Co-Articulation between Consonant and Vowel in Cantonese Syllables

Wai-Sum Lee

Abstract—This study investigates C-V and V-C co-articulation in Cantonese monosyllables of the CV, VC or CVC structure, with C = one of the three stop consonants [p, t, k] and V = one of the three corner vowels [i, a, u]. Five repetitions of each test syllable on a randomized list were elicited from Cantonese young adult speakers in their early-20s. A research tool, EMA AG500, was used to record the synchronized audio signals and articulatory data at three different locations of the tongue – tongue tip, tongue middle, and tongue back – and the positions of the upper and lower lips during the test syllables. The main findings based on the articulatory data collected from two male Cantonese speakers are as follows: (i) For the syllable-initial [-p-], strong co-articulation is observed when [-p-] preceding the high vowel [i] or [u], but not the low vowel [a]. As for the syllable-final [-p], it is strongly co-articulated with the preceding vowel, even when the vowel is [a]. (ii) The co-articulation between the initial [-t-] and the following vowel of any type is weak. In the syllable-final position, the degree of co-articulatory resistance of [-t] is also large when following the vowel [u], but [-t] is largely co-articulated with the preceding vowel when the vowel is [i] or [a]. (iii) The strength of co-articulation differs when the initial [-k-] precedes the different types of vowel. A stronger co-articulation between [-k-] and [i] than between [-k-] and [u], and the strength of co-articulation is much reduced between [-k-] and [a]. However, in the syllable-final position, there is strong co-articulation between [-k] and the preceding vowel [a]. (iv) Among the three types of stop consonants in the syllable-initial position, the decreasing degree of co-articulatory resistance (CR) is [-k] > [-k-] > [-p-], and the degree of CR is reduced during all three types of stop in the syllable-final position. In general, the data on co-articulation between consonant and vowel in the Cantonese monosyllables are similar to those in other languages reported in previous studies.

Keywords—Cantonese, co-articulation, consonant, vowel.

I. INTRODUCTION

This study is an investigation of co-articulation between consonant and vowel in Cantonese syllable production. The phenomenon of co-articulation has drawn the attention of phoneticians as early as the late 1800s. For instance, it is described in [1] that speech sounds like “points in a stream of incessant change,” suggestive of co-articulation of segments due to the transitional movements that bring together different articulatory targets. In explicating co-articulation, the ‘overlapping innervation wave theory,’ analogizing segment as an invariant wave proposed in [2] that ‘waxes and wanes’ and waves for successive sound segments overlap in time. In [3], [4] the investigations of vowel reduction in Swedish CVC syllables show the influence of neighboring segments, in that as duration decreases, the vowel formants undershoot the target values towards the values of the neighboring consonants resulted from co-articulation. It is further revealed in [5] that the extent of the consonantal influence on vowel formant frequencies varies depending on the place of articulation, manner of articulation, and voicing characteristics of the consonants. And, the ‘articulatory syllable model’ or ‘C2V model’ (n = number) proposes based on the observation that segments are co-articulated within, but not across, Cv sequences [6].

In the past decades, there have been a number of studies of co-articulation between consonant and vowel in different languages based on spectral data. Acoustically, co-articulation between initial consonant and following vowel (C-V) is commonly investigated in terms of the formant transition pattern at the C-V boundary, e.g., [3], [7]-[9]. The strength of C-V co-articulation is quantified on the basis of F2 transition measured at the vowel onset and vowel mid-point. It is reported in [7]-[9] that for Swedish voiced stop consonants [-b- d- g-] in CV syllables, the strength of co-articulation is larger for [-b-] and lesser for [-d-]. The co-articulation pattern for [-g-] is similar to that for [-d-] when preceding a front vowel, but similar to that for [-b-] when preceding the other vowel types. This is considered in relation to the fact that velar consonants commonly undergo palatalization before a front vowel as reported in [3], [7]-[11], which results in a greater resistance to co-articulation. Similar data on C-V co-articulation are also reported in languages, such as American English [10]-[13], Canadian English [14], Catalan [15], French [16], Thai, Cairene Arabic, and Urdu [17], Spanish [18], Yanyuwa and Yindjibarndi [19], and Persian [20].

