GESTURAL REDUCTION OF HONG KONG CANTONESE SYLLABLE-FINAL ORAL STOPS

Jonathan C. Yip

The University of Hong Kong
yipjonat@hku.hk

ABSTRACT

Previous work on Cantonese unreleased syllable-final stops have shown that these sounds are cued primarily by their spectral transitions during the preceding vowel (Khouw & Ciocca, 2006). However, it remains unclear whether speakers have lost a place distinction between such sounds. This study investigates the articulation of syllable-final /t, k/ that precede consonants differing by place (labial, alveolar, velar), which could result in varying degrees of gestural reduction. Productions of disyllabic Cantonese words by five speakers were recorded using ultrasonic tongue imaging and analyzed for lingual contour and minimal lingual aperture. Results for all but one speaker show systematic patterns of reduction of the anterior or posterior constrictions for /t, k/ according to the place of the following consonant. Unexpectedly, the pre-labial contexts elicited the most gestural reduction and deletion/assimilation, despite the fact that this context should have permitted the greatest amount of simultaneity and coarticulation.

Keywords: production, gestural reduction, Cantonese, articulatory coordination, ultrasound.

1. INTRODUCTION

This study investigates the effect of phonetic context on the magnitude of the anterior and dorsal lingual gestures for unreleased syllable-final oral stop sounds [t, k] by speakers of Hong Kong Cantonese. Previous work on the series of Cantonese unreleased coda oral stops /p, t, k/ has focused on what acoustic cues to place are present in the acoustic stream in the obligatory absence of plosive release cues for these sounds in most phonetic environments. [2] and [5] found that cues to unreleased stop place are provided solely by F1, F2 , and F3 transitions during the preceding vowel into the stop coda constriction. However, as noted by [14] and [7], younger-generation speakers of Hong Kong Cantonese tend to produce these syllables with full glottal closure in lieu of or immediately preceding the oral stop gesture. Because of the acoustic nature of unreleased stops, i.e. silence, it is difficult to ascertain from the acoustic findings the extent to which the articulatory actions involved during unrealized coda stop productions occur in normal productions by Cantonese speakers. It is potentially the case that the production goals of younger speakers do not require an oral stop gesture at all for these sounds, as a consequence of an ongoing process of phonological change. Hence, the aim of this study is to shed light on this phenomenon by using ultrasonic-imaging technology to track changes in the shape and movement of the tongue and determine whether and when Cantonese speakers produce physical gestures for such sounds.

2. METHODS

2.1. Data Collection

2.1.1. Participants

The participants were two male and three female native speakers of Cantonese who were born, grew up, and received their formal education in Hong Kong. Their ages ranged from 18 to 26 years (mean: 20). No participants reported any hearing or speaking deficits.

2.1.2. Test Materials

Target stimuli were a set of disyllabic spoken Cantonese lexical items, each constructed by combining each of the four stop-final morphemes 發 [faːt^3], 法 [faːt^3], 拍 [p^{h}aːk^3], and 白 [paːk^2] with a following morpheme that contained a labial, alveolar, or velar onset consonant (two words for each place context). The four initial morphemes were chosen based on three criteria: 1) they contained a labial obstruent onset, 2) they contained the low central long vowel [aː], and 3) they were among the set of medium- to high-frequency morphemes found in the Hong Kong Cantonese Adult Language Corpus (HKCAC) [8]. The full list of target words contained 24 lexical items, listed in Table 1. Target items were combined with 18 filler items to obscure the aim of the study during the production task.

Table 1: The full sets of target disyllabic items with initial morphemes ending in a /t/ or /k/ coda.

