International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 2, Issue 10 (October 2013), PP. 15-20

A Study on the Effect of Prosody on the Temporal Realization of Segments in Chinese

Maolin Wang, Guangling Zi
(College of Chinese Language and Culture, Jinan University, China)

Abstract:- In this study the prosodic effect on the temporal realization of segments in Chinese is analyzed. The test segments are in the form of `$C_1V_1\#C_2V_2$', with `14' varied to be prosodic word boundary and prosodic phrase boundary, and C2 varied to be labial, alveolar and velar stop. The subjects are fourteen native speakers of standard Chinese. The VOT of the consonants and the durations of the vowels are analyzed. Results show that the effect of prosody on the duration of the preceding vowel is great, and there is also an effect of place of articulation of the second consonant on the duration of the first vowel, the VOT of the second consonant and the duration of the second vowel.

Keywords:- Duration, prosody, speech, VOT, vowel.

I. INTRODUCTION

It has been shown from researches on prosodic organization that phonological units are organized in hierarchically nested prosodic structure, where lower prosodic constituents are grouped into immediately higher levels [1, 2], and it has been shown that prosodic structure affects the fme acoustic details of segment realizations. A well-known acoustic correlate of prosodic structure is 'final lengthening': domain-final syllables are longer than the medial ones. Phrase-final lengthening has been extensively discussed in the speech production literature [3]. Typically, the last vowel before a large phrasal boundary is lengthened, and other units such as final consonants, final syllables, and final words have been identified as subject to domain-final lengthening as well.

Besides final lengthening, there is also an effect on the initial segment: consonants show more linguo-palatal contact, stops show longer closures and longer Voice Onset Time (VOT), and vowels are more often glottalized and show greater resistance to vowel-to-vowel coarticulation [4-7]. A lot of studies on acoustic effects of prosodic structure have investigated duration. Fougeron and Keating [8] reported that the duration of individual segments at the edges of prosodic boundaries increases with increasing strength of the boundary. In terms of temporal effects of prosodic and linguistic structure in general, research in this area has a relatively long tradition. For example, Lehiste [9] examined what has come to be known as phrase-final lengthening, whereby syllables are lengthened at the end of a prosodic phrase.

Regarding Chinese prosody, there has been much research work in recent years. For example, Li [10] pointed out that in Chinese, the prosodic units of an utterance are intonational phrase, prosodic phrase and prosodic word. Lin [11] reported that at higher prosodic boundaries, there tends to be a pause, syllable lengthening, or FO reset. Wang et al. [12] found that the acoustics correlates of prosodic word boundary are pre-boundary lengthening and discontinuation of the bottom line of intonation. And those of prosodic phrase and intonational phrase are pitch reset of bottom line of intonation and insertion of silence.

In nowaday speech synthesis technology, improving the naturalness of synthesized speech so as to have the expressive variability of actual speech is an important issue in speech engineering. The study of prosodic feature is the crucial work for synthesizing speech samples of high naturalness and high intelligibility, as well as for recognizing natural speech. In this study, the effect of prosody on the temporal realization of segments in different prosodic contexts will be investigated. To be specific, the VOT of the consonants and the durations of the vowels in the prosodic word (PW) boundary and the prosodic phrase (PP) boundary will be analyzed.

II. METHOD

2.1 Speakers, stimuli, and recording

Fourteen native speakers of Standard Chinese, seven male and seven female, were recorded in a sound-Proof recording room. The acoustic data were recorded directly into the computer at a sampling rate of 16 kHz using the software of Cool Edit Pro, and saved as wa^y files.

The key syllables of the stimuli are in the form of `$C_1V_1C_2V_2$', where 'CNC is always `da', and C2 varies to be one of the three consonants of /b d g/ in Chinese, with V2 varying to be one of the two vowels of /a u/. They are embedded in two prosodic contexts, one in PW boundary context and the other in PP boundary context. The carrying sentences are as follow, with the type of prosodic contexts in parentheses and the translation below.

(1) Jingli chuanda bushu le gongzuo jihua. (PW boundary) The manager transmitted and disposed the working plan.

(2) Wu ban de nage Li Da bu shi Sichuan ren. (PP boundary) Li Da in Class 5 is not from Sichuan.

The consonant in bold is varied to be one of /b d g/, and the vowel following the consonant is varied to be one of /a u/. Therefore, there are totally 6 combinations (3 consonants x 2 vowels), with a total of 12 sentences (2 prosodic contexts x 6 combinations). The speakers produce 3 repetitions of the corpus (with the recording order randomized), giving a total of 504 utterances (12 sentences x 3 repetitions x 14 speakers).

