The Effect of Anticipatory Coariticulation in VCV Sequences in Chinese

Maolin Wang, Xuejiao Huang

College of Chinese Language and Culture, Jinan University, Guangzhou, China

Abstract: *In this study, the anticipatory Vowel-to-Vowel (V-to-V) coarticulatory effect in the Vowel-Consonant-Vowel (VCV) sequences is investigated. The subjects are twelve native speakers of standard Chinese, and the F_2 offset value of the first vowel is analyzed. Results show that, in the trans-segment context, anticipatory coarticulation exists in Chinese. The coarticulatory effect in the context of labial is greater than that of alveolar, which is in line with the degrees of articulatory constraint model. The articulatory strength is great for aspirated consonants, so the coarticulatory effect is great in the context of aspirated stops.*

Keywords: *Aspiration, coarticulation, formant, place of articulation, vowel*

I. INTRODUCTION

This study deals with the acoustic signaling of coarticulation, which refers to the articulatory modification of a speech sound under the influence of adjacent segments. For example, all else being equal, the back vowel [u] in 'two' is produced farther forward than the same vowel in 'who' due to the influence of the adjacent coronal consonant. Coarticulation effect may vary with its specific context or the phonological system of a language. In his classic spectrographic study of VCV sequences in three languages, Öhman [1] found that F_2 values of target vowels varied more due to vowel context in English and Swedish than in Russian. He attributed the coarticulatory differences to the languages' consonant systems, arguing that the requirements on the tongue body imposed by contrastive palatalization in Russian, but not in English or Swedish, restricted transconsonantal coarticulation in Russian.

Consonant restrictions on V-to-V coarticulation have also been reported by Recasens [2], who found less V-to-V coarticulation across the velarized lateral of Catalan than across the 'clear' lateral of Spanish and German. He and his colleagues ascribed the coarticulatory differences to different lingual constraints for these laterals. Bladon and Al-Bamerni [3] originated the concept of 'coarticulatory resistance' that claimed phonetic segments possess inherent properties that limit the extent to which they can be influenced by neighboring segments. Using this concept within a coarticulatory approach to speech production, Recasens [4] developed the 'degrees of articulatory constraint' (DAC) model to account for coarticulatory effects of both vowels and consonants. Recasens' model predicts that the more a specific region of the tongue is involved in the occlusion for the C, the more the C affects V, but the less it can be shaped by the vowel, and the less the transconsonantal V-to-V coarticulation.

There have been a number of studies on the coarticulatory effect of segments in Chinese, including the analysis of the acoustic coarticulatory patterns of voiceless fricatives in CVCV [5], the study of vowel formant pattern and the coarticulation in the voiceless stop initial monosyllables [6], the acoustic study of intersyllabic anticipatory coarticulation of three places of ariculation of C_2 in CVCV [7], vowel segmental coarticulation in read speech in Standard Chinese [8], and anticipatory coarticulation in V1#C2V2 sequences [9]. It is found that coarticulation exists in segment adjacent and trans-segmental contexts in Chinese.

Coarticulation is a common phenomenon in languages, and it is believed that coarticulation affects the smoothness and naturalness of the synthesized speech in Text-to-Speech. Therefore, the naturalness of synthesized speech will be greatly improved if speech coarticulation is properly solved [10]. The research presented in this paper aims to investigate the V-to-V coarticulation in VCV sequences in Chinese. Coarticulation may be generally classified as carry-over (left-to-right) or anticipatory (right-to-left) ones [11], and the present study will focus on anticipatory coarticulation.

II. METHODOLOGY

2.1 Speakers, stimuli and recording

Twelve native speakers of Standard Chinese, six male and six female, participated in the recording. Regarding the stimuli, disyllabic words, in the form of $C_1V_1C_2V_2$, are used, with V_2 providing the 'changing' vowel context, V_1 the 'fixed' vowel, which is designed for the changing vowels to affect the fixed vowel. The fixed vowel is /a/, and for the changing vowel context, vowels /i/ vs. /u/ are used to influence the offset of the F_2 frequencies of the fixed vowel. The intervocalic consonant C_2 includes /b, p, d, t/, two unaspirate stops /b/, /d/ and two aspirated ones /p/, /t/. All the words used are in normal stress, without neutral tone syllables. An example of a

pair of words used are 'dadi' and 'dadu', which mean 'archenemy' and 'to bet' respectively. Two sets of words of identical combinations are used, so there are 16 words in the word list (4 stops × 2 changing vowel contexts × 2 sets).