Context	**Initial-Syllable Coda Articulator**			
	Tongue Tip		*Tongue Dorsum*	
Labial	發佈 faːt^{3}.pou^{33} 'publish' 發火 faːt^{3}.fɔ^{25} 'get angry'	法文 faːt^{3}.mɐn^{2*5} 'French (lang.)' 法網 faːt^{3}.mɔŋ^{23} 'the arm of the law'	拍板 p^{h}aːk^{3}.paːn^{25} 'clapperboard' 拍片 p^{h}aːk^{3}.pin^{2*5} 'shoot a film'	白板 paːk^{2}.paːn^{25} 'whiteboard' 白飯 paːk^{2}.faːn^{22} 'plain cooked rice'
Alveolar	發達 faːt^{3}.taːt^{2} 'get rich' 發動 faːt^{3}.toŋ^{22} 'start, initiate'	法定 faːt^{3}.teŋ^{22} 'legal, statutory' 法典 faːt^{3}.tin^{25} 'legal codex'	拍檔 p^{h}aːk^{3}.tɔŋ^{33} 'partner' 拍拖 p^{h}aːk^{3}.t^{h}ɔ^{55} 'go out on a date'	白豆 paːk^{2}.tɐu^{2*5} 'soybean' 白糖 paːk^{2}.t^{h}ɔŋ^{21} 'white sugar'
Velar	發覺 faːt^{3}.kɔk^{3} 'realize, discover' 發掘 faːt^{3}.k^{w}ɐt^{2} 'dig, exhume'	法國 faːt^{3}.k(^{w})ɔk^{3} 'France' 法官 faːt^{3}.kun^{55} 'judge (n.)'	拍擊 p^{h}aːk^{3}.kek^{5} 'smack, beat' 拍劇 p^{h}aːk^{3}.k^{h}ɛk^{2} 'shoot a soap opera'	白鴿 paːk^{2}.kaːp^{2*5} 'dove, pigeon' 白金 paːk^{2}.kɐm^{55} 'platinum'

2.1.3. Procedure

Speakers produced two randomized iterations of each target stimulus in the carrier phrase "再講___個字" [tsɔi^{33} kɔŋ^{25}____ kɔ^{25} kɔ^{33} tsi^{22}] 'Say the word ____ again.' Audio recordings of speakers' productions were recorded with an Oktava MK-012 condenser microphone with cardioid polar capsule attachment. Articulatory data were collected using a Telemed ClarUs ultrasonic beamformer and MC4-2R20N convex transducer (2–4 MHz frequency range), in conjunction with Echo Wave II ultrasound imaging software [13]. During the production task, the speaker rested his/her forehead onto the softly padded ends of two freely posable camera arms, keeping the head fixated, while a third camera arm held the transducer at stable position for a midsagittal scan of the surface of the speaker's tongue. Continuous ultrasound video was captured with Fraps [1] real-time video capture software at a rate of 60 frames per second.

2.2. Data Processing

The acoustic signal was synchronized with the ultrasound video signal by gently tapping the ultrasound probe against the microphone during simultaneous video and audio capture, allowing for the acoustic event in the acoustic stream to be aligned temporally to the brief shaking of ultrasonic gel in the ultrasound video. Ultrasound data were extracted as sequences of consecutive frames taken from the acoustic intervals for /...aːt.CV.../ and /...aːk.CV.../ in the relevant word items. For each analysis frame, a smoothing spline was fit to the lower edge of the visual lingual contour in the ultrasound image using EdgeTrak [9].

From each frame, measures of aperture distance at the tongue-tip and the tongue-dorsum were assessed based on the location of the palate as imaged during the speaker's swallowing of a water bolus during scanning. Because the flexible soft palate is normally pushed upward by the tongue during swallow, palate traces were based on frames taken just at the end of swallowing. Regions along the palate contour relevant to tongue-tip and -dorsum constriction were identified based on the location of constriction locations during 5 randomly selected tokens of each constriction place. Aperture distance values were calculated by taking the minimum Euclidean distance between the region-of-interest points along the palate to any point along the tongue contour. Times for each aperture distance value were taken from the temporal offset of voicing of the vowel in the first syllable.