As C2 is varied to be one of the three consonants of /b d g/, which correspond to three places of articulation: labial, alveolar, and velar, the effect of place of articulation on the temporal realization of the key segments, C_1, V1, C2, V2, will be analyzed.

Acoustic data were first segmented and labeled by a segmenting program, and then manually verified in Praat [13], and all the data were organized according to the prosodic contexts and places of articulation. Statistic analysis was done in SPSS.

2.2 Measurements

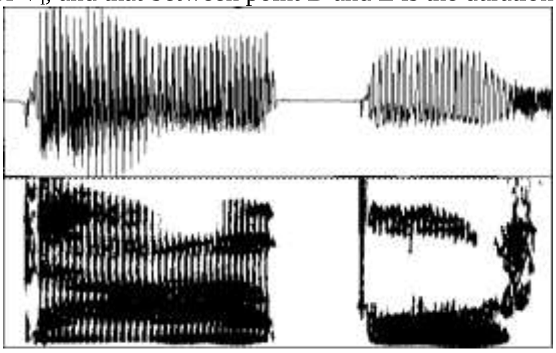
The measurements of this study are as follow,

1) VOT of the consonants

VOT is a feature of the production of stop consonants. It is defined as the length of time between the release of a stop and the onset of voicing. Fig. 1 displays the waveform and spectrogram of the key syllables \`da' and \`bu', in the PP boundary context. In Fig. 1, the interval between point 0 and A is the VOT of C_1, and that between point C and D is the VOT of C2.

2) Duration of the vowels

The durations of V_1 and V2 are also measured in this study. As is shown in Fig. 1, the interval between point A and B is the duration of V_i, and that between point D and E is the duration of V2.



0 A C D
da bu

Fig. 1 The waveform and spectrogram of the key syllables \`da' and \`bu'

RESULTS

3.1 VOT of C1

Fig. 2 graphs the VOT of C_1, broken down by prosodic contexts, PW and PP boundary, and contexts of place of articulation of C2, labial, alveolar and velar. It is shown from repeated measures ANOVA results that there is no significant effect for place of articulation of C2: $F(2, 166) = 2.53$, $p = 0.083$. However, the effect of prosody is significant: $F(1, 83) = 77.9$, $p < 0.001$, with VOT preceding PP boundary longer than that preceding PW boundary.

There is no significant prosody x place of articulation interactive effect either: $F(2, 166) = 2.73$, $p = 0.068$. No interactive effect exists for the two factors.

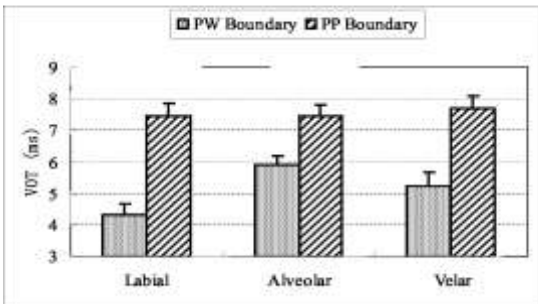


Fig. 2 VOT of C_1 broken down by contexts of prosody and place of articulation

3.2 Duration of V1

Fig. 3 displays the duration of V_1, broken down by contexts of prosody and of place of articulation. From Fig. 3 it can be seen that the prosodic effect is obvious. Repeated measures ANOVA results show that the effect of prosody is significant: $F(1, 83) = 301$, $p < 0.001$. Vowels preceding the PP boundary are longer than that in preceding the PW boundary.

As for the context of place of articulation of C2, for the duration of V_i, there is significant effect: $F(2, 166) = 26.9$, $p < 0.001$. Vowel preceding the alveolar consonant is longer than that preceding the labial: $F(1, 83) = 15.3$, $p <0.001$; and that preceding the velar consonant is longer than that preceding the alveolar: $F(1, 83) = 12.1$, $p = 0.001$.

Further observation shows that there is a significant prosody x place of articulation interaction: $F(2, 166) = 11.4$, $p < 0.001$, which is attributable to the disproportionate effect of prosody in different contexts of place of articulation.

