Cues of Prosodic Boundaries in Chinese Spontaneous Speech

LIU Yabin and LI Aijun
Institute of Linguistics, Chinese Academy of Social Sciences
Lyabin@sina.com Liaj@linguistics.cass.net.cn

ABSTRACT

This paper analyzes the acoustic features of prosodic boundaries at all levels in Chinese spontaneous speech using statistical method, and manages to find the cues of prosodic phrase boundaries in spontaneous speech.

1. INTRODUCTION

What are the acoustic cues of prosodic boundaries in continuous speech? This is a fundamental problem in the study of prosody. There have been a lot of studies on this problem and many useful results are achieved. Colin W. Wightman [1] reported that pre-boundary lengthening is significantly correlated with the perceived size of a boundary and at least four distinct types of boundaries can be distinguished on the basis of this lengthening. Statistical results of Chiu-yu Tseng [2] showed pause is a major cue of prosodic structure in continuous speech. It is a most popular opinion that pause, pre-boundary lengthening and F0 reset are the acoustic cues associated with prosodic boundaries in continuous speech, and they are correlated with each other. [4,5,6,7]

However, most of these results are achieved based on read speech rather than spontaneous speech. Our previous work [8] has shown that spontaneous speech is largely different from read speech. So it is uncertain that whether the results based on read speech are still available on spontaneous speech. The purpose of this article is to find the acoustic cues of prosodic boundaries in spontaneous speech based on a Chinese spontaneous speech corpus CADCC. The collecting and labeling of the corpus are introduced in section 2. Statistical analyses on the characteristics of prosodic boundaries are shown in section 3. Finally, conclusions are given in section 4.

2. CORPUS AND ANNOTATION

CADCC (Chinese Annotated Dialogue and Conversation Corpus) is a spontaneous speech corpus that was recorded in the ordinary rooms [9]. Thirteen pairs of speakers recruited are colleagues or classmates who speak Standard Chinese and have common interesting topics so that they can change their topics freely and naturally during an hour's conversation for each pair.

In this paper, we selected about 4 hours of speech balanced phonetically with 7 tiers of segmental and prosodic annotations. The data of F0 and intensity are extracted by Praat [10], and the F0 values are manually modified to get rid of some mistakes.

Segmental and prosodic annotations have been made by using SAMPA-C, and C-ToBI [11] respectively. There are 7 tiers labeled: canonical pinyin and tone tier, initial and final tier (including sound variability), sentence mode tier, paralinguistic phenomena tier, turn-taking tier, break index tier and stress index tier.

Four types of boundaries are marked in break index tier: prosodic word (B1), minor prosodic phrase (B2), major prosodic phrase (B3) (intonation group boundary), and turn-taking boundaries (B4). The default boundary is syllable boundary (B0). However, turn-taking boundary (B4) is not considered in this paper because its interactive function is so complicated that detailed transcription is needed.

3. THE ACOUSTIC ANALYSIS ON PROSODIC BOUNDARIES

3.1 Pause

There are two kinds of perceived pauses: silent pause (SP) and filled pause (FP: a perceived pause without silent interval). All the SPs of 3 level boundaries and their durations are figured out. Fig.1 gives the occurrence frequencies of SPs following every type of boundaries (*number of SP/number of (SP+FP)*). The mean durations of SPs are shown in fig.2.

It can be easily seen from fig.1 and fig.2 that the occurrence frequencies and durations of SPs are both relevant to boundary levels. The higher the boundary level is, the higher the percentage of the number of SPs is and the longer the duration of post-boundary SP is. The SPs following B0 are always very short while the SPs following B3 are longest. There are distinct distances between each two adjacent boundary levels. It should be illuminated that in the analysis the inherent silent gap before a stop is also regarded as a part of silence preceding it, which should be cut out in precise measurements in our further work.

Furthermore, some spoken phenomena (such as breathing) that often occur following B3 should be counted as a part of pause. So the occurrence frequency and duration of pause following B3 should be still longer than our results.

3.2 Duration of Prosodic Units

3.2.1 Normalization

In our study, the duration of every initial or final is normalized using z-score method [12], and the durations of syllables and other prosodic units are achieved by summing up normalized durations of corresponding initials and finals.

