Temporal Organisation of Syllable Production in Cantonese Chinese

Eric Zee
City University of Hong Kong

ABSTRACT

The study is an investigation of the temporal organisation of syllable production in Cantonese Chinese. Results show (i) temporal compensation does not take place between C and V in the CV syllables, however, the duration of V is affected by the initial C type; (ii) tone plays a part in determining the temporal pattern of the syllables; and (iii) the reduction of diphthong duration or vowel+nasal sequence is contributed by the reduction of the 1st target vowel not the transition or the 2nd target vowel, or the vowel not the nasal.

INTRODUCTION

The temporal aspect of speech production has been extensively studied [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. The purpose of this study is to demonstrate certain regularities that have been found to be characteristics of the temporal structure of syllable production at the local time control level in Cantonese. By local time control, Allen [1] refers to "the specification of segment durations within syllables and possibly syllable durations within rhythmic phrases", and by global time control, he refers to "speech rate or tempo" (p. 222). The paper also shows whether temporal compensation takes place within sequences of successive component segments in the Cantonese monosyllables. Temporal compensation has been used to refer to the cases where the segment durations are inversely correlated [8], [9], [10] as an effort by the speaker to maintain an invariant syllable duration [7] or macrostructural invariance [16]. An inverse or negative correlation between the durations of two successive segments implies that "articulatory events are programmed at some higher level not in terms of single phonemes, but in terms of high-level articulatory units" and temporal compensation occurs when the durations of segments in a sequence are negatively correlated or the production of the segments takes place in terms of phonological units [9].

In Cantonese, four types of syllable structure are distinguished: $(C_1)V$:, $(C_1)D$, $(C_1)V:C_2$, and $(C_1)VC_2$, where C_1 may be an aspirated or unaspirated stop, an aspirated or unaspirated affricate, a fricative, a nasal, a liquid, or a glide; D a diphthong; V a short vowel; V: a long vowel; and C_2 an "-N", i.e., a syllable-final nasal [-m, -n, -ŋ] or "-S", i.e., a syllable-final unreleased stop [-p, -t, -k]. In Cantonese two types of diphthongs, D_1 and D_2, may be distinguished in terms of their internal temporal structure. D_1 refers to the diphthongs, such as [ui, iu, ai, au, ɔi], and D_2 to the diphthongs, such as [ei, ɐi, ɐu, ou, ɵy]. In Cantonese, six long citation tones, ˥, ˧˥, ˧, ˨˩, ˩˧, ˨, or 55, 33, 22, 21, 35, 24, respectively, occur on the open syllables or syllables with a nasal ending, $(C_1)V$:, $(C_1)D$, $(C_1)V:N$, and $(C_1)VN$. Short variants of the tones, ˥, ˧˥, ˧, i.e., 5, 3, 2, realise on the syllables with a stop ending, such as $(C_1)V:S$ and $(C_1)VS$.

The present study investigates the temporal structures in syllables of different types associated with different tones in Cantonese. Due to page limit, the only temporal data to be discussed in this paper are those of syllables, CV:, CV:S, CVS, CD, and CV:N, where V: = [a], V = [ɐ], D = [ai], and N = [ŋ]. More specifically, the paper shows (1) whether temporal compensation takes place between a syllable-initial consonant, C, and the post-consonantal -V:, -V:S, -VS, -D, and V:N in CV:, CV:S, CVS, CD, and CV:N syllables; (2) whether temporal compensation takes place among the first target vowel, the transition, and the second target vowel in D, i.e., the diphthongs; (3) and whether tone plays a part in determining the temporal structure of the syllables under question.