2.3. Analysis

The degree of achievement for each coda constriction gesture was assessed by extracting the minimum distance of lingual aperture during the voiceless intervals of the sequences /...aːt.C.../ and /...aːk.C.../ from the total set of aperture data. These minimum distance values were pooled across the two speakers and compared in a linear mixed-effects model (LMM), with Coda Articulator and Phonetic Context as fixed effects and Speaker and Item as random effects. This model was computed using the lmer() function of the lmerTest package in R [6]. Additionally, comparisons of the variation among lingual contour extracted at V1 offset in the planned phonetic contexts were made using Smoothing Spline ANOVA (SS ANOVA) models. SS ANOVAs for each articulator of interest (tongue tip in coda-/t/ contexts, tongue dorsum in coda-/k/ contexts) were computed for each speaker using the ssanova() function of the gss package in R [3].

3. RESULTS

3.1. Minimal Lingual Aperture

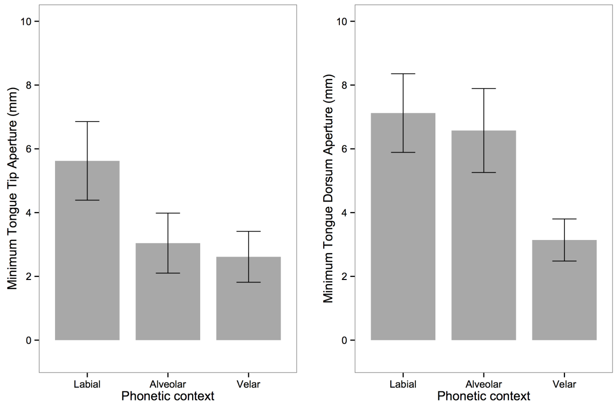
Comparisons of minimal lingual aperture distance during the intervals /...t#C.../ and /...k#C.../ are shown in Fig. 1. LMM results indicate that in the coda-/t/ context, tongue-tip aperture was greater in the pre-labial context than in the pre-alveolar context ($\beta = -2.5798$, $t = -3.669$) and in the pre-velar context ($\beta = -3.0049$, $t = -4.255$). However, tongue-tip aperture in the pre-alveolar and pre-velar contexts did not differ significantly ($t = 0.602$). Correspondingly, results for the coda-/k/ context indicated that minimum tongue-dorsum aperture during the /...k.C.../ interval was greatest in the pre-labial context. While dorsal aperture in the pre-labial context was significantly greater than that in the pre-velar context ($\beta = -3.9814$, $t = -5.662$), dorsal aperture in the pre-alveolar context was equally as large as that in the pre-labial context ($\beta = -0.5477, t = -0.779$).

Together, the results for coda-/t/ and coda-/k/ suggest that the lingual gestures for coda-position oral stops are heavily reduced or deleted when the onset of the following syllable is labial regardless of (lingual) articulator. When the following onset is alveolar, the degree of gestural reduction depends on whether the articulator and the onset-context are homorganic, with gestural reduction occurring when the coda is /k/. However, both coda-/t/ and coda-/k/ gestures are preserved when the onset of the following syllable is velar, suggesting that any type of lingual raising—whether anterior or posterior—might facilitate the formation of an ensuing dorsal constriction. Importantly, coda-/k/ in the pre-alveolar context resulted in a pronounced reduction of the tongue-dorsum raising gesture.

3.2. Lingual Contours at V1 Offset

SS-ANOVA results show a degree of variation in the actual contour of the tongue during the offset of the vowel preceding the stop codas /t/ and /k/. Among the five speakers, three distinct patterns emerged:

Figure 1: Barplot of minimum tongue-tip aperture during /...t.C.../ (left) and minimum tongue-dorsum aperture during /...k.C.../ (right).



1. tongue-tip and -dorsum constriction achievement only when the following consonant is lingual (speakers S1 and S2)
2. tongue-tip constriction achievement when the following context is lingual, but tongue-dorsum constriction achievement only when the following context is velar (speakers S3 and S4)
3. full constriction of the tongue-tip and -dorsum gestures regardless of the following context, i.e. no effect of place of the following consonant (speaker S5).

Example SS-ANOVA plots for patterns 1 and 2 are presented in Figs. 2 and 3 (respectively). In these plots, it is important to note that, although an oral constriction for coda /t/ or /k/ may not have been achieved, the interval of stop-to-stop closure is likely to have been initiated with a full constriction at the glottis.

