The Coarticulatory Effect of Bi-Syllabic Words in Chinese

Maolin Wang and Shengnan Xiong

Abstract—In this study, trans-consonantal vowel-to-vowel coarticulation in Chinese is investigated. The target words are in the form of ‘bV1ba’, and the subjects are eight native speakers of standard Chinese. Vowel formants are examined at the onset, middle and offset points of the target vowel. Results show that trans-segmental coarticulation exists in Chinese, especially at the onset point of the target vowel. Coarticulation is more likely to occur on F2, and in Chinese, coarticulatory effect does not extend to the offset point of the vowel.

Index Terms—Speech, coarticulation, vowel, formant.

I. INTRODUCTION

Coarticulation refers to the fact that speech sounds are not produced as isolated gestures, but are superimposed in a complex, context-dependent fashion. While it is clear that all natural speech is coarticulated, the magnitude and temporal extent of coarticulation in different contexts has been difficult to explain. The need to adequately describe the varieties of coarticulation has been a key factor driving the development of competing models of speech production and also perception [1].

Researches on speech production reveal systematic differences between languages in the spatiotemporal characteristics of coarticulation [2]. Öhman [3] compared the coarticulatory effects in three languages, and found that the F2 values of target vowels varied more in English and Swedish than in Russian, due to vowel context. He attributed the coarticulatory differences to the languages’ consonant systems, arguing that the requirements on the tongue body imposed by contrastive palatalization in Russian restricted trans-consonantal coarticulation.

Following Öhman’s [3] ground breaking work on transconsonantal vowel-to-vowel coarticulation, researchers have sought to understand how different factors influence this effect, and it is shown that a language’s system of vowel contrasts may influence V-to-V effect. Beddor et al. [4] conducted three experiments to test the hypothesis that V-to-V coarticulatory organization differs in Shona and English. An acoustic study of Shona and English trisyllables shows that the two languages differ in the coarticulatory effect of stressed and unstressed vowels on each other, and the relation between the production and perception data suggests that listeners are attuned to native-language coarticulatory patterns.

It is shown from research results that languages with larger vowel systems tend to exhibit weaker V-to-V coarticulatory effects than those with smaller systems. Weaker effects have been shown for English compared to the much smaller five-vowel systems of Shona, Swahili [5]. Cho [6] examined how the degree of vowel-to-vowel coarticulation varies as a function of prosodic factors, and results show that vowels in prosodically stronger locations are coarticulated less with neighboring vowels, but do not exert a stronger influence on the articulation of neighboring vowels. An examination of the relationship between coarticulation and duration reveals that accent-induced coarticulatory variation cannot be attributed to a duration factor, and some of the data with respect to boundary effects may be accounted for by the duration factor. He proposed that prosodically conditioned V-to-V coarticulatory reduction is another type of strengthening that occurs in prosodically strong locations. The prosodically driven coarticulatory pattern can be taken as part of the phonetic signaling of the hierarchically nested structure of prosody.

Studies on various languages have shown vowel-to-vowel coarticulatory effects not only in transitions, but extending into the steady-state period of the transconsonantal vowel both in palatographic data [7], [8] and in acoustic data [9], [10]. While there is ample evidence of the existence of vowel-to-vowel coarticulatory effects, factors have been cited which affect the extent of these effects. For instance, these effects may be constrained by intervocalic palatals and velars, whose production requires use of the tongue body in conflict with the production of vowels, thereby restricting vowel-to-vowel coarticulation [11] and [7].

Despite early indications that V-to-V coarticulation might be a relatively local phenomenon [12], subsequent work has shown that this is not always the case. Instances of long-distance coarticulation, involving effects crossing two or more intervening segments have been found [13], [14]. Magen [15] analyzed [bVvabVvb] sequences produced by four English speakers and found evidence of coarticulatory effects between the first and final vowel, meaning that such effects can cross foot boundaries and multiple syllable boundaries. More recently, Grosvald [16] studied long-distance vowel-to-vowel coarticulation in English, and found that anticipatory V-to-V coarticulation can occur over at least three vowels’ distance in natural discourse, and that even such long-distance effects can be perceived by some listeners.

There has been some research work on the coarticulation of segments in Chinese. Wu and Sun [17] analyzed the acoustic coarticulatory patterns of voiceless fricatives. The fricatives are designed in CVCV contexts in which three peripheral vowels are combined with the fricatives in C1 and C2. Acoustic data, including frequencies and durations of the lower margins, concentration bars, vowel formants and the onset and offset transitions, are measured. It is found that there are three types of coarticulation effects: homorganic,
heterorganic and contiguous.