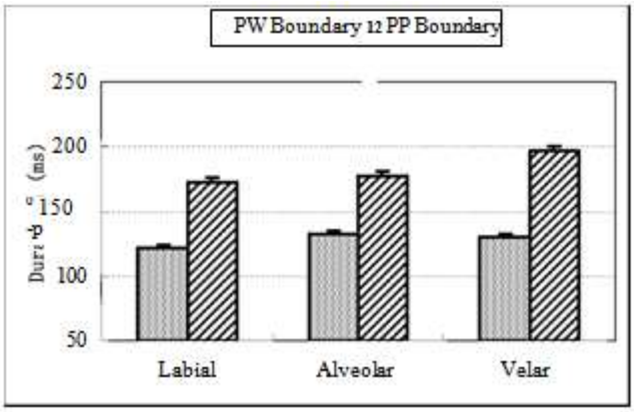


Fig. 3 Duration of V_1 broken down by contexts of prosody and place of articulation

3.3 VOT of C2

Fig. 4 graphs the VOT of C2, broken down by contexts of prosody and place of articulation. From Fig. 4 it is shown that the effect of place of articulation is obvious. Repeated measures ANOVA results display that there is significant effect for place of articulation: $F(2, 166) = 89.5$, $p < 0.001$. The VOT of velar is longer than those of labial and alveolar, velar vs. labial: $F(1, 83) = 90.4$, $p < 0.001$; velar vs. alveolar: $F(1, 83) = 108$, $p < 0.001$

In regard to prosody, as far as VOT of C2 is concerned, there is no significant effect: $F(1, 83) = 1.77$, $p = 0.187$. VOT of C2 is not affected by prosody.

As for the prosody x place of articulation interactive effect, it is significant: $F(2, 166) = 5.18$, $p = 0.007$, which is due to the asymmetric effect of prosody in different contexts of place of articulation.

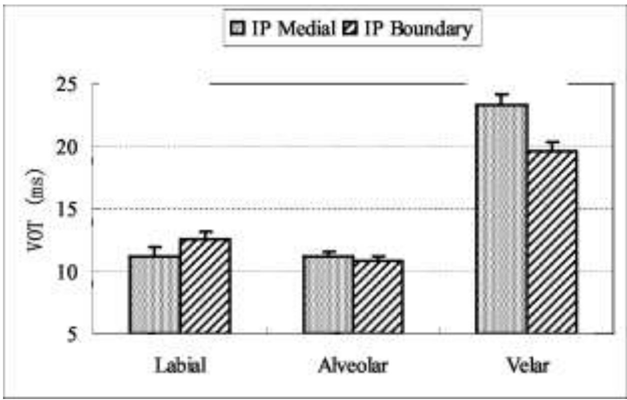


Fig. 4 VOT of C2 broken down by contexts of prosody and place of articulation

3.4 Duration of V2

Fig. 5 displays the duration of V2, broken down by contexts of prosody and of place of articulation. Repeated measures ANOVA results show that the effect of prosody is significant: $F(1, 83) = 71.3$, $p < 0.001$. Vowels following PW boundary are longer than those following PP boundary.

Regarding the context of place of articulation, there is also significant effect: $F(2, 166) = 11.5$, p <

0.001. Vowels following labial are longer than those following velar: $F(1, 83) = 17.6$, $p < 0.001$, while there is no significant difference between the durations of vowels following labial and alveolar.
There is also a significant prosody x place of articulation interactive effect: $F(2, 166) = 49.1$, $p < 0.001$, which is due to the asymmetrical effect of prosody in different contexts of place of articulation.

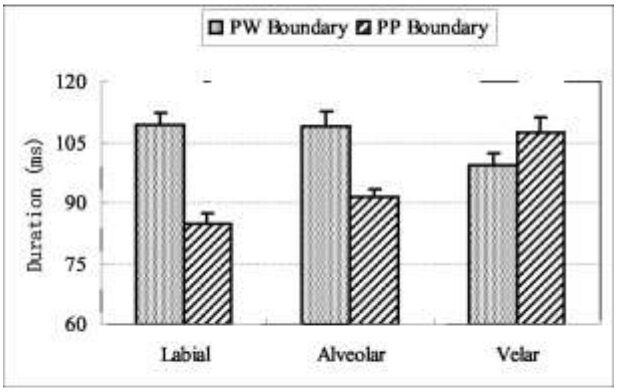


Fig. 5 Duration of V2 broken down by contexts of prosody and place of articulation

IV. DISCUSSION

4.1 VOT of Ci

Results of this experiment show that, at first, the VOT of C_1 is not affected by the place of articulation of C2, which is due to distance effect. In this study, the key segments are in the form of `$C_1V_1$#$C_2V_2$', with 'PP varied to be PW and PP boundaries, and C2 varied to be labial, alveolar and velar. The immediate neighbor of C2 is the prosodic boundary `#', and preceding which there is also vowel V_1. As a result, C_i is far from C2, and the effect of C2 on C_i will be blocked by the boundary and Vi.