The SS-ANOVA results for speakers S3 and S4 pattern most closely to the LMM results for maximum lingual aperture during the stop+stop interval. These speakers seem to prefer anterior (apical) raising over posterior (dorsal) raising during coordination of consecutive stop gestures. Speakers S1 and S2, on the other hand, achieved tongue-tip and -dorsum constrictions whenever an upcoming sound gesture was also lingual (regardless of place along the tongue). Lastly, speaker S5 (not shown) always achieved /k/ and /t/ constriction targets, without influence of upcoming sound gestures.

4. DISCUSSION

The results presented in Section 3 present interactions between the nature of the coda oral-stop articulator and the place of the following consonantal gesture. Namely, it is not simply the case that when

Figure 2: SS-ANOVA results for speaker S2 for the tongue contour at the onset of the voiceless interval associated with coda /t/ (top) and coda /k/ (bottom).

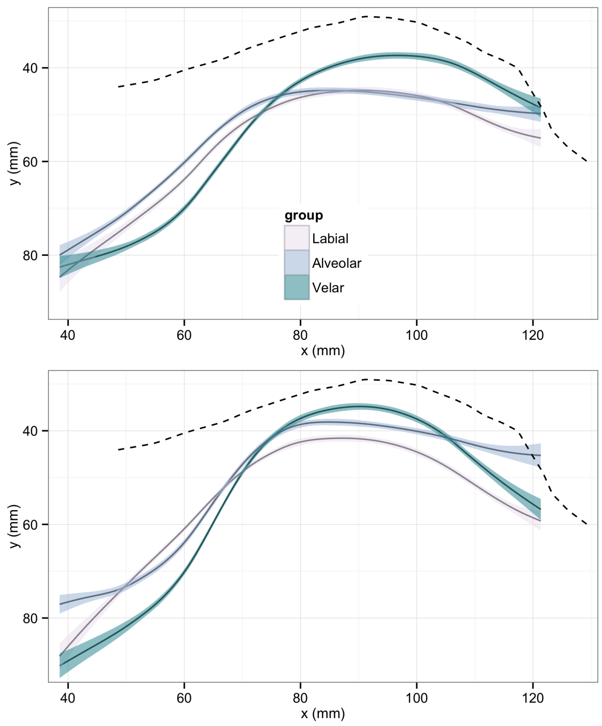
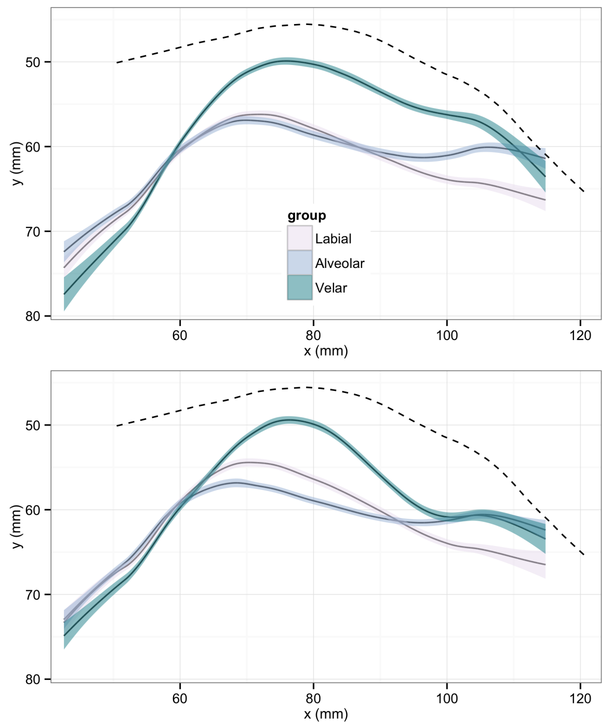


Figure 3: SS-ANOVA results for speaker S3 for the tongue contour at the onset of the voiceless interval associated with coda /t/ (top) and coda /k/ (bottom).



an oral-stop coda is followed by a non-homorganic onset consonant, the magnitude of the constriction gesture is reduced, as this is only partially true, i.e. when the context is pre-labial. When the gesture involves tongue-tip raising, the coda gesture is preserved when the place of the following onset consonant is either velar or alveolar. However, as seen from the SS-ANOVA results for speakers S1 and S2, the completion of this gesture might not occur until during the voiceless interval, during which there are no acoustic cues to gestural completion.

