ARTICULATION OF CANTONESE AND ENGLISH SIBILANTS AMONG BILINGUAL SPEAKERS IN HONG KONG

Jonathan Havenhill, Ming Liu, Tak Wang Li

The University of Hong Kong
jhavenhill@hku.hk, mingliu_7@connect.hku.hk, u3574384@connect.hku.hk

ABSTRACT

English influence has been argued to be responsible for the recent emergence of an allophonic split between alveolar [t͡s, t͡sʰ] and alveolo-palatal [t͡ɕ, t͡ɕʰ] alternants for the affricates of Hong Kong Cantonese (HKC). However, the phonetic similarity and phonological relationship between English and HKC sibilants has not been empirically established. This study uses ultrasound tongue imaging with synchronized audio and lip video to examine the production of English and Cantonese sibilants among native HKC speakers with varying levels of English proficiency. Participants recited a Cantonese word list containing Cantonese /s, t͡s, t͡sʰ/ and an English word list containing /s, ʃ, t͡ʃ, d͡ʒ/. While L1 English, L2 English, and L1 HKC sibilants are largely similar in terms of spectral center of gravity, a high degree of interspeaker variability is observed in articulation, primarily with respect to lip rounding. These findings are considered with respect to their implications for theories of bilingual phonological representation and of contact-induced change.

Keywords: ultrasound tongue imaging, articulation, sibilants, Hong Kong Cantonese, bilingualism

1. INTRODUCTION

Hong Kong is a multilingual environment where both English and Cantonese are in common use and have been in continuous contact for several centuries. Although Hong Kong English (HKE) has received increasing attention as a distinct variety of English [1], English is predominantly learned through schooling as a second language by native speakers of Cantonese [2]. HKE therefore exhibits a number of phonological patterns that derive from the sound system of Hong Kong Cantonese (HKC). These include alternation between [n] and [l], *th*-fronting, and consonant cluster reduction, among others [3]. Substitutions of [s] for [ʃ] (and vice versa) have also been reported. HKC is canonically described as having a single series of alveolar sibilants: the voiceless fricative /s/ as well as aspirated and unaspirated affricates /t͡s, t͡sʰ/ [4]. Given the lack of an English-like alveolar vs. post-alveolar contrast, Chan and Li [5] report that Hong Kong speakers show a merger of English minimal pairs like *save-shave* and *sip-ship* in which both words are produced with [s]. They note that this pattern is reversed before back round vowels, such that "soup" is pronounced as [ʃuːp], although the opposite ([sʊd] for "should") has also been observed [6]. The (typically rounded) English affricates [t͡ʃ, d͡ʒ] are likewise reported to be replaced by [t͡sʰ, t͡s], described as alveolar with lip spreading [5].

However, a handful of Cantonese studies (e.g., [7, 8]) report the emergence of an allophonic split between alveolar [t͡s, t͡sʰ] and alveolo-palatal [t͡ɕ, t͡ɕʰ] or post-alveolar [t͡ʃ, t͡ʃʰ]. Recent acoustic and articulatory data confirm that young HKC speakers use distinct tongue gestures for producing the two allophones, and that the allophones are conditioned by vowel rounding [9]. This change has typically been attributed to English contact, with Cheung [7] asserting that "speakers equate the rounded Cantonese sibilants (basically alveolopalatals) with the English palatoalveolars" and that they "substitute the English palatoalveolars for the usual realization of Cantonese sibilants when these are followed by rounded vowels" (p. 202). Despite this claim, phonetic similarity of the HKC and English sibilants has not been explicitly tested.

Transfer from English to Cantonese is to some extent plausible. Most theories of L2 perception and learning (including PAM/PAM-L2 [10] and SLM-r [11]) assume that L2 learners take as their starting point the phonetic and phonological system of their L1. Phonological acquisition begins with mapping L2 sounds to equivalent L1 sounds according to their phonetic similarity [11]. Such combined L1-L2 categories may subsequently experience phonetic shifts under influence from both L1 and L2 input.

The phonetic properties that comprise such categories, however, have not been fully established. Most studies of L2 production rely on acoustic data, although PAM-L2 and SLM-r differ significantly

with respect to the phonological status of articulatory gestures. Whether the phonetic targets of nonnative speech production are acoustic or articulatory has been examined by Oakley [12], who finds that L2 speakers can adopt a range of distinct articulatory configurations in producing nonnative sounds, and that speakers may target both acoustics and articulation.