Recording was done in a sound-treated room, and the acoustic data were recorded directly into the computer at a sampling rate of 16 kHz using the recording software of Cool Edit Pro. The speakers were asked to read the word list three times, in random order for each repetition, in normal pace, so each speaker produced 48 tokens: 16 words × 3 repetitions. In total, 576 tokens were acoustically analyzed (48 tokens × 12 speakers).

2.2. Procedure and measurements

1) F_2 offset value: This study aims at investigating the extent of V-V coarticulation in VCV sequences, and vowel formant is examined. Formant values are extracted using Praat [12], and the effect of trans-consonantal coarticulation is analyzed by measuring the F_2 offset value of the fixed vowel. F_2 offset frequency is taken at the offset of the fixed vowel V_1.

Fig. 1 displays the waveform and spectrogram of 'dadi' (archenemy), with C_2 as an unaspirated, alveolar consonant. For the purpose of this study, F_2 offset value is taken at the offset point of the vowel, that is, point 'A' on the graph.
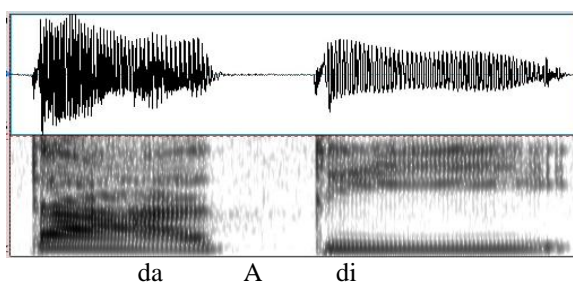

da　　A　　di

Fig. 1 Waveform and spectrogram of 'dadi'

2) F_2 delta: In order to compare the extent of coarticulatory effects under various consonant contexts, besides the F_2 offset values, their differences caused by the changing V_2 contexts are also calculated. Coarticulation effects due to changing V_2 contexts are indexed by F_2 delta values obtained at the V_1 F_2 offset, and F_2 delta (Hz) is derived by computing the difference in offset frequencies of the fixed vowel in each sequence pair, as is shown in formula (1)

$$\Delta F_2 = F_{2i} - F_{2u} \qquad (1)$$

In (1), F_{2i} and F_{2u} refer to the V_1 F_2 offset values preceding vowel /i/ and /u/ respectively, and ΔF_2 is the F_2 delta at the offset of V_1.

Fig. 2 displays the F_2 contours of the sequence pair 'dadi' (archenemy) and 'dadu' (to bet), with the contour of 'dadi' in solid line, and that of 'dadu' in dashed line. In this sequence pair, for the changing vowel context, F_2 of /i/ is high and that of /u/ is low. If V_1 F_2 offsets differ in this pair, then it is reasonable to attribute the frequency difference to the high vs. low F_2 contexts in V_2. The greater the F_2 delta value is, the greater the coarticulatory effect of V_2 is on V_1.

Fig. 2 F_2 contours of the sequence pair 'dadi' and 'dadu', with the former in solid line, and the latter in dashed line

A repeated measures ANOVA was performed with two within-subjects factors— aspiration (unaspirated, aspirated) and place of articulation (labial, alveolar).
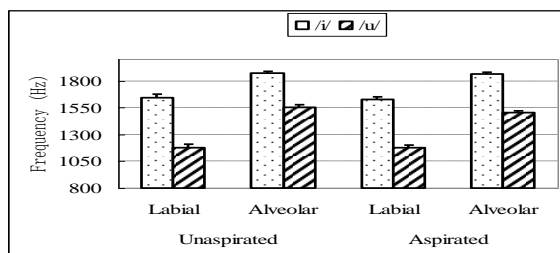
III.　RESULTS

3.1 F_2 value

Fig. 3 graphs the F_2 offset values for male speakers (Fig. 3a) and female speakers (Fig. 3b), broken down by the contexts of aspiration, place of articulation and changing vowels. The changing vowel contexts are indicated by /i/ and /u/, which refer to changing vowel context of /i/ and /u/ respectively. Repeated measures ANOVA results show that, as far as main effect is concerned, there are significant effects for all the three factors: place of articulation: $F(1, 71) = 1084$, $p < 0.001$; aspiration: $F(1, 71) = 7.39$, $p = 0.008$; changing vowel

context: $F(1, 71) = 726$, $p < 0.001$. The V_1 F_2 offset values are comparatively great in the context of alveolars, unaspirated consonants, and preceding vowel /i/.
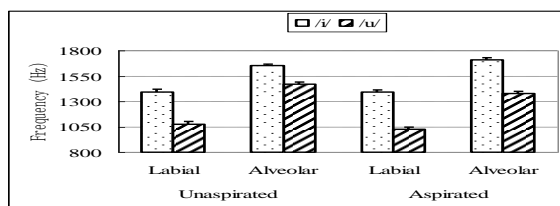
(a) F_2 values of male speakers