Yan [18] studied the vowel formant pattern and the coarticulation effect in the voiceless stop on monosyllables, and it is found that there is no effect of tones on formant values of the vowel. However, there is significant effect of aspiration on the formant values at the onset point of the vowel. Chen [19] investigated the intersyllabic anticipatory coarticulation in CVCV sequences, and found that there is no effect of articulation manner on the formant values. There is coarticulatory effect at different vowel points in Chinese. Besides, there is also the study of aspiration on the formant values of the target vowel. Chen [19] investigated the intersyllabic coarticulation effect in the voiceless stop onset contexts are significantly different, it can be concluded that there is no effect of coarticulation at that point. In this study, comparison will be made for both F1 and F2 values at each of the three points, the onset, middle and the offset point.

II. METHODOLOGY

A. Speakers and Stimuli

The speakers for this experiment were eight native speakers of Standard Chinese, four males and four females. The stimulus list is comprised of 6 stimuli, embedded in three carrier sentences of Chinese, each containing an item from a set of two target words. The target words are ‘Biba’ and ‘Baba’, which are supposed to be two persons’ names. They are in the form of ‘bV1ba’, with /a/ as the target vowel, and V1 providing the ‘changing’ vowel contexts, namely, the contexts of /a/ and /u/.

Labial /b/ is chosen to minimize the effects of consonant articulation on lingual articulations [3]. Within the carrier sentences, the target item is located at the sentence initial position. One example is shown as follow:

Biba de jiejie lai le.

Biba’s sister has come.

B. Procedure and Measurements

The speakers were asked to read the sentences three times, in random order for each time, in normal speed, so each speaker produced 18 tokens: 6 sentences × 3 repetitions. In total, 144 tokens were acoustically analyzed (18 tokens × 8 speakers).

This study aims at investigating the extent of V-to-V coarticulation in VCV sequences, and vowel formants were examined. Formant values were extracted using Praat [23], and the extent of trans-consonantal coarticulation was analyzed by comparing the formant values of the target vowel at three points: the onset point, the middle point and the offset point. That is, formant values at the onset, middle and offset points of the target vowel /a/ were extracted, and the values at different vowel contexts were compared. As is mentioned in the previous subsection, there are two contexts for the target vowel, /i/ and /a/. Coarticulatory effect exists if there is significant difference between the formant values in the two vowel contexts. On the contrary, there is no coarticulatory effect if there is no significant difference. A repeated measures ANOVA was performed for the comparison, and statistic analysis was done in SPSS.

Fig. 1 displays the waveforms and formant contours of the two key words ‘Biba’ (a) and ‘Baba’ (b). In the graphs, the second syllable ‘ba’ is the key syllables, and the preceding syllables ‘bi’ and ‘ba’ provide the changing vowel contexts. Formant values are extracted from the onset, middle and offset points of the second syllable ‘ba’, which correspond to point A, B and C in the graphs. Comparison is made for formant values of the target values in the two contexts at each of the three points. Taken the onset point, point A, as an example, since the contexts shown in graph (a) and graph (b) are different, if the formant values at point A in the two contexts are significantly different, it can be concluded that there is an effect of coarticulation at that point. In this study, comparison will be made for both F1 and F2 values at each of the three points, the onset, middle and the offset point.

Fig. 2 presents the mean F1 (a and b) and F2 (c and d) values of the target vowel /a/ for male (a and c) and female (b...
and d) speakers, in the contexts of /a/ and /i/, measured at the onset, middle and offset points. One of the aims of this study is to investigate the extent of the coarticulation, so formant is measured at three points, the onset, middle and the offset points of the target vowel. Analysis is given in the following subsections.

A. Onset Point

Results from a repeated measures ANOVA show that, at the onset point of the target vowel, with data of both male and female speakers pooled together, the effect of changing vowel is significant for both F₁ and F₂; F₁(1, 71) = 40.09, p < 0.001; F₂(1, 71) = 78.31, p < 0.001. That is, coarticulatory effect exists at the onset point of the target vowel.

B. Middle Point

Analysis shows that, at the middle point of the target vowel, the effect of changing vowel is significant for F₂, but not for F₁. F₁: F(1, 71) = 0.63, p = 0.431; F₂: F(1, 71) = 4.65, p = 0.034. Coarticulatory effect exists for F₂, but not for F₁.

C. Offset Point

Coming to the offset point of the target vowel, it is shown that the effect of changing vowel is not significant for either F₁ or F₂; F₁(1, 71) = 0.01, p = 0.912; F₂(1, 71) = 0.58, p = 0.45. Coarticulatory effect does not exist at that point.