This study aims to determine how Cantonese-English bilingual speakers produce both L1 Cantonese and L2 English sibilants. Questions to be addressed include to what extent HKC speakers re-use L1 phonetic targets in producing L2 sibilants; whether L1 and L2 sibilants are similar in acoustics, articulation, or both; and whether native HKC speakers show similarity to native English speakers.

2. METHODS

2.1. Participants

8 L1 speakers of HKC (1 man, 7 women) and 4 native US/UK English (NE) speakers (all women) participated in the study. NE speakers were aged 21-32 years (mean 26.5; SD 3.9). HKC speakers were born between 1996-2004 (mean age 20 years; SD 2.1) and raised in Hong Kong at least through age 18. All began acquiring English between ages 2-6, but show varying levels of English proficiency and use. Proficiency was tested via two pre-screening tasks, C-Test [13] and LexTale [14], and through collection of standardized English exam scores. 4 high and 4 low proficiency speakers were recorded, but results do not clearly pattern according to proficiency, which is not further considered here.

2.2. Materials

HKC participants were first asked to read an HKC wordlist, followed by an English wordlist. The HKC wordlist, adapted from [9], included 72 disyllabic words with target onsets /s t͡s t͡sʰ/ followed by the vowels /i, ɛ, ɐ, a, y, œ, u, ɔ/, as well as 30 filler words with onsets /kʰ, kʰʷ, k, kʷ, w/. Target onsets appeared in the first syllable with either Tone 1 (high level) or Tone 3 (mid level). Each word was presented in the carrier phrase [ŋɔ:˧ sɛ:˥ ____ jɐt̚˥ tsʰi:˧] (“I write ____ one time”), written in traditional Chinese characters. The English wordlist contained 96 words with the onsets /s, ʃ, t͡ʃ, d͡ʒ/ and the vowels /i, ɪ, e, ɛ, u, ʊ, o, ɔ/, appearing in the first syllable of a stress-initial word. Words were embedded in the carrier phrase ‘say ____ each time’. NE speakers produced the same English wordlist as HKC speakers.

2.3. Procedure

The experiment was conducted in a sound-attenuated booth at the authors’ university. Stimuli were presented to each speaker in a unique pseudo-random order using Articulate Assistant Advanced (AAA) [15]. Each prompt was repeated three times. The HKC wordlist yielded 216 target tokens per speaker (excluding filler items), while the English wordlist generated 288 tokens. The total number of target tokens was 5,184 across 12 participants.

Ultrasound data were collected using an Articulate Instruments SonoSpeech Micro ultrasound system with a 20mm radius 2–4MHz transducer, which captured ultrasound images at an average frame rate of 84 frames per second (fps). Lip video was captured at 60 fps with front-view and side-view cameras. Both the ultrasound transducer and lip cameras were held in place with a stabilizing headset worn by the participant. Audio was recorded at a 48kHz sample rate and 16-bit sample depth on a Denon F650R solid state recorder, using a Sound Devices USBPre2 preamplifier and an Earthworks Ethos cardioid condenser microphone placed 5-10 cm from the side of speaker’s mouth. Audio was simultaneously recorded in AAA, which was used to synchronize the acoustic and articulatory recordings. At the beginning and end of each session, an image of the occlusal plane was captured using a plastic biteplate and a palate trace was recorded. For HKC speakers, both word lists were recited in a single session with a short break, but without removal or adjustment of the headset.

2.4. Analysis

Acoustic recordings were segmented using the Montreal Forced Aligner [16] and manually corrected. For the affricates /t͡s, t͡sʰ, t͡ʃ, d͡ʒ/, the first four spectral moments [17] were measured at the midpoint of frication, typically corresponding to approximately 75% of overall constriction duration. For the fricatives /s, ʃ/, spectral measurements were taken at 75% of frication duration, such that all target sibilants were measured at a consistent distance from the vowel onset.

Ultrasound tongue contours were automatically tracked with DeepLabCut [18] using the pre-trained MobileNet1.0-based neural network implemented in AAA [19]. Tongue splines were extracted for each token at a single time point corresponding to the acoustic measurements. Tongue contours were analyzed using polar SSANOVA [20], with splines fit to each combination of language, sibilant, and vowel rounding. Front-view lip video was

also tracked with DeepLabCut using a pre-trained MobileNetV2 network [19]. Rounding was quantified by calculating the degree of horizontal lip opening between the oral commissures [21] for frames corresponding to the acoustic measurements.