(b) F_2 values of female speakers

Fig. 3 F_2 offset values for male speakers (a) and female speakers (b), broken down by the contexts of aspiration, place of articulation and changing vowel

For the purpose of elaborating the coarticulatory effect under various consonant contexts in detail, F_2 delta value will be analyzed in the next section. To be specific, the extent under the contexts of place of articulation and aspiration will be presented.

3.2 F_2 delta value

Table 1 presents the F_2 delta means and significance results for the main effects. From Table 1 it can be seen that, in terms of overall main effect, there is significant effect for both place of articulation and aspiration, with the effect in the labial contexts greater than that of alveolar, and the effect of aspirated stop contexts greater than that of the unaspirated ones.

Table 1 F_2 delta means (in Hz) and statistical results for the main effects

			Mean	Statistical result
Place of articulation		Labial	347.9	$F(1, 71) = 22.0$, $p < 0.001$
		Alveolar	275.6	
Aspiration		Unaspirated	281.7	$F(1, 71) = 16.3$, $p < 0.001$
		Aspirated	341.6	

When interactive effects are examined, it is shown that the place of articulation × aspiration interaction is significant: $F(1, 71) = 13.3$, $p = 0.001$, which is attributable to the inconsistent effect of place of articulation under different contexts of aspiration. In the next subsection the effects of the factors will be described in detail to help inform and elaborate on the main effects.

3.2.1 The effect of place of articulation:
1) The unaspirated stop contexts

Fig. 4 shows the F_2 delta under the effects of place of articulation and aspiration. Result from repeated measures ANOVA shows that, in the unaspirated stop contexts, the effect of place of articulation is significant: $F(1, 71) = 50.6$, $p < 0.001$, with the extent of labial context exceeding that of alveolar context.
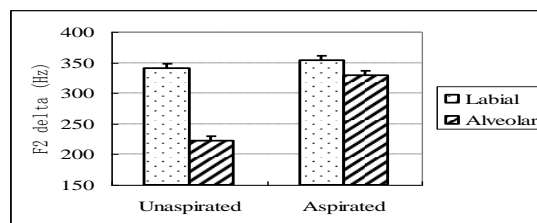
Fig. 4 F_2 delta under the effects of place of articulation and aspiration

2) The aspirated stop contexts

When the intervocalic stops are aspirated, there are no significant difference between the two place of articulation contexts: $F(1, 71) = 1.24$, $p = 0.27$. That is, there is no significant difference in the coarticulatory effect of the contexts of place of articulation in the case of aspirated stop context.

3.2.2 The effect of aspiration:
1) The context of labials

It is shown from repeated measures ANOVA result that, in the context of labials, there is no significant difference between the unaspirated and aspirated stop contexts: $F(1, 71) = 0.64$, $p = 0.428$. In this context, there is no significant effect of aspiration on the trans-consonantal coarticulation.

2) The context of alveolars: In the context of alveolar stops, the difference between unaspirated and aspirated stop contexts is significant: $F(1, 71) = 22.8$, $p<0.001$. The aspirated stop context exceeds the unaspirated one in the extent of coarticulation.

IV.   DISCUSSION

Analysis in the previous section shows that, when V_1 F_2 offset value is analyzed, the effect of the changing vowel context on it is significant, with F_2 offset value preceding vowel /i/ higher than that preceding /u/. This study aims at investigating the anticipatory coarticulatory effect of V_2 on V_1, with the changing vowels of V_2 as /i/ and /u/. As is mentioned above, the F_2 value of /i/ is high, while that of /u/ is low. The V_1 F_2 offset values are significantly different when preceding /i/ and /u/. This implies that trans-consonantal anticipatory vowel to vowel effect exists in Chinese.

