SA participants were able to distinguish the contrasts of vowel length and vowel rounding. Nevertheless, their distinction of these contrasts was still worse than that of the Cantonese native speakers. For the vowel length contrast, the [a:] and [ɐ] shadowed by the SA participants, compared to their attempted production, were durationally closer to the Cantonese [a:] and [ɐ] respectively, suggesting that the SA participants were able to produce the durational contrast by imitating the experimenter. Regarding the vowel rounding contrast, the SA speakers had better separation of the rounding contrast for the high vowel pair (e.g. [i y]), which is especially so for participants who are dominant in Punjabi and Urdu.

Furthermore, there are variations in difficulty of a particular contrast in different phonetic environments. For example, the vowel length contrast is most difficult in diphthongs, while the vowel rounding contrast is more difficult in closed syllables (pre-stop and pre-nasal environments). Possible explanations for the difficulties include assimilation caused by the absence of vowel length contrast in diphthongs in Punjabi and Urdu and a change in quality of the rounded vowels caused by the presence of final consonants. In addition, even if the SA participants can make a phonological distinction in Cantonese, they make use of cues that are different from that of native Cantonese speakers, e.g. they only used durational difference to distinguish [a:] vs. [ɐ] in closed syllables while Cantonese speakers use both duration and quality. Our study displays the specific features of how SA students produce and perceive the Cantonese vowel system, providing a useful basis for further studies of advancing Cantonese learning by the ethnic minorities in Hong Kong.

5. REFERENCES

- [1] Bauer, R. S., Benedict, P. K. 1997. *Modern Cantonese Phonology*. Berlin; New York: Mouton de Gruyter.
- [2] Best, C. T. 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). Cambridge: MIT Press.
- [3] Bhardwaj, M. 2016. *Punjabi: A comprehensive grammar*. London: Routledge.
- [4] Carmichael, S. 2009. Language rights in education: a study of Hong Kong's linguistic minorities. In *Occasional Paper No. 19*. Hong Kong: Centre for Comparative and Public Law, The University of Hong Kong.
- [5] Census and Statistics Department. 2017. *Thematic report: Ethnic minorities*. Hong Kong: 2016 Population By-census Office, Census and Statistics Department.
- [6] Cheung, H.-N, S. 1972. *A grammar of Cantonese as spoken in Hong Kong*. Hong Kong: The Chinese University of Hong Kong.
- [7] Cheung, K. H. 1986. *The phonology of present-day Cantonese*. University College of London.
- [8] Fant, G. 1960. *Acoustic theory of speech production*. The Hague: Mouton.
- [9] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issue in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press.
- [10] Flemming, E. S. 2002. *Auditory representations in phonology*. New York: Routledge.
- [11] Gill, H. S., Gleason, H. A. 1969. *A reference grammar of Punjabi*. Patiala: Department of Linguistics, Punjabi University.
- [12] Gu, C. 2013. *Smoothing spline ANOVA models*. New York: Springer New York.
- [13] Johnson, K. 1997. *Acoustic and auditory phonetics*. Cambridge, Mass.: Blackwell Publishers.
- [14] Kachru, Y. 1990. Hindi-Urdu. In B. Comrie (Ed.), *The major languages of South Asia, the Middle East, and Africa*. London: Routledge.
- [15] Kao, D. 1971. *Structure of the syllable in Cantonese*. The Hague: Mouton.
- [16] Koul, O. N. 2008. Modern Hindi Grammar. *Journal of Asian Studies* (Vol. 19).
- [17] Lee, T. H. 1983. The vowel system in two varieties of Cantonese. *UCLA Working Papers in Phonetics*, (57), 97–114.
- [18] Macmillan, N. A., Creelman, C. D. 1991. *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- [19] Schmidt, R. L. 2007. Urdu. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 286–350). London: Routledge.
- [20] Shackle, C. 2007. Punjabi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages*. London: Routledge, pp. 581–621.
- [21] Shapiro, M. C. 2007. Hindi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 250–285). London: Routledge.
- [22] Stevens, K. N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.
- [23] West, P. 1999. The extent of coarticulation of English liquids: An acoustic and articulatory study. In *Proceedings of 14th ICPhS* (pp. 1901–1904). Citeseer.
- [24] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America*, 88(1), 97–100.
- [25] Wong, S. L. 1941. *A Chinese syllabary pronounced according to the dialect of Canton*. Hong Kong: Chung Hwa Book.
- [26] 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)*.
- [27] Yue-Hashimoto, O. 1972. *Phonology of Cantonese*. Cambridge: Cambridge University Press.
- [28] Zee, E. 1999. An acoustical analysis of the diphthongs in Cantonese. In *Proceedings of the XIVth International Congress of Phonetic Sciences*.