3.2.2 Syllable duration

Fig.3 gives the mean durations of all the syllables preceding

ISBN 1-876346-48-5

and following boundaries. It is shown that the syllable duration preceding B1 is a little shorter than that preceding B0, and both of the durations preceding B2 (longest) and B3 are much longer. On the other hand, the durations of post-boundary syllables are also lengthened but the lengthening is not so large as pre-boundary syllables. In general, only the durations of the syllables preceding high level boundaries are lengthened distinctly.

3.2.3 Durations of different tones

In order to observe the influences of boundaries on different tones, we analyzed the durations of pre- and post-boundary syllables with different tones, the results are shown in fig.4 and fig.5. Tone 0, 1, 2, 3, 4 represent neutral, high, rising, low and falling tone respectively.

According to the figures, all tones including neutral tone are clearly lengthened when preceding B2 and B3, and all tones preceding B1 are shorter than those preceding B0 except neutral tone. On the other side, almost all tones are lengthened following B1, B2, B3, except that the low tone following B1 is a little shorter.

It is also shown that the post-boundary low tone is always the shortest tone in the four lexical tones no matter which type of boundary it precedes. As is known to all, low tone is the longest tone in four lexical tones but it always undershoots when it doesn't occur at the end of the prosodic units. So low tone is always greatly shortened among the four tones at the same position. It also indicates that the duration of low tone is more easily influenced by boundaries.

3.2.4 Initials and Finals

In traditional phonetic theory, it is the durations of finals that are subjective to change in sentences, and few changes occur to the durations of initials. Fig.6~fig.9 give the durations of initials and finals preceding and following different boundaries. We classified initials into 6 classes by articulatory method: stops, fricatives, affricates, nasals, lateral and voiced fricatives, which respectively correspond to Class 1 to 6 in fig.6 and fig.7; however, finals are classified into 3 classes: monophthongs (Mo), diphthongs (Di) and triphthongs (Tri).

No clear rules can be found about the pre-boundary initial durations (fig.6). However, all classes of initials are largely lengthened following three boundaries especially B2 and B3, only except that nasals and laterals following B3 are relative short (fig.7).

All 3 classes of finals preceding B2 (longest) and B3 are obviously lengthened (fig.8, 9). Finals following B2 are also a little lengthened but it is not as great as those at the pre-boundary.

Therefore, the main influence of boundaries on initials is the post-boundary lengthening and that on finals is the pre-boundary lengthening. In addition, the duration change pattern of initials and finals here are much similar with that of syllables.

3.2.5 Durations of Prosodic Words

In order to analyze the rhythmic cues of larger units, we calculate the mean durations of pre- and post-boundary prosodic words shown in fig.10. It can be easily seen that prosodic words preceding B2 and B3 are much longer than those preceding B1, and no distinct changes occur to the durations of post-boundary prosodic words. This means the pre-boundary influence also plays an important role on the prosodic word durations.

3.3 F0

F0 changes are observed for both registers and ranges of syllables and words between boundaries. F0 data extracted include mean F0 (MF0: approximate to F0 register, except neutral tone and high level tone syllables), high F0 (HF0), low F0 (LF0) and F0 range (RF0=HF0-LF0). Semitone (St) instead of Hertz (Hz) is used in the measurement with reference frequency of 100Hz.

3.3.1 F0 of syllables

Fig.11 gives the MF0 changes of pre- and post-boundary syllables. It is shown that the pre-boundary MF0 is higher than post-boundary MF0 at B0 and B1, but contrary phenomenon is found at B2 and B3. And the difference is increasing with the boundary level increasing. The F0 reset at B0 and B1 is negative, and it is positive at B2 and B3. And we also found the HF0 and LF0 changing tendencies are similar to MF0.

Then, it indicates that F0 reset is surely present at high level boundaries such as B2 and B3, and the degree of F0 reset increases while the boundary level increases. Here in our analysis, F0 reset is only observed between the pre- and post-boundary syllables, and will be observed in a larger unit in 3.3.2.