3. RESULTS

3.1. Acoustic Results

Combined spectral measurements for all speakers are provided in Figure 1. The patterns observed here are representative of individual speaker results, with only minor variance. Given the reported equivalence between HKC [t͡sʰ, t͡s] and English [t͡ʃ, d͡ʒ] [5, 7], L1-L2 affricate pairs are displayed together for HKC speakers, as are L1 and L2 [s].

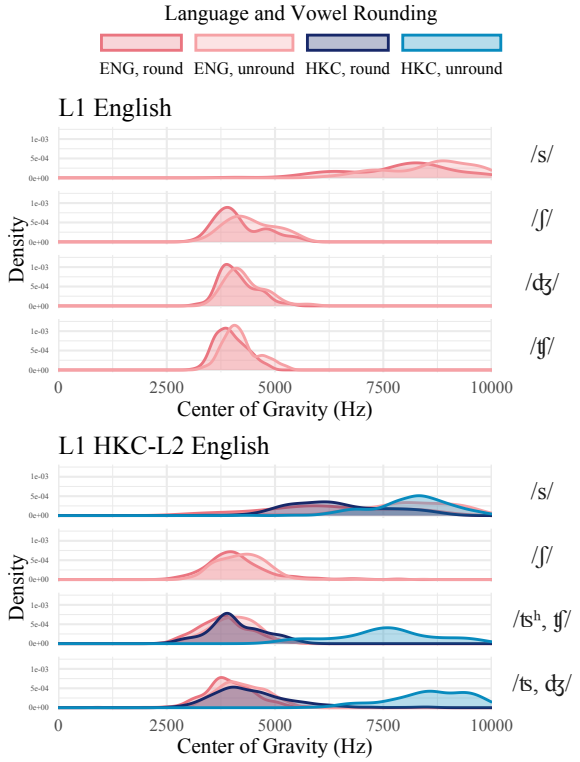

Figure 1: Spectral COG for all participants

In accordance with recent studies [8, 9], HKC speakers show evidence of an allophonic split for their L1 affricates, which differ significantly in COG according to vowel rounding ($p < 0.001$). Although the COG for /s/ (both L1 and L2) also varies significantly by vowel rounding ($p < 0.001$), the effect is much smaller.

With respect to acoustic similarity between L2 English and L1 HKC, it is observed that all three English post-alveolars pattern like the posterior allophones of the HKC affricates that occur in round environments. This is the case regardless of vowel rounding, and the L2 English affricates do not exhibit the same allophonic split observed in HKC. This result suggests an equivalence between the HKC allophones [t͡ʃʰ, t͡ʃ] and the English phonemes /t͡ʃ, d͡ʒ/. The new L2 sound [ʃ] is likewise produced with a low COG similar to that of the [t͡ʃʰ, t͡ʃ] allophones, and does not show an allophonic split.

Finally, the HKC and L2 English affricates, as well as [ʃ], show strong acoustic similarity to the L1 English post-alveolars. HKC and L1 English /s/ show similarly high COG in unround vowel environments, but HKC and L2 English /s/ show lower COG in round environments.

3.2. Ultrasound Results

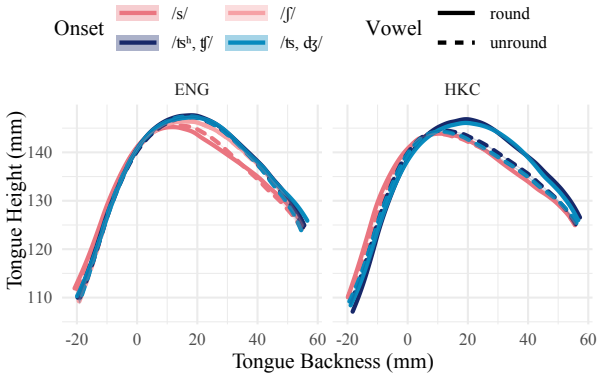

Figure 2: SSANOVA for HKC05. Tongue front to the right.