IV. DISCUSSION

In reviewing the results reported in the previous section we note that, as far as the data in this study are concerned, trans-segmental coarticulation does exist in Chinese, especially at the onset point of the target vowel. When formant values are examined at the onset point of the target vowel, it is found that coarticulatory effect exists for both F₁ and F₂.

In this experiment, coarticulatory effect is examined at the onset, middle and offset points of the target vowel respectively. It is found that, as far as carry-over coarticulation is concerned, the coarticulatory effect is strong at the onset point, reduced at the middle point, and disappeared at the offset point. To be specific, at the onset point of the target vowel, coarticulatory effect exists for both F₁ and F₂. However, at the middle point of the target vowel, coarticulatory effect exists for F₂, but not for F₁.

The effect on the first and the second formants are not consistent with each other at the middle point of the target vowel, with F₂ affected, while F₁ unaffected. We speculate that the reason for this is that the difference between /a/ and /i/ for F₂ is larger than that for F₁. According to the report of Bao [24], the mean formant values by 8 male speakers are as follow, /a/: F₁ = 984 Hz, F₂ = 1157 Hz; /i/: F₁ = 283 Hz, F₂ = 2350 Hz. The differences between /a/ and /i/ for F₁ and F₂ are 701 Hz and 1193 Hz respectively. The difference of F₂ is much larger than that of F₁. If the formant difference is large, the force for changing the course of formant contour is also large. Vowel coarticulatory effect is more likely to occur on cases with great formant difference, therefore, it occurs on F₂, not on F₁.

When the offset point of the target vowel is investigated, it is found that coarticulatory effect does not exist. That is, for either F₁ or F₂, there is no coarticulatory effect. This result is caused by the ‘distance effect’: at the onset point, the distance from the measured point to the changing vowels is close, and the effect is great; at the middle point, the distance gets farther, and the effect is reduced; while at the offset point, the distance gets even farther, and the effect disappears.

Magen, [15] investigated the extent of vowel-to-vowel coarticulation in English trisyllabic utterances, and it was found that coarticulatory effects can, in some instances, extend beyond the bounds that previous research had assumed; coarticulatory effect can extend from one full vowel, through the medial schwa, and into the midpoint of the next full vowel. He proposed that foot does not define the domain over which coarticulatory effects operate. However,
in the present study, it is found that coarticulatory effect does not extend to the end of the vowel. We speculate this is because that Chinese is of different language typology from English.

One of the well-known properties of language typology is the rhythm unit of a language. Languages have been categorized as mora-timed, such as Japanese, stress-timed, like English and German, and syllable-timed, such as Chinese and French [25], [26]. English is a stress-timed language, and the unstressed syllables are quite weak, so it is possible for coarticulatory effect to extend to the third syllable in English. Chinese is a syllable-timed language, in which syllables are rarely as weak as those in English. Compared to English, the degree of articulatory constraint (DAC) in Chinese is high. Therefore, the coarticulatory effect in Chinese is not as great as that in English.

V. CONCLUSION

In this study, the vowel-to-vowel coarticulation effect in bisyllabic words in Chinese is analyzed, and it is found that trans-segmental coarticulation exists in Chinese, especially at the onset point of the target vowel. Because of the ‘distance effect’, coarticulatory effect is great at the onset point, reduced at the middle point, and disappeared at the offset point of the target vowel. Coarticulation is more likely to occur on F2 because the difference between /a/ and /ɪ/ of F2 is larger than that of F1. In Chinese, coarticulatory effect is not as great as that in English, as the two languages are of different prosodic typology.

This study is significant for speech engineering. In speech synthesis, the effect of trans-segmental coarticulation must be taken into consideration, especially at the onset point of the target vowel. At the middle point of the vowel, only F2 is affected, so it is not necessary to mind about the effect on F1. Nor is it necessary to consider the trans-syllabic coarticulatory effect at the end of the vowel, as in Chinese, coarticulatory effect does not extend to the end of the syllable. Therefore, this study is helpful for speech engineering technology.

ACKNOWLEDGMENT

This work was supported in part by the Social Science Research Project of Guangdong Province, Grant No. GD11CWW04, as well as the Humanity and Social Science Research Project for Colleges in Guangdong Province, Grand No. 11WYXM012. And the research reported here was carried out in the Laboratory of Phonetics and Speech Signal Processing of the Department of Applied Linguistics, Jinan University.

REFERENCES


Maolin Wang was born in Hebei, China, in 1965. Ph.D. from the Chinese Academy of Social Sciences in 2003. Associate professor in Jinan University, Guangzhou, China. Research interest is phonetics and speech signal processing.