3.3.2 F0 of prosodic words

MF0 changes of pre- and post-boundary prosodic words are similar to those of syllables (fig.12). But something is found when we compared it with RF0 changes in fig.13. It can be seen that there is a certain relationship between RF0 and MF0. At B1, the pre-boundary RF0 is lower than post-boundary RF0 while the contrary is to MF0; at B2, the pre-boundary RF0 is higher than post-boundary RF0 but MF0 is quite the contrary; and at B3, both of the pre-boundary values are lower than the post-boundary values. That means, there is little F0 reset at B1 and distinct F0 reset at B3, and B2 looks like a transitional state at which a little F0 reset occurs. However, according to the analyses of durations of syllables and finals, the duration lengthening preceding B2 is greater than that preceding B3, here we can explain why. Because there is a stronger F0 reset at B3, it is satisfied to achieve the perception of B3. It is not necessary to lengthen the duration so much as at B2. Similarly, there is no distinct F0 reset at B2, so it should depend on longer duration to carry it out. In a word, the pre-boundary lengthening and F0 reset are compensative cues for prosodic boundary perception.

3.4 Intensity

Few prosodic studies have focused on the function of intensity in boundary perception. "Intensity increases/ decreases with the tone register rising/falling and it is just a supplemental element in the prosody." [13] However,

ISBN 1-876346-48-5

something interesting is found in our study when it is analyzed to the larger unit of prosodic word.

Fig.14 gives the mean intensity values of prosodic words preceding and following boundaries. The mean intensity of pre-boundary prosodic words falls while the boundary level increases, but the intensity of post-boundary prosodic words rises while the boundary level increases. In other words, the higher the boundary level is, the lower the pre-boundary prosodic word intensity is, and the higher the post-boundary prosodic word intensity is.

4. DISCUSSIONS AND CONCLUSIONS

Obviously, pause is exactly a cue of boundaries based on the strong evidences in 3.1.

Based on the analysis on the durations of syllables, tones, initials and finals and prosodic words, we can find that the pre-boundary syllable lengthening is generally present no matter which tone it is, and the pre-boundary lengthening is mostly achieved by final lengthening while the lengthening of initials can be only seen after boundaries. The pre-boundary lengthening of prosodic words indicates that the pre-boundary lengthening is not only present in a syllable but also in a whole word preceding the boundary. Then the pre-boundary lengthening is also proved to be a cue of boundaries.

According to the analysis in 3.3, F0 reset is also present at both syllable and prosodic word boundaries but with different characteristics. F0 reset is present in mean F0 no matter it is a syllable or a prosodic word, while in prosodic words F0 range also changes correspondingly. So F0 reset (mean F0) and F0 range are both the cues of boundaries.

There are also regular intensity changes with the boundary level changes but only B3 can be distinguished from other boundaries, so it is not as strong as other cues and only can be used as a referential cue at high level boundaries.

In a whole, with an array of statistical analyses on the acoustic performance in the vicinity of the boundaries, this study investigated the suprasegmental characteristics, including the duration of all segments and prosodic units, F0 reset, and intensity changes. And we found that pause, pre-boundary syllable lengthening, F0 reset (F0 register) and F0 range are major cues of boundaries in Chinese spontaneous speech. In addition, distinct intensity rising is also a cue for strong boundary strength.

REFERENCES

[1] Colin W. Wightman, Stefanie Shattuck-Hufnagel, et al, Segmental durations in the vicinity of prosodic phrase boundaries, *JASA, 91: 1707-1717*, 1992.

[2] Chiu-yu Tseng, Major cues of prosodic structures in continuous speech, NCMMSC6, 2001.

[3] CAO Jianfen, Phonetic and linguistic clues in Chinese prosodic segmentation and grouping, *Proceedings of 5th National Conference On Modern Phonetics*, 2001.

[4] LI Aijun, An acoustic analysis on prosodic phrases and sentence accents in Mandarin dialogues, *Proceedings of 4th National Conference On Modern Phonetics*, 1999.

[5] QIAN Yao, et al, An acoustic analysis on prosodic unit boundaries in Standard Chinese, *Proceedings of 5th National Conference On Modern Phonetics*, 2001.

[6] WANG Bei, et al, The acoustic relevant of prosodic hierarchical structures in Chinese, *Proceedings of 5th National Conference On Modern Phonetics*, 2001.

[7] XIONG Ziyu, Pitch reset and breaks, *Proceedings of 5th National Conference On Modern Phonetics*, 2001.

[8] LIU Yabin, LI Aijun, Comparative analysis between read and spontaneous speech, *Journal of Chinese Information Processing, Vol. 16, No. 1*, 2002.

[9] LI Aijun, XU Bo, et al, Spontaneous speech corpus CADCC and the phonetic research, *Proceedings of 5th National Conference On Modern Phonetics*, 2001.