Two patterns of lingual articulation are observed among the HKC speakers. The predominant pattern, observed for six speakers, is shown in Figure 2. Corroborating the acoustic results, all HKC speakers clearly show an allophonic split for the two affricates. /t͡sʰ, t͡s/ show distinct apical vs. laminal tongue gestures according to vowel rounding. The same apical tongue gesture is also used for [s], although [s] does not vary by vowel rounding. Acoustic differences observed for [s] in round environments can therefore likely be attributed to anticipatory vowel rounding. In L2 English, these speakers show similarity both to their L1 HKC and to the native English speakers. L2 [s] is produced with an apical alveolar tongue gesture, while the three post-alveolars use a laminal post-alveolar gesture similar to that of pre-round HKC [t͡ʃʰ, t͡ʃ]. This pattern is identical to that observed for all native English speakers in this study.

The second pattern is shown in Figure 3. These speakers use similar tongue gestures for their L1 and L2 [s], and for [t͡sʰ, t͡s] in unround environments.

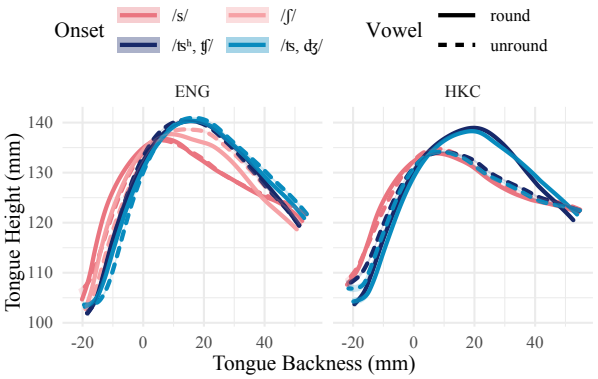

Figure 3: SSANOVA for HKC06.

Likewise, the tongue gestures for L2 [t͡ʃ, d͡ʒ] are similar to L1 [t͡ʃʰ, t͡ʃ] in round environments. Most notably, however, these speakers produce English [ʃ] with a tongue position intermediate to that of [s] and that of [t͡ʃ, d͡ʒ]. Moreover, [ʃ] shows an unexpected split according to vowel rounding, with a higher tongue blade in unround environments.

3.3. Lip Video Results

All native English speakers were consistent in their use of secondary lip rounding on the post-alveolar (but not alveolar) sibilants. For all HKC speakers, the HKC sibilants are variably rounded in accordance with the following vowel. However, the degree of interspeaker variability in rounding among HKC speakers is greater than in either acoustics or tongue position, with several distinct strategies. In L2 English, three speakers show a native English-like pattern with consistent rounding on all post-alveolars and coarticulatory rounding for [s].

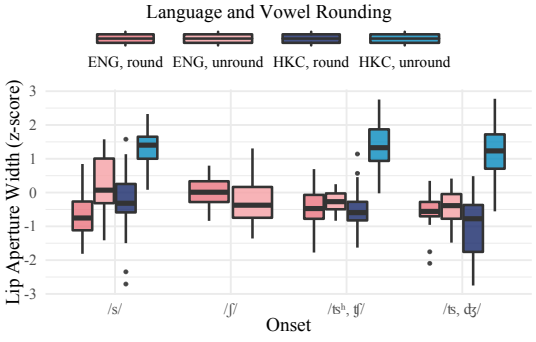

Figure 4: Lip rounding for HKC06

The speaker in Figure 4 (and two others) shows consistent rounding for the English affricates, similar to to the degree of rounding for HKC sibilants in round environments. However, she shows less consistent rounding for for English [s, ʃ], frequently rounding English [s] even in unround environments.

The speaker in Figure 5 (and one other), on the other hand, show the expected coarticulatory rounding for both HKC and English [s], and for the HKC affricates. However, these speakers use significantly less rounding for English post-alveolars, suggesting they have not fully acquired the secondary rounding typical of US/UK English.