The articulatory patterns presented here may be driven by known observations of tongue dynamics, as reported in previous studies ([12], [11], [10], and [4]), in which the tongue shows a general tendency to move in a sequential posterior-to-anterior "looping" motion to achieve a combination of two oral constriction gestures that differ in their lingual articulators. It is possible that for the coordination of Cantonese unreleased oral stops with a following consonantal gesture, there exists some degree of articulatory compatibility when multiple lingual gestures are involved, such that if the gesture for the first (coda) consonant requires any kind of lingual raising, the tongue may be in an optimal position to produce full constrictions for an ensuing, lingually-articulated sound.

Notably, in pre-labial contexts, the lingual gestures were reduced to a significantly higher degree than the same gestures in the other place contexts. From a mechanical perspective, there should be little to no tradeoff between the achievement of consecutive—or even simultaneous—lingual and labial gestures. This effect suggests that the loss of coda-stop gestures before labial sounds has been phonologized.

Additional work (not reported here) involves an exploration of the acoustic cues to anterior and dorsal raising available during preceding vowel in order to understand how such cues map onto the articulatory observations. Because heavy gestural reduction of both lingual gestures were found in the pre-labial context, it is expected that there will be the fewest acoustic cues to coda-/t/ or coda-/k/ in this context.

5. REFERENCES

[1] Beepa, 2004–2015. Fraps [computer program]. Version 3.5.99. http://www.fraps.com. (accessed 15-Nov-14).

[2] Ciocca, V., Wong, L., So, L. 1994. An acoustic analysis of unreleased stop consonants in word final positions. *Proceedings of the International Confer-*

ence on Spoken Language Processing, Yokohama volume 21 1131–1134.

[3] Gu, C. 2014. gss: General smoothing splines. [R Package]. Version 2.1-4. http://cran.r-project.org. (last accessed 26-Jan-15).

[4] Iskarous, K. 2005. Patterns of tongue movement. *Journal of Phonetics* 33, 363–381.

[5] Khouw, E., Ciocca, V. 2006. An acoustic and perceptual study of final stops produced by profundly hearing impaired adolescents. *Journal of Speech, Language, and Hearing Research* 49, 172–185.

[6] Kuznetsova, A., P., B., Christensen, R. 2012. lmerTest: Tests in linear mixed effect models. [R Package]. Version 2.0-20. http://cran.r-project.org. (last accessed 25-Jan-15).

[7] Law, S.-P., Fung, R. S.-Y., Bauer, R. 2001. Perception and production of Cantonese consonant endings. *Asia Pacific Journal of Speech, Language and Hearing* 6, 179–195.

[8] Leung, M.-T., Law, S.-P. 2001. HKCAC: The Hong Kong Cantonese Adult Language Corpus. *International Journal of Corpus Lingusitics* 6(2), 305–325.

[9] Li, M., Akgul, Y., Kambhamettu, C. 2005. Edgetrak [computer program]. http://vims.cis.udel.edu/EdgeTrak. (accessed 15-Dec-14).

[10] Mooshammer, C., Hoole, P., Kühnert, B. 1995. On loops. *Journal of Phonetics* 23, 3–21.

[11] Munhall, K., Jones, J. 1995. The spatial control of speech movement. In: Bell-Berti, F., Raphael, L., (eds), *Producing speech: contemporary issues. For Kathrine Safford Harris*. AIP Press.

[12] Munhall, K., Ostry, D., Flanagan, J. 1991. Coordinate spaces in speech planning. *Journal of Phonetics* 19, 293–307.

[13] Telemed, 2014. Echo Wave II [computer program]. Version 3.3.2. http://www.pcultrasound.com. (accessed 2-Dec-14).

[14] Zee, E. 1999. Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics* 27, 120–167.