METHOD

In this study, meaningful Cantonese monosyllables of the types $(C_1)V$:, $(C_1)D$, $(C_1)V:C_2$, and $(C_1)VC_2$, associated with different tones, such as ˥, ˧˥, ˧, ˨˩, ˩˧, and ˨ (i.e., 55/5, 33/3, 22/2, 21, 35, and 24, respectively), were used as test syllables: [pa55], [pa33], [pa22], [pa21], [pa35]; [pʰa55], [pʰa33], [pʰa21], [pʰa35]; [tsa55], [tsa33], [tsa22], [tsa35], [tsa21]; [tsʰa55], [tsʰa33], [tsʰa24], [tsʰa21]; [sa55], [sa33], [sa35], [sa21]; [pak5], [pak3], [pak2]; [pʰak5], [pʰak3], [tsak5], [tsak3], [tsak2]; [tsʰak5], [tsʰak3], [tsʰak2]; [sap5], [sap3], [sap2]; [pɐt5], [pɐt2], [pʰɐt5]; [tsɐt5], [tsɐt2], [tsʰɐt5]; [sɐt5], [sɐt2]. Temporal organization of the test syllables is determined by measurements of the durations of the successive component segments contained in the syllables. Using Kay Elemetrics' CSL 4300B speech analysis software on a 486 PC, durations of the successive component segments in the test syllables were measured directly from the speech waveforms. In the cases where the test syllables contain a diphthong or a nasal ending, durations were measured from the speech waveforms as well as from their formant trajectories. Two native Hong Kong Cantonese college students, one male and one female, provided the speech data. The speakers were instructed to utter each of the test monosyllables in a carrier sentence, [ŋɔ jiu tɔk _____ pei lei tʰɐŋ] "I want (to) read _____ for you (to) listen", at normal rate of speech. The test words in the word-lists were randomised. Five repetitions of each of the test syllables were recorded. The recording was performed in a sound-proof booth.

RESULTS

The mean vowel or diphthong durations (n = 5) in CV: or CD syllables, where C is [p], [pʰ], [ts], [tsʰ], or [s], tend to be similar if the tone on the syllable is 55, 33, 22, 35, or 24; and the mean vowel durations (n = 5) in CV: or CD syllables, where C is [pʰ], [ts], [tsʰ], or [s], tend to be similar if the tone on the syllable is 21.

The overall mean vowel [a] duration in all the CV: syllables, where C is [p], [pʰ], [ts], [tsʰ], or [s], is longer (329.16 ms (n = 85; s.d. = 22.18)) if the tone on the syllable is 55, 33, 22, 35, or 24, and shorter (247.00 ms (n = 25; s.d. = 23.63)) if the tone on the syllable is 21. This is also true for CD syllables. The overall mean diphthong [ai] duration in all the CD syllables associated with tone 21 is 274.30 ms (n = 10, s.d.= 13.44) and the overall mean diphthong duration associated with tones 55, 33, 22, and 35 is 348.07 ms (n = 75, s.d. = 24.45). Such a correlation between duration and tone type is not present in CV:S and CVS syllables, due probably to the fact that the tone on these syllables does not reach the level of 1.

Regardless of the tone type on the CV:S or CVS syllables, the mean vowel [a] or [ɐ] durations in these syllables tend to be similar. The overall mean vowel [a] duration in all the CV:S syllables asso-ciated with tones 5, 3, and 2 is 186.61 ms (n = 70, s.d. = 20.07). The overall mean vowel [ɐ] duration in all the CVS syllables associated with tones 5 and 2 is 120.03 ms (n = 40, s.d. = 10.57).

Regardless of the tone type on the syllables, the mean vowel or diphthong durations in CV:, CD, CV:S, and CVS syllables are shorter if C is an aspirated stop or affricate; or longer if C is an unaspirated counterpart or a fricative. The overall mean vowel duration of all the CV: syllables, associated with tones 55, 33, 22, 24, and 35, is 344.57 ms (n = 35, s.d. = 18.49) if C is an unaspirated [p] or [ts]; 312.87 ms (n = 30, s.d. = 20.03) if C is an aspirated [pʰ] or [tsʰ]; and 344.47 ms (n = 15, s.d. = 18.42) if C is [s]. The overall mean diphthong duration of all the CD syllables, associated wtih tones 55, 33, 22, and 35 is 363.66 ms (n = 35, s.d. = 18.59) if C is an unaspirated [p] and [ts]; 328.04 ms (n = 25, s.d. = 19.04) if C is an aspirated [pʰ] or [tsʰ]; and 346.73 ms (n = 15, s.d. = 20.07) if C is [s]. The overall mean vowel duration of all the CV:S syllables, associated with tones 5, 3, and 2, is 202.63 ms (n = 30, s.d. = 14.72) if C is an unaspirated [p] or [ts]; 169.56 ms (n = 25, s.d. = 15.65) if C is an aspirated