However, there is an effect of prosody on the VOT of C_1, with VOT preceding PP boundary longer than that preceding PW boundary. Generally speaking, segments at the end of a prosodic domain tend to be lengthened, especially vowels. In this study, there are two prosodic contexts, PW and PP boundaries, and it is shown that VOT of the stop in the final syllable can also be affected by the following prosodic boundary. Therefore, the effect of prosody on the preceding segments is great.

4.2 Duration of Vi

The effect of prosody on the duration of V_1 is strong. Vowels preceding PP boundary are longer than those preceding PW boundary. In Chinese, there is no length contrast for vowels. In English, there are 'long' vs 'short' vowels, like 'sheep' vs 'ship' and 'short' vs 'shot'. With this contrast, the lengthening effect of vowels in English may be limited to some extent. As there are no phonologically contrastive 'long' or 'short' vowels in Chinese, there is no limitation for the lengthening effect. Therefore, the phrase-final lengthening effect is strong in Chinese.

The final-lengthening effect has the demarcative function: the lengthening of the final segment indicates the end of a prosodic phrase. This effect is communicatively functional, so it has been observed by a lot of researchers. As is mentioned in the first section of this paper, the effect of final lengthening has been extensively reported [3, 8, 9].

There is also an effect of place of articulation of C2 on the duration of V_1. This is not strange, as the two segments are immediate adjacent to each other. We speculate the reason for this is the starting time of the following segment. The duration of a vowel may be affected by the starting time of the following segment. A vowel may be a little shorter if the following segment starts quickly, whereas it may a little longer if the following segment starts slowly.

In this study, C2 is varied to be /b d g/, which belongs to labial, alveolar and velar stop. When a stop is produced, at first, a closure is created in the vocal tract by the lips, tongue blade, or tongue body. The closure will be maintained for an interval of time, and then a release is followed. Generally speaking, for a small, flexible articulator such as the lip or the tongue tip, it is likely to take less time to create the closure, whereas for a larger, less flexible articulator such as the back of the tongue, it will require more time.

The articulators for the labial, alveolar, and velar are the lips, tongue tip and the back of the tongue respectively. The time intervals for the creation of closure at these places are not the same: short for the lips and alveolar, and long for the back of the tongue. Just because of this, the duration of the preceding vowel is affected: short for that preceding labial and alveolar, and long for that preceding velar. As the lips are at the outer part of the mouth, and the closure of the lips is quicker than the tongue tip, vowel preceding labial is shorter than that preceding alveolar.

4.3 VOT of C2

There is an effect of place of articulation on the VOT of C2. VOT is short for labial and alveolar, and long of velar. The reason for this is similar to the effect on the duration of V_1. VOT is affected by the release time. The quicker the release is, the shorter the VOT will be. On the contrary, the slower the release is, the longer the VOT will be. As is mentioned above, for a small, flexible articulator such as the lip or the tongue tip, it is likely to take less time to create the closure, whereas for a larger, less flexible articulator such as the back of the tongue, it will require more time. The temporal pattern for the release is similar to that of closure creation. It may take less time to finish the release by a small, flexible articulator such as the lip or the tongue tip, while it may take more time to do this by a larger, less flexible ariticulator like the back of the tongue. Therefore VOT of labial and alveolar is short, and that of velar is long.

Results in the previous section show that there is no effect of prosody on the VOT of C2. The reason for this is that C2 is varied to be three unaspirated stops /b d g/, and there is few articulatory and acoustic degree of freedom for unaspirated stops in Chinese. VOT of unaspirated stops are short in Chinese, and they are affected by place of articulation. Because of this, there is fewer variability for the unaspirated stops in Chinese. As a result, VOT of C2 is not affected by prosody.

4.4 Duration of V2

The duration of V2 is affected by the place of articulation of the preceding consonant, with vowels following labial and alveolar longer than those following velar. We speculate that this is because of the compensation principle [14]. It is generally approved that languages can be classified typologically into 'stress-timed' languages and 'syllable-timed' ones [15]. Languages like English and German are of the first type, and Chinese belongs to the second type. It is also believed that languages observe the isochrony principle [16]. For stress-timed languages, intervals between the stressed syllables tend to be equal, and for syllable-timed languages, syllables tend to be of equal length.