In regard to the F_2 delta values, when main effects are examined, there is significant effect for place of articulation, with the effect in the labial contexts greater than that of alveolar. The DAC model [11] predicts that in VCV sequences, an increase in the degree of constraint for the consonant should yield an increase in the prominence of the C-to-V effects and a decrease in the strength of the V-to-C and V-to-V effect. According to the DAC model, the diversity in the involvement of the articulators in the production of different obstruents results in the variation of the degree of articulatory constraint. Obstruents, particularly alveolopalatals, that maximally engage the tongue dorsum for the occlusion gesture would reduce V effects, that is, stops like /d/ and /t/ exhibit reduced extents of V-V coarticulation.

With respect to coarticulation, coarticulatory sensitivity, which is the magnitude and temporal extent of the coarticulatory effect at a given articulator, is shown to be inversely related to the degree of articulatory constraint: highly constrained phonetic segments are generally more resistant to coarticulation than those specified for a lower degree of articulatory constraint, and thus less sensitive to coarticulatory influence from neighboring segments. The model also predicts that coarticulatory dominance is positively related to the degree of articulatory constraint: phonetic segments with high DAC value and coarticulation resistant usually have prominent coarticulatory effects on neighboring phonetic segments.

In this study, it is shown that the coarticulatory pattern is consistent to the DAC model: as far as main effect is concerned, the coarticulatory effect is greater in the context of labials than that of alveolar. The coarticulatory resistant for alveolars is comparatively great, which is in line with the DAC model.

As for the effect of aspiration, analysis in the previous section shows a general tendency for coarticulation effect to be greater for aspirated stops than the unaspirated ones, which is due to the high articulatory strength of aspirated obstruents. Generally speaking, phonetic segments can be classified into two groups: 'fortis' and 'lenis', which refer to consonants that are produced with greater and lesser energy respectively, such as in energy applied, articulation, etc. Fortis and lenis were coined as less misleading terms to refer to consonantal contrasts in languages that do not employ actual vocal fold vibration in their voiced consonants, but instead involved amounts of articulatory strength. For example, in English there are fortis consonants, as in 'come' and 'put' that exhibit a longer stop closure and shorter preceding vowels than their lenis counterparts, as in 'grass' and 'bed'. In Chinese, aspirated consonants are fortis, while unaspirated ones are lenis.

As the aspirated consonants in Chinese is fortis, their articulatory strength is great. Generally speaking, consonants with high articulatory strength tend to exert great effect on the preceding vowels. In Chinese, consonants and vowels are combined into one unit: syllable. Syllables with consonants of high articulatory strength may exert great effect on the preceding vowel. As a result, the anticipatory coarticulatory effect is great when C_2 is aspirated consonant.

Regarding the effect of place of articulation, results from the previous section show that, in the context of unaspirated consonants, coarticulatory effect in the trans-labial context is greater than that in the trans-alveolar context, which is in accordance to the DAC model. As for the effect of aspiration, when C_2 is alveolar, coarticulatory effect is comparatively great in the context of aspirated consonant, which is due to the high articulatory strength of the aspirated consonants.

However, when C_2 is aspirated, there is no significant effect of place of articulation on coarticulation. The cause for this result comes from the effect of aspiration on the V_1 F_2 offset values. In section 3.1, Fig. 3 displays the V_1 F_2 offset values, from which it is shown that V_1 F_2 offset value is comparatively small when C_2 is aspirated. Further analysis shows that when V_2 is vowel /i/, there is no significant effect of aspiration on the V_1 F_2 offset values, labial: $F(1, 71) = 0.47$, $p = 0.494$; alveolar: $F(1, 71) = 2.05$, $p = 0.157$. Only when V_2 is /u/, there is significant effect of aspiration on the V_1 F_2 offset values, labial: $F(1, 71) = 6.03$, $p = 0.016$; alveolar: $F(1, 71) = 16.4$, $p < 0.001$. The V_1 F_2 offset value is comparatively great in the context of unaspirated consonants. That is to say, when V_2 is /u/, the V_1 F_2 offset value is reduced in the context of aspirated consonants.

The significant level of the effects are diverse under different contexts of place of articulation: for labial, it is comparatively low, with $p = 0.016$; for alveolar, it is high, with $p < 0.001$. The variability of V_1 F_2 offset value is great in the context of alveolar. The F_2 delta is the difference of V_1 F_2 offset values between the contexts of subsequent vowels /i/ and /u/. In the context of alveolar, the V_1 F_2 offset value is reduced when preceding vowel /u/, so the F_2 delta will increase. As a result, in the context of aspirated consonants, there is no significant difference between the contexts of alveolars and labials.