[p^h] or [ts^h]; and 182.53 (n = 15, s.d. = 7.51) if C is [s]. The overall mean vowel duration of all the CVS syllables, associated with tones 5 and 2, is 126.85 ms (n = 20, s.d. = 5.28) if C is an unaspirated [p] or [ts]; 111.00 ms (n = 10, s.d. = 12.55) if C is an aspirated [p^h] or [ts^h]; and 115.40 ms (n = 10, s.d. = 7.07) if C is a fricative. Similar results are obtained for CV:N syllables.

The reduction of the total diphthong durations in CD syllables which are associated with 55, 33, 22, and 35 tones and when initial C is an aspirated stop or affricate is contributed mainly by the reduction of the duration of the first target vowel [a], not the transition or the second target vowel in the diphthong. The mean duration of the first target vowel [a] in the CD syllables, associated with tones 55, 33, 22, and 35 is 197.34 ms (n = 35, s.d. = 14.98) if C is an unaspirated [p] or [ts], and 158.60 ms (n = 25, s.d. = 13.10) if C is an aspirated [p^h] or [ts^h]. The mean duration of the transition in the CD syllables, associated with tones 55, 33, 22, and 35 is 130.77 ms (n = 35, s.d. = 12.45) if C is [p] or [ts], and 135.36 (n = 25, s.d. = 11.88) if C is [p^h] or [ts^h]. And, the mean duration of the second target vowel [i] in the CD syllables, associated with tones 55, 33, 22, and 35 is 35.54 ms (n = 35, s.d. = 12.20) if C is [p] or [ts], and 34.08 ms (n = 25, s.d. = 11.57) if C is [p^h] or [ts^h].

The further reduction of the total diphthong duration in CD syllables, as a result of the effect the 21 tone, where C is an aspirated stop or affricate is contributed by the reductions of durations of the first target vowel and the transition, or the second target vowel in the diphthong.

As for the temporal structure in the diphthongs in those CD syllables, which are associated with tones 55, 33, and 35 and where C is a fricative, [s], the relatively smaller reduction of the total diphthong duration is contributed by both the reductions of durations of the first and second target vowels.

There is a tendency for the syllable-initial aspirated stops and affricates in CV: and CD syllables to have a slight increase in duration if the tone on the syllable is 21, as compared with the durations associated with a non-21 tone. The mean duration of [p^h] in [p^ha] associated with a 21 tone is 93.00 ms (n = 5, s.d. = 13.06), whereas the mean durations of [p^h] in [p^ha] associated with tones 55, 33, and 35 are 69.80 ms (n = 5, s.d. = 5.07), 73.21 ms (n = 5, s.d. = 9.05), and 85.20 ms (n = 5, s.d. = 5.02), respectively. The mean duration of the [ts^h] in [ts^ha] associated with a 21 tone is 140.40 ms (n = 5, s.d. = 20.07), whereas the mean durations of [ts^h] in [ts^ha] associated with tones 55, 33, and 24 are 113.40 ms (n = 5, s.d. = 12.22), 133.00 ms (n = 5, s.d. = 22.66), and 125.00 ms (n = 5, s.d. = 10.40), respectively. Similar results are obtained for CD, CV:S, CVS and CV:N syllables.

The syllable-initial consonant [s] in CV:, CD, CV:S, and CVS syllables has the longest duration, followed by [ts^h], [p^h], [ts], and [p] in descending duration, for example, in CV: syllables which are associated with a 55 tone, the mean durations (n = 5) of [s], [ts^h], [p^h], [ts], and [p] are 144.80 ms (s.d. = 12.44), 113.40 ms (s.d. = 12.22), 69.80 ms (s.d. = 5.07), 54.00 ms (s.d. = 12.69), and 8.40 ms (s.d. = 2.51), respectively