In addition to C-V co-articulation, co-articulation between final consonant and preceding vowel (V-C) has been investigated [3], [12], [20]-[22] by measuring the F2 transition at the vowel offset and vowel mid-point. The co-articulation data for the final stop consonants [-b- d- g-] in the V-C context of the languages, such as Swedish [3], American English [12], and Persian [20] demonstrate different patterns of V-C co-articulation across languages and across studies. For instance, in Swedish [3], the extent of co-articulation is reduced for the final [-b-], but increases for the final [-d] and final [-g-], as compared with the initial [-b- d- g-] counterparts. In American English [12] and Persian [20], the strength of co-articulation is reduced for both [-b-] and [-d-], although the reduction is lesser for [-d-]. As for [-g-], the strength of co-articulation increases in English [12], but decreases in Persian [20].

Wai-Sum Lee is with the Department of Linguistics and Translation, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong (phone: 852-3442-6605; fax: 852-3442-0358; e-mail: w.s.lee@cityu.edu.hk). This research was supported by the International Scholarly Exchange Research Grants - Asia Pacific Region (RG003-P-14/CityU 9500023) from Chiang Ching-kuo Foundation in Taiwan (Republic of China).
Co-articulation effects have also been observed in the movements of the articulators. As reported in [23], [24] co-articulation resistance (CR) increases when the neighboring consonant and vowel compete for the articulatory targets using the same articulator(s), especially the tongue body. The articulatory data for the Catalan consonants reported in [25] show that CR in tongue position is greatest for the alveolopalatal consonants and smallest for the labial consonants, with the alveolar and velar consonants coming in between. In American English, a larger CR due to a less varied tongue body position is observed for the coronal consonants than the labial consonants [26], [27].

To our knowledge, there is a paucity of information on the co-articulation between consonant and vowel in Chinese, although there is a recent study of the co-articulation between the stop consonants [p, t, k] and neighboring vowels [i, a, u] in CVC syllables of Taiwanese [28]. The articulatory data on tongue position collected from three young adult speakers of Taiwanese show that in the C-V context, CR is larger for [i] than [a] and [u]. In the V-C context, CR is reduced and the degree of co-articulation increases for all the three Taiwanese final stops [-p -t -k] with the preceding vowels.

The present study investigates the co-articulation of the stop consonants in the syllable-initial and syllable-final position with different types of neighboring vowels in the monosyllables of Cantonese. It determines (i) both (a) the C-V co-articulation between the initial stop consonant and the following vowel and (b) the V-C co-articulation between the final stop consonant and the preceding vowel, and (ii) the degree of co-articulation resistance of the different types of syllable-initial and syllable-final stop consonants to the neighboring vowels through collection and analysis of the articulatory data on the tongue and lip positions during the Cantonese syllables.

II. METHOD

A. Test Materials

In Cantonese, the unaspirated stops [p, t, k] occur in the initial and final positions of monosyllables. Table I presents the total of 25 Cantonese monosyllables of the CV, VC or CVC structure that were used for investigation. As presented in the table, the initial or final C is any one of the three stop consonants [p, t, k] and V is any one of the three corner vowels [i, a, u]. Only meaningful monosyllables were analyzed in this study. All the test monosyllables presented in Table I are commonly used in daily communication for speech samples were collected from 10 native Cantonese speakers, five male and five female, who were born and grew up in the monolingual Cantonese-speaking families in Hong Kong with no history of speech and hearing problems. All the speakers were undergraduate students of the ages ranging from 18 to 22 years, studying at the City University of Hong Kong.

<table>
<thead>
<tr>
<th>CV</th>
<th>VC</th>
<th>CVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>[p]</td>
<td>[a]</td>
<td>[t]</td>
</tr>
<tr>
<td>[p]</td>
<td>[a]</td>
<td>[k]</td>
</tr>
<tr>
<td>[a]</td>
<td>[u]</td>
<td>[t]</td>
</tr>
<tr>
<td>[a]</td>
<td>[u]</td>
<td>[k]</td>
</tr>
</tbody>
</table>

C. Data Collection and Analysis

Recording and analysis of the test materials were performed using the research tool, Electromagnetic Midsagittal Articulography (EMA) Model AG500, by Carsten of Germany available in the Phonetics Lab at the City University of Hong Kong. EMA AG500 performs synchronized digital recordings of (i) the movements and positions of the articulators (the tongue, lips, and jaw) and (ii) the audio signals during the test syllables. For background information on the technique and details of operation and setup of EMA, see [29]-[31].