[10] www.praat.com

[11] LI Aijun, et al, A national database design and prosodic labeling for speech synthesis, *Oriental COCOSDA'99,* Taipei.

[12] YIN Zhigang, A study of duration normalization based on corpus, *Proceedings of 5th National Conference On Modern Phonetics*, 2001.

[13] WU Zongji, Appliance of Chinese phonology and phonetics in Chinese speech synthesis, *Language Teaching and Research*, 2002.

Figures:

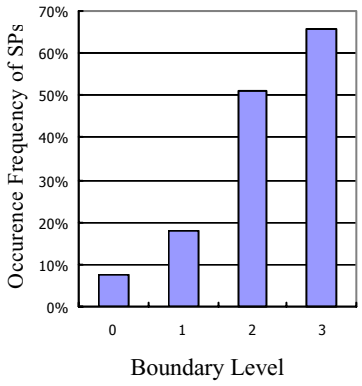

Fig1: Occurrence Frequency of SPs

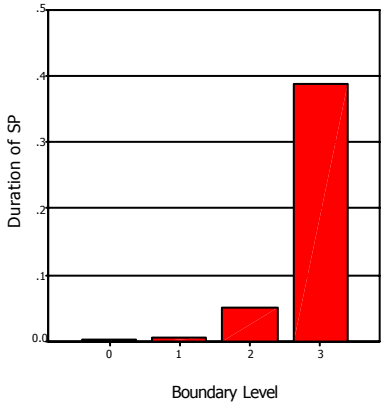

Fig2: Durations of Silences

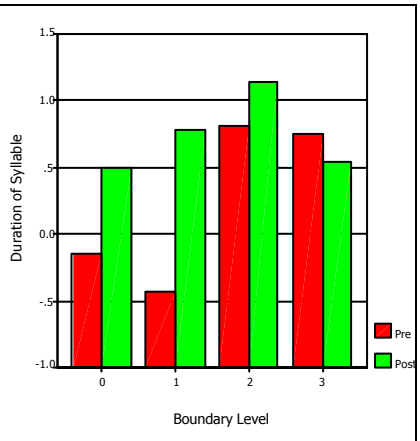

Fig3: Durations of Syllables

ISBN 1-876346-48-5

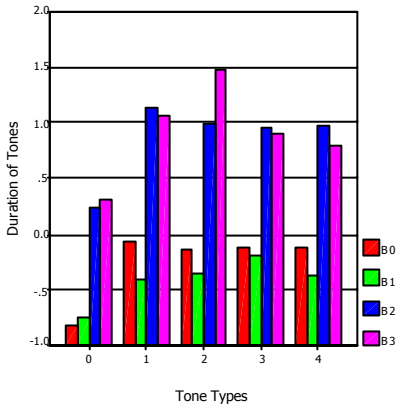

Fig4: Durations of pre-boundary tones