CONCLUSION

Temporal compensation does not seem to take place between syllable-initial consonant and the post-consonantal vowel or diphthong in the Cantonese CV:, CD, CV:S, and CVS syllables. The observation is evidenced by the fact that invariant syllable duration in Cantonese is not maintained, for example, the mean durations (n = 5) of the syllable-initial consonants [p], [ts], [ts^h], and [s] in CV: syllables associated with a 33 tone are 8.40 ms (s.d. = 2.30), 52.80 ms (s.d. = 10.62), 133.00 ms (s.d. = 22.66), and 150.60 ms (s.d. = 14.05), respectively, whereas the mean durations (n = 5) of the following vowel [a] in the syllables are 338.20 ms (s.d. = 14.91), 350.40 ms (s.d. = 18.48), 348.40 ms (9.61), and 353.80 ms (s.d. = 23.78)

The reduction of the vowel or diphthong duration in the syllables where the initial consonant is aspirated is not viewed as an effort of the speaker to maintain invariant syllable duration, rather a result of reduced subglottal pressure caused by the production of aspiration of the syllable-initial consonant.

The further reduction of the duration of the vowel or diphthong in CV: and CD syllables which are associated with a 21 tone and where C is an aspirated stop or affricate is assumed to be contributed by both the reduced subglottal pressure and the 21 tone. The duration of the vowel or diphthong in CV: syllables which are associated with a 21 tone is shorter than the duration associated with a non-21 tone even if the syllable-initial is zero or an unaspirated stop or affricate. This shows that tone does play a part in determining the temporal structure of the syllables. That the vowel or diphthong duration in the CV:, CD, CV:S, CVS, and CV:N syllables where the initial consonant is a fricative, [s], is not reduced or only slightly reduced is probably due to the fact that the production of [s] does not reduce the subglottal pressure for the production of the post-consonantal vowel or diphthong as much as an aspirated consonant does. It seems that temporal compensation takes place among the first target vowel, the transition, and the second target vowel in some syllables, as the total mean durations of the syllables [pai˥] and [pai˦] as well as [pai˧] and [pai˨] are almost the same, although the mean durations of the first target vowel, transition, and the second target vowel differ. The durational data by the female Cantonese speaker have not been presented, although the pattern of the temporal organisation is similar.

REFERENCES

[1] Allen, G. (1973), Segmental timing control in speech production. *Journal of Phonetics*, 1: 219-237.

[2] Bell-Berti, F. & K.S. Harris (1981), A temporal model of speech production. *Phonetica*, 38:9-20.

[3] Cao, J. F. (1992), Temporal distribution of bisyllabic words in Standard Chinese: an evidence for relational invariance and variability for natural speech. *Communication of COLIPS*, 2:58-66.

[4] Crystal, T.H. and A.S. House (1988a), The duration of American-English vowels: an overview. *Journal of Phonetics*, 16: 263-284.

[5] Crystal, T.H. and A.S. House (1988b), The duration of American-English stop consonants: an overview. *Journal of Phonetics*, 16: 285-294.

[6] Fowler, C.A. (1977), Timing Control in Speech Production. Bloomington, Indiana University Linguistics Club.

[7] Fowler, C.A. (1981), A relationship between co-articulation and compensatory shortening. *Phonetica*, 38:35-50.

[8] Kozhevnikov, V.A. & Chistovich, L.A. (1965), Speech: Articulation and Perception. Washington, DC., U.S. Department of Commerce, Joint Publications Research Service, No. 30: 543.

[9] Lehiste, I. (1971), Temporal organisation of spoken language. In L.L. Hammerich, et al, eds., *Form & Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, pp. 159-169.

[10] Lehiste, I. (1972), The timing of utterances and linguistic boundaries. *Journal of Acoustical Society of America*, 51:2018-2024.

[11] Lindblom, B.E.F. (1968), Temporal organization of syllable production. *STL-QPSR* 2-3, 1968, 15.

[12] Lindblom, B.E.F. & K. Rapp (1973), Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm*, 21:1-59.

[13] Nooteboom, S.G, (1972), *Production and Perception of Vowel Duration*. Utrecht, Utrecht University Dissertation.

[14] Peterson, G. & E. Lehiste (1960). Duration of syllable nuclei in English. *Journal of Acoustical Society of America*, 32:693-703

[15] [*Phonetica*, 1981]

[16] Port, R.F., S. Al-Ani, & S. Maeda (1980), Temporal compensation and universal phonetics. *Phonetica*, 37: 235-252. [End]