During the recording, frequency receivers (or sensors) in small coils are fixed on a speaker’s articulators, with (i) three of the sensors being fixed evenly on the tongue tip (TT), tongue middle (TM) and tongue back (TB), (ii) two at the vermilion borders of the upper and lower lips, and (iii) one on the jaw. Three additional sensors, one fixed on the nose bridge and two at the back of each ear, function as reference points of fixed positions for tracking head movement. MATLAB scripts were used to perform on the obtained raw articulatory EMA data for correction of head movement. A data visualization and analysis tool, MVIEW, was used to obtain and display the position values for the sensors attached to the articulators.

The articulatory data on (i) the tongue positions at TT, TM and TB and (ii) the positions at the upper and lower lips were extracted for the 'x, y' coordinates which indicate the 'front-back' and 'up-down' positions of the tongue and lips in the vocal tract. By making reference to the waveforms and spectrograms of the acoustic signals, the tongue and lip positions were determined at three time points – onset, mid-point and offset – of the vowels. The onset and mid-point of the vowels are for deciding the strength of C-V co-articulation between the initial stop and the following vowel, whereas the offset and mid-point of the vowels are for deciding the strength of V-C co-articulation between the final stop and the preceding vowel. A large or small difference in tongue and/or lip position between the two time points is taken to indicate a lesser or greater strength of co-articulation between the stop consonant and the neighboring vowel.

III. RESULTS

The articular data from two male Cantonese speakers were analyzed and the main findings are presented as follows. Due to the striking similarities between the data for the two speakers, only the figures for one speaker are presented in this paper for illustration purposes.

Figs. 1-6 show the superimposed mid-sagittal tongue
contours and tongue positions (i) at the vowel onset (in red dotted lines) and vowel mid-point (in blue solid lines) for the initial stops [p- t- k-] preceding the vowels [i, a, u] in CV and CVC syllables and (ii) at the vowel offset (in green dash lines) and vowel mid-point (in blue solid lines) for the final stops [-p -t -k] following the vowels [i, a, u] in VC and CVC syllables for a male speaker. The tongue contours were obtained by connecting the positions of TT, TM, and TB, with the front-back position on the x-axis against the up-down position on the y-axis. The figures with the origin at the top right corner are on the same scale (in mm). An increase in value on the x- or y-axis denotes a forward shift or a lowering of the tongue position. The thick dark line in the figures is the tracing of the mid-sagittal palate contour serving as the reference, relative to which the tongue positions in both the front-back (x) and up-down (y) dimensions during the different time points are defined and compared.

A. Bilabial Stop [p]

As shown in Fig. 1, in the syllable-initial position where the initial [p-] precedes the high vowel [i] (Fig. 1 (a)) or [u] (Fig. 1 (c)), the tongue positions at the vowel onset and vowel mid-point are nearly the same, indicating strong co-articulation between [p-] and the following vowel. This is because the articulation of [p] does not involve the tongue and the tongue is thus free to assume the gesture for the following vowel. For [p-] when preceding [i] (Fig. 1 (a)), the position of TB tends to be further raised and closer to the palate at the vowel mid-point than at the vowel onset, indicating that the tongue has yet to reach the target position at the vowel onset. Such articulatory actions are not observed when [p-] precedes [u] (Fig. 1 (c)). The difference in the strength of co-articulation with the Cantonese vowels [i] and [u] may be due to the difference in tongue position between the two vowels. A comparison of Figs. 1 (a) and (c) shows that the tongue position is slightly higher and closer to the palate during [i] than during [u]. It is thus assumed that more time is needed for the tongue to reach the target position for [i] than for [u].