Results from the previous section show that VOT of labial and alveolar are short, and that of velar is long. Chinese is a syllable-timed language, so syllables tend to be of same duration. The VOT of labial and alveolar are short, vowels following them may tend to get longer to compensate the length. Therefore, vowels following labial and alveolar are longer than those following velar.

The effect of prosody on the duration of V2 is significant, with vowels following PW boundary longer than those following PP boundary. We speculate this is also due to the compensation principle. Vowel preceding PP boundary is longer than that preceding PW boundary. In Chinese, syllables tend to be of equal length, so the duration of the vowel following PP boundary will get reduced to guarantee isochrony of the syllables.

V. CONCLUSION

In this study, the prosodic effect on the temporal realization of segments is analyzed. It is found that, as the obstruction of the prosodic boundary and V_1, the VOT of C_1 is not affected by the context of place of articulation. However, the effect of prosody is strong, and its effect extents to the consonant of preceding syllable. Because of the final-lengthening effect, vowels preceding the PP boundary are much longer than those preceding the PW boundary. As the time intervals for the creation of closure for labial, alveolar and velar are different, there is also an effect of contexts of place of articulation on the duration of V_1. The VOT of C2 is affected by the place of articulation, but as there is little articulatory and acoustic variability for the unaspirated stops, there is no effect of prosody on it. Because of the compensation principle, the duration of V2 is affected by the place of articulation of the preceding consonant, and the duration of V2 is also affected by prosody.

ACKNOWLEDGEMENTS

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.

REFERENCES

[1]. M. Nespor and I. Vogel, *Prosodic phonology.* Dordrecht: Foris, 1986.

[2]. S. Shattuck-Hufnagel and A. E. Turk, "A prosody tutorial for investigators of auditory sentence processing," *Journal of Psycholinguistic Research,* vol. 25, pp. 193-247, 1996.

[3]. C. W. Wightman, S Shattuck-Hufnagel, M. Ostendorf and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *Journal of the Acoustical Society of America,* vol. 91, pp.1707-1717, 1992.

[4]. C. Fougeron, "Articulatory properties of initial segments in several prosodic constituents in French," *Journal of Phonetics,* vol. 29, pp. 109-135, 2001.

[5]. P. A. Keating, T. Cho, C. Fougeron and C.-S. Hsu, "Domain-initial articulatory strengthening in four languages," In *Papers in laboratory phonology,* vol. VI, J. Local, R. Ogden, and R. Temple Eds.

Cambridge: Cambridge University Press, 2003, pp. 145-163.

[6]. T. Cho, "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English," *Journal of Phonetics*, vol. 32, pp. 141-176, 2004.

[7]. M. Tabain, "Effects of prosodic boundary on /aC/ sequences: Acoustic results," *I Acoust. Soc. Am.* vol. 113, no. 1, pp. 516-531, 2003.

[8]. I. Lehiste, "The timing of uttera nces and linguistic boundaries," *J. Acoust. Soc. Am.*, vol. 51, pp. 2018-2024, 1972.

[9]. C. Fougeron and P. Keating, "Articulatory strengthening at edges of prosodic domains," *I Acoust. Soc. Am.*, vol. 101, pp. 3728-3740, 1997.

[10]. A. Li, "The acoustic representation of the prosodic feature in Chinese dialogue," *Chinese Linguistics*, vol. 6, pp. 525— 535, 2002.

[11]. M. Lin, "Prosodic structure and lines of FO top and bottom of utterances in Chinese," *Contemporary Linguistics*, no. 4, pp. 254-265, 2002.

[12]. B. Wang, Y. Yang and S. Lu, "Acoustic analysis on prosodic hierachical boundaries of Chinese," *Acta Acustica*, vol. 29, no. 1, pp. 29-36, 2004.

[13]. P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341-345, 2001.

[14]. L. Feng, "The duration of onsets a nd rhymes in connected speech in Chinese," in *Experimental Studies in Chinese*, Beijing: Beijing University Press, 1985.

[15]. K. L. Pike, *The intonation of American English*, Ann Arbor: University of Michigan Press, 1945.

[16]. J. Laver, *Principles of phonetics*, Cambridge: Cambridge University Press, 1994.