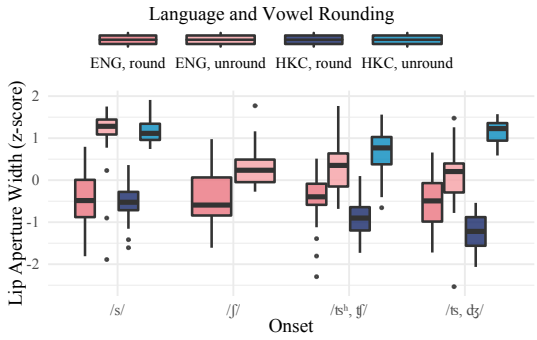

Figure 5: Lip rounding for HKC01

4. DISCUSSION AND CONCLUSION

This study finds substantial similarity between L1 HKC, L2 English, and L1 English sibilants, particularly in spectral center of gravity and tongue position. For the "similar" [11] sounds /t͡sʰ, t͡ʃ/ and /t͡s, d͡ʒ/, nearly all HKC speakers use the same tongue gestures in both their L1 and L2, yielding similar acoustic output for the affricate pairs. For the new sound /ʃ/, most speakers use the same posterior tongue gesture used for /t͡ʃ, d͡ʒ/, but two use novel gestures intermediate to [s] and [t͡ʃ]. For the identical /s/ sounds, speakers use the same tongue gesture in both languages, again with similar COG.

However, substantial interspeaker variation is observed with respect to lip rounding. Thus, like [12], this study finds that speakers do not necessarily use the same *combinations* of articulatory gestures in their L1 and L2. Rather, speakers may recruit individual gestures to achieve an L2 target. Yet these findings also contrast with [12] in that lip gestures were found to be more variable than tongue gestures, perhaps due to their association with coarticulatory vowel rounding in Cantonese and their status as secondary features in English.

Despite the Cantonese-English similarities, it cannot yet be determined to what extent allophonic split for the HKC affricates can be attributed to English influence. Real and apparent-time data (collection of which is underway) are needed to observe the early stages of that split, while sociolinguistic studies are necessary to determine whether the split is associated with Cantonese-English bilingualism. Nevertheless, this study provides phonetic data needed for such future work.

REFERENCES

[1] C. C. M. Sung, "Hong Kong English: Linguistic and sociolinguistic perspectives," *Language and Linguistics Compass*, vol. 9, no. 6, pp. 256–270, 2015.

[2] J. Setter, C. S. Wong, and B. H. Chan, *Hong Kong English*. Edinburgh University Press, 2010.

[3] T. T. N. Hung, "Towards a phonology of Hong Kong English," *World Englishes*, vol. 19, no. 3, pp. 337–356, 2000.

[4] S. Matthews and V. Yip, *Cantonese: A Comprehensive Grammar, Second Edition*. London: Routledge, 2011.

[5] A. Y. Chan and D. C. Li, "English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems," *Language Culture and Curriculum*, vol. 13, no. 1, pp. 67–85, 2000.

[6] K. K. Luke and J. C. Richards, "English in Hong Kong: Functions and status," *English World-Wide*, vol. 3, no. 1, pp. 47–64, 1982.

[7] K. H. Cheung, "The phonology of present-day Cantonese," doctoral dissertation, University of London, 1986.

[8] K.-L. Chan, "The sound change of [ts, tsʰ, s] to [tʃ, tʃʰ, ʃ] in Hong Kong Cantonese," Master's thesis, The Hong Kong University of Science and Technology, Hong Kong, 2007.

[9] P. H. Yeung and J. Havenhill, "Acoustic ambiguity and articulatory re-analysis: Variation and change in the Hong Kong Cantonese sibilants," to appear.

[10] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. Philadelphia, PA: John Benjamins, 2007, pp. 13–34.

[11] J. E. Flege and O.-S. Bohn, "The revised speech learning model (SLM-r)," in *Second Language Speech Learning*, R. Wayland, Ed. Cambridge: Cambridge UP, 2021.

[12] M. Oakley, "Articulating non-native vowel contrasts," doctoral dissertation, Georgetown University, 2021.

[13] T. Eckes and R. Grotjahn, "A closer look at the construct validity of C-tests," *Language Testing*, vol. 23, no. 3, pp. 290–325, 2006.

[14] K. Lemhöfer and M. Broersma, "Introducing LexTALE: A quick and valid lexical test for advanced learners of English," *Behavior research methods*, vol. 44, no. 2, pp. 325–343, 2012.

[15] Articulate Instruments Ltd., *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd., 2012.

[16] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

[17] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *The Journal of the Acoustical Society of America*, vol. 108, pp. 1252–1263, 2000.

[18] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," *Nature neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.

[19] A. Wrench and J. Balch-Tomes, "Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut," *Sensors*, vol. 22, 2022.

[20] J. Mielke, "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2858–2869, 2015.

[21] J. Havenhill and Y. Do, "Visual speech perception cues constrain patterns of articulatory variation and sound change," *Frontiers in Psychology*, vol. 9, p. 728, 2018.