As for the initial [p-] that precedes [a] (Fig. 1 (b)), both the positions of TT and TM tend to be higher at the vowel onset than at the vowel mid-point, suggesting a small strength of co-articulation between [p-] and [a]. This is attributed to the competition between the lip closure in [p] and the required degree of mouth opening for the articulation of the open vowel [a]. The measurements of the positions of the upper and lower lips show that the distance between the two lips increases at the vowel mid-point as compared to the vowel onset for the initial [p-] when preceding any vowel type. The increase in the distance between the two lips at the vowel mid-point is significantly larger for [p-] when preceding [a] (with an increase over 70%) than preceding [i] (with an increase around 30%) and [u] (with an increase below 20%). Thus, the data on the size of lip opening indicate a larger co-articulatory resistance of the initial [p-] when followed by the low vowel [a] than the high vowels [i] and [u].

In the syllable-final position, the strength of co-articulation in terms of tongue or lip gesture does not vary significantly between the final stop [-p] and the preceding vowel of any type. As shown in Fig. 2, there is a marked similarity between the tongue positions at the vowel mid-point and vowel offset for the final stop [-p] when following the high vowel [i] (Fig. 2 (a)) or low vowel [a] (Fig. 2 (b)), indicating strong co-articulation between [-p] and the preceding vowel. There are minor differences in the positions of the three tongue points between the vowel mid-point and vowel offset in each case. When [-p] follows [i] (Fig. 2 (a)), the positions of TT and TM tend to be slightly higher at the vowel mid-point than at the vowel offset. When [-p] follows [a] (Fig. 2 (b)), the positions of the all three tongue points, TT, TM, and TB, are slightly retracted at the vowel offset as compared to the vowel mid-point. The data indicate that [-p] is not in total co-articulation with the preceding vowel in terms of tongue gesture, even though the tongue is free and not involved in the articulation of [-p].

![Fig. 1](image1.png)

![Fig. 2](image2.png)

Fig. 1 (a)-(c) Superimposed mid-sagittal tongue contours and tongue positions at the vowel onset (in red dotted lines) and vowel mid-point (in blue solid lines) for the initial [p-] preceding the vowels [i, a, u] in CV and CVC syllables and palate contour (in thick dark line) for a male speaker

Fig. 2 (a), (b) Superimposed mid-sagittal tongue contours and tongue positions at the vowel offset (in green dash lines) and vowel mid-point (in blue solid lines) for the final [-p] following the vowels [i] and [a] in VC and CVC syllables and palate contour (in thick dark line) for a male speaker

With regard to the co-articulatory resistance of the lip gesture of [-p], the data show that the distance between the upper and lower lips reduces at the vowel offset as compared to the vowel mid-point. The extent of reduction in lip distance...
for [-p] when preceded by [i] (about 40%) is similar to that for [-p] when preceded by [a] (about 30%), which indicates similar co-articulatory resistance of the final [p-] to the preceding vowel of different types.

B. Alveolar Stop [t]

The strength of co-articulation between the initial [t-] and the following vowel is small, regardless of the vowel type. As shown in Fig. 3, the position of TT at the vowel onset is raised close to the anterior part of the palate compared to the lowered position of TT at the vowel mid-point. This is attributed to the articulatory resistance of [t], where TT must be raised to form the alveolar closure. When the initial [t-] is followed by the high vowel [i] (Fig. 3 (a)), the positions of TM and TB are raised close to the palate at the vowel onset, similar to the positions of TM and TB at the vowel mid-point, while TT is lowered at the vowel mid-point than at the vowel onset. When [t-] is followed by the low vowel [a] (Fig. 3 (b)), the positions of TT, TM, and TB are higher at the vowel onset than at the vowel mid-point, while the difference between the vowel onset and vowel mid-point is less pronounced in the position of TB. The data indicate that co-articulation is stronger for [t-] with [i] than [t-] with [a]. This is attributed to the raising of the tongue tip involved in the articulation of [t], which is more compatible with the tongue gesture for the articulation of the low vowel [i] than with the tongue gesture for the articulation of the high vowel [i].