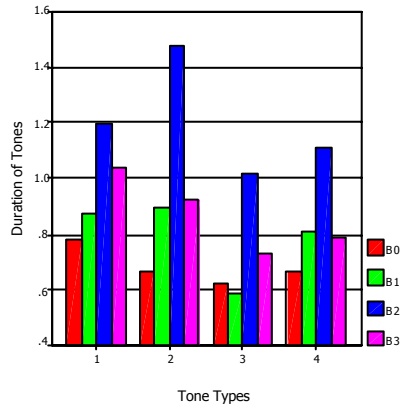

Fig5: Durations of post-boundary tones

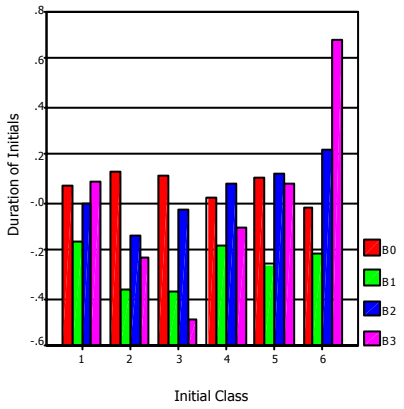

Fig6: Durations of pre-boundary initials

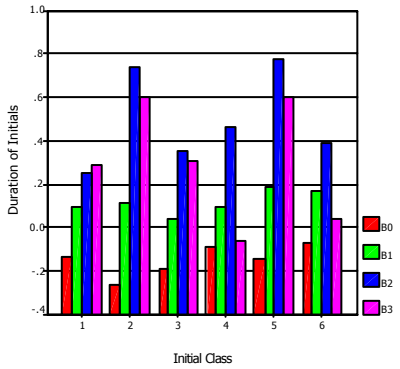

Fig7: Durations of post-boundary initials

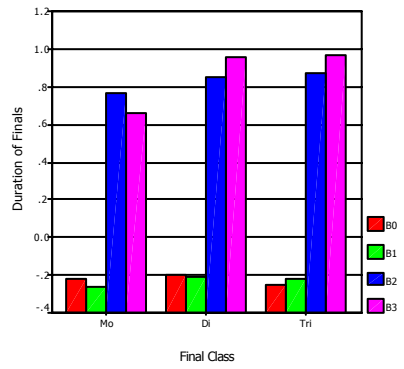

Fig8: Durations of pre-boundary finals

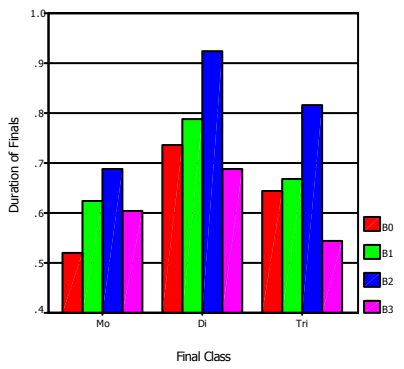

Fig9: Durations of post-boundary finals

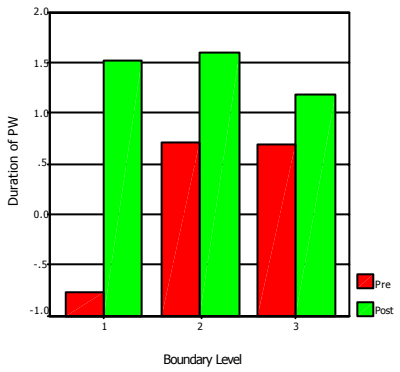

Fig10: Durations of Prosodic Words

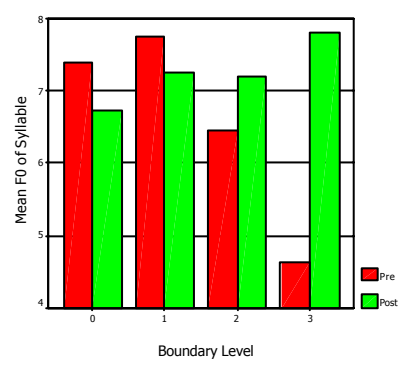

Fig11: MF0 of Syllables

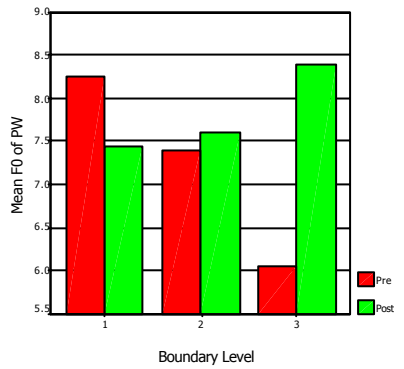

Fig12: MF0 of Prosodic Words

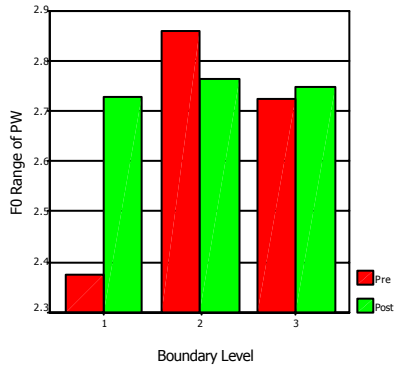

Fig13: RF0 of Prosodic Words

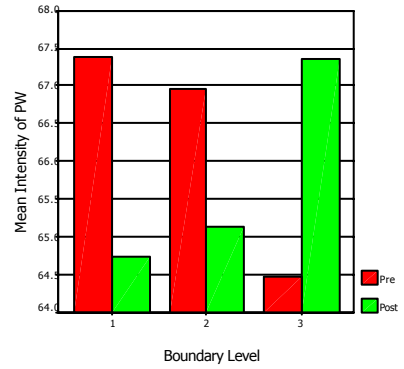

Fig14: Mean Intensity of Prosodic Words