It is also shown in the previous section that, when C_2 is labial, there in no significant effect of aspiration on the extent of coarticulation. As is mentioned above, the articulatory strength of aspirated consonant is high, but as the degree of articulatory constrait of labial is low, the effect of aspiration for labials on the V_1 F_2 offset value is small. When V_2 is /i/, there is no significant effect of aspiration on the V_1 F_2 offset value, while when V_2 is /u/, the significant level for the effect of aspiration on the V_1 F_2 offset value in the context of labial is low. As a result, when C_2 is labial, there in no significant effect of aspiration on the extent of coarticulation.

V.   CONCLUSION

In this study, the V-to-V coarticulation in the VCV sequences is investigated, and it is found that there is significant difference between the V_1 F_2 offset values in the contexts of subsequent vowels /i/ and /u/, which means that trans-segmental anticipatory coarticulation exists in Chinese. As far as main effect is concerned, coarticulatory effect is greater in the context of labial than of alveolar, which is consistent to the DAC model. As the articulatory strength of aspirated consonant is high, the effect is great when C_2 is aspirated consonant. The V_1 F_2 offset value will be reduced in the context of aspirated consonants when preceding vowel /u/, and this effect intensifies when C_2 is alveolar consonant. Therefore, in the context of aspirated consonants, there is no significant difference between the two place of articulation contexts. The effect get when C_2 is labial, so in this case there is no effect of aspiration on the extent of coarticulation.

This study is significant in speech engineering. In speech synthesis, the effect of trans-consonantal coarticulation must be taken into consideration. The extent of coarticulation in the context of labial exceeds that of alveolar, and the extent of aspirated consonants exceeds that of unaspirated ones, so much attention should be paid in these contexts. However, in some case, the difference of coarticulatory effects can be neglected, as it disappears. Therefore, this study is helpful in speech engineering technology.

VI.   ACKNOWLEDGEMENTS

REFERENCES

[1]. S. E. G. Öhman, "Coarticulation in VCV utterances: spectrographic measurements," Journal of the Acoustical Society of America, vol. 39, pp. 151–168, 1966.

[2]. D. Recasens, "An acoustic analysis of V-to-C and V-to-V coarticulatory effets in Catalan and Spanish VCV sequences," Journal of Phonetics, vol. 15, pp. 299–321,1987.

[3]. R. A. W. Bladon and A. Al-Bamerni, "Coarticulation resistance in English /l/," Journal of Phonetics, vol. 4, pp. 137–150,1976.

[4]. D. Recasens, "V-to-C coarticulation in Catalan VCV sequences," Journal of Phonetics, vol. 12, pp.61–73, 1984.

[5]. Z. Wu and G. Sun, "Acoustic coarticulatory patterns of voicelese fricatives in CVCV in Standard Chinese", Report on Phonetic Research, Institute of Linguistics, CASS, 1990, pp. 1–19.

[6]. J. Yan, "A study of the vowel formant pattern and the coarticulation in the voiceless stop initial monosyllable of Standard Chinese", Report on Phonetic Research, Institute of Linguistics, CASS, 1990, pp. 30–54.

[7]. X. Chen, "A study on the coarticulation of segments in Chinese, Chinese Linguistics", vol. 5, pp. 345–350, 1997.

[8]. G. Sun, "Vowel segmental coarticulation in read speech in Standard Chinese", Report on Phonetic Research, Institute of Linguistics, Chinese Academy of Social Sciences, 1998.

[9]. Y. Li and J. Kong, "Anticipatory coarticulation in V1#C2V2 sequences in standard Chinese," *Journal of Tsinghua University*, Vol. 51, No. 9, pp 1220-1225, 2011.

[10]. M. Wang, P. Chai and M. Xie, "Research on the Coarticulation of Two-syllable Word for Mandarin Speech Synthesis," Computer Engineering, vol. 28, no. 5, pp. 182-184, 2002.

[11]. G. Modarresi, H. Sussman, B. Lindblom and E. Burlingame, "An acoustic analysis of the bidirectionality of coarticulation in VCV utterances," Journal of Phonetics, vol. 32, pp. 291–312, 2004.

[12]. P. Boersma, "Praat, a system for doing phonetics by computer", Glot International, 5:9/10, pp. 341–345, 2001.

[13]. D. Recasens, "An EMA study of VCV coarticulatory direction," Journal of the Acoustical Society of America, vol. 111, pp. 2828–2841, 2002.