In the syllable-final position, the co-articulatory resistance of [-t] is large when preceded by the back vowel [u], but not the front vowel [i] or low vowel [a]. As shown in Fig. 4 (c), when [-t] is preceded by [u], the position of TT at the vowel offset is raised toward the anterior part of the palate as compared to the position of TT at the vowel mid-point. The positions of TM and TB also differ between the two time points, being more frontward at the vowel offset than at the vowel mid-point. The data indicate weak co-articulation between the final [-t] and the preceding back vowel [u]. When the final [-t] is preceded by [i] (Fig. 4 (a)) or [a] (Fig. 4 (b)), no raise of TT at the vowel offset is observed, different from the case where [-t] is preceded by [u] (Fig. 4 (c)), which suggests the disappearance of the final stop [-t] and presumably [-t] turns into a glottal sound [-ʔ]. Furthermore, when [-t] is preceded by [i] (Fig. 4 (a)) or [a] (Fig. 4 (b)), the tongue positions at the vowel mid-point and vowel offset are almost the same, suggesting that [-t] is closely co-articulated with the preceding vowel.

C. Velar Stop [k]

For the syllable-initial velar stop [k], the co-articulation is stronger when [k-] precedes the vowel [i] or [u] than precedes the vowel [a]. When [k-] precedes the high front vowel [i] as shown in Fig. 5 (a), the positions of TT, TM, and TB at the vowel onset and vowel mid-point are nearly the same. Similarity in tongue position is also observed when [k-] precedes the high back vowel [u] (Fig. 5 (c)), although there is a tendency for the tongue to lower and retract slightly at the vowel mid-point than at the vowel onset. It is considered that a stronger co-articulation between [k-] and [i] than between [k-] and [u] is attributed to the similarity in tongue position between [k-] and [i] than between [k-] and [u] due to the palatalization of [k-] before [i].

The strength of co-articulation is reduced when [k-] precedes the low vowel [a]. As shown in Fig. 5 (b), the tongue position, in particular the position of TB, is close to the palate at the vowel onset, but distant from the palate at the vowel mid-point. The variation in the extent of co-articulation between [k-] and the following vowel of different types is attributed to the compatibility between the tongue gestures for the articulation of [k] and the neighboring vowel. During [k], TB must be raised towards the velum to form a velar closure, which is compatible with the tongue gesture for the articulation of the high vowels [i] and [u], but in competition with the tongue gesture for the articulation of the low vowel [a].

In the syllable-final position, the strength of co-articulation between the final [k-] and the preceding vowel [a] increases. As shown in Fig. 6, there is close similarity in tongue position between the vowel mid-point and vowel offset, suggesting strong co-articulation between the final [k-] and the preceding
vowel [a]. At the vowel offset, the positions of TT, TM, and TB remain distant from the palate as compared to the positions of TT, TM, and TB at the vowel mid-point, which suggests the disappearance of the final stop [-k], presumably turning into a glottal stop [-ʔ].

The degree of CR for any type of Cantonese stop is reduced in the V-C context, resulting in more similar patterns of co-articulation among the final stops [-p -t -k]. A reduction in the degree of CR is also reported during the Taiwanese final stops [-p -t -k] as compared with the initial counterparts [p- t- k-] [28]. The finding of a higher strength for V-C co-articulation than C-V co-articulation supports the phonological structure of a syllable, in that the vowel and the final consonant are the daughter constituents of the rhyme, whereas the initial consonant is a sister constituent of the rhyme in a syllable.

For further analysis, the strength of co-articulation and the degree of CR for the three types of Cantonese stop with different neighboring vowels in both C-V and V-C contexts will be quantified using the statistical formulae proposed in [25], [27] by calculating the variations in the tongue and lip positions in the Euclidean distance during a given stop across different vowel contexts. The articulatory data collected from the other speakers in this study will also be analyzed to expand the data pool. Further discussions in connection with the findings of co-articulation in other languages reported in previous studies will be made in order to reach some cross-language generalizations.

**REFERENCES**

14. T. M. Nareay and S. E. Shammass, “Formant transitions as partly distinctive invariant properties in the identification of voiced stops,”


