ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2012
Sixth International Conference
Shanghai, China
May 22-25, 2012

Prosodic Boundaries and Prosodic Word in Chengdu Dialect: a Durational Perspective

Zuxuan Qin

School of Foreign Languages, Tongji University, China
simonqzx@163.com

Abstract

This paper will present a study on the prosodic structure of Chengdu Dialect (CD) spoken in Chengdu, capital city of southwest China's Sichuan province. First, we will distinguish different levels of prosodic boundary of CD through perception. Second, we will examine the acoustic manifestation of these boundaries and Prosodic Word from the durational perspective. Finally, we discuss the role of duration in marking prosodic boundaries and determining tone sandhi patterns in CD. Our analysis suggests that listeners might be able to distinguish different levels of prosodic boundaries by attending to the duration of pause and syllables adjacent to them, and that the duration pattern of PW is closely related to tone sandhi patterns in CD.

Index Terms: Chengdu dialect, prosodic word, prosodic boundaries

10.21437/SpeechProsody.2012-150

1. Introduction

This paper aims to present a preliminary study on the prosodic structure of CD. First, it will distinguish six levels of prosodic boundaries perceptually. Second, it will examine the acoustic manifestation of these prosodic boundaries from the perspective of duration. The remainder of this paper will be arranged as follows. Section 2 will give an introduction to the speech data to be analyzed. Section 3 will give a statistics analysis of the duration of syllables before and after the perceptually obtained boundaries and the duration pattern of PW. Section 4 will discuss the role of duration in distinguishing different levels of prosodic boundary and the close relationship between the tone sandhi patterns and the duration pattern of PW in CD. Section 5 concludes this paper by giving a summary of its major findings.

2. About the speech data

Since the present study aims to investigate the prosody of spontaneous speech of CD, we make a recording of spontaneous speech by three native speakers of CD, whose information is shown in Table 1. All the three informants are linguistically naive. Each informant was told to tell a story about their life experience for about thirty minutes in a relaxed state. The recordings were made using Cool Edit Pro 2.0 and a portable computer in quiet rooms in May 2011. The speech waveform was sampled at the rate of 44,100Hz and the resolution was 16 bit.

Table 1. *Basic information about the informants*

Informants	Year of birth	Educational background	Occupation
M1	1933	junior high school	retired civil servant
F1	1928	junior high school	accountant
F2	1941	junior high school	retired civil servant

The transcription of the speech data is made by Praat 5.2.26. It includes the five tiers of Word, Pinyin, Syllable, Sandhi and Break, with the former two transcribing the standard pronunciation of each syllable in Chinese characters and IPA, respectively. Syllable tier transcribes the real pronunciation of each syllable in IPA, and Sandhi tier the actual pitch value of each syllable in Chao's five-digit letters. The Break tier labels the level of perceptual break. Details about the different levels of break are given in Table 2.

Table 2. *Break indices and the corresponding descriptions*

Break index	Corresponding boundary	The corresponding descriptions
B-1	reduced syllabic boundary	between reduced syllables: ni+men—ni<B-1>m
B0	normal syllabic boundary	default case within a polysyllabic word: tsoŋ<B0>kue 'China'
B1	prosodic word (PW)boundary	the minimally perceptible break
B2	prosodic phrase (PPh)boundary	short break larger than B1 but smaller than B3
B3	intonational phrase (IP)boundary	larger than B2; a sense of being non-final about what is going on, something to be continued
B4	utterance (U)boundary	the largest break，a sense of being final about something being said

The transcription was made by a student of linguistics (T1) and the author (T2), both knowing CD well. Before the transcription of the speech data used for the acoustic analysis, we labeled the same set of speech data from the three informants for comparison of the break indices. In the case of inconsistency, a new standard which both transcribers agree on will be set up. After a few such sessions, the inter-transcriber

consistency was found to be over 85%. Only the present author annotated all the speech data to be analyzed in the present study and for limitation of time the other transcriber labeled only two minutes of them so as to check again the inter-transcriber consistency. See Table 3 for detailed statistics.

Table 3. *The inter-transcriber consistency*

T1 / T2	B0	B1	B2	B3	B4
B0	97.7%	2.3%	0%	0%	0%
B1	5.7%	93.1%	1.2%	0%	0%
B2	0%	10.3%	86.2%	3.4%	0%
B3	0%	0%	0%	93.9%	6.1%
B4	0%	0%	0%	0%	100%

Table 3 shows a high inter-transcriber consistency of more than 90% for all break indices but B2, which has a consistency of 86.21%. The fact that 10.3% of what T2 labeled as B2 is annotated as B1 by T1 suggests that it is difficult to distinguish between B1 and B2. Anyhow, the overall high inter-transcriber consistency ensures the validity of the prosodic transcription.

3. The durational perspective of prosodic boundaries and PW

We measure the duration of the syllables before and after different types of boundary to see whether the latter have influence over the former, as claimed in previous studies [1, 2, 3, 4, 5, 6, 7]. Listed below are statistics for a five-minute speech data from F1.

Table 4. *The syllable duration (ms) before and after boundary*

Break index	before boundary			after boundary		
	N	Mean	Std. D	N	Mean	Std. D
B1	399	167	52	399	205	73
B2	93	235	119	93	229	97
B3	63	221	66	63	199	88
B4	33	191	63	33	157	58

There are 1, 270 syllables in the five-minute speech data and the mean duration (hereafter M) is 188 ms (Std Deviation=75). Different types of boundary have varying influence over nearby syllables. The mean duration of syllables before B1 (167) is significantly shorter than M ($p=0.000<0.001$), whereas the duration of syllables after it (205) is significantly longer than M ($p=0.000<0.001$). This indicates that the PW boundary tends to shorten the syllables preceding it but lengthen those following it. The duration of syllables both before and after B2 (235, 229) is significantly longer than M ($p=0.000<0.001$), suggesting that the PPh boundary tends to lengthen syllables both before it and after it. Although the average duration of syllables following B3 (199) is 11 ms longer than M, the difference is not significant ($p=0.318>0.10$). However, the mean duration of syllables before B3 (221) is significantly longer than M ($p=0.000<0.001$). Thus the IP boundary tends to lengthen syllables before it but exerts no obvious influence over the duration of syllables after it. The U boundary has a strong shortening effect on the following syllable, since the mean duration of the latter (157) is far shorter than M ($p=0.005<0.05$). However, it has no effect whatever on the syllable before it, given that the duration of the latter (191) is almost identical with M ($p=0.766> 0.1$).

Perceptually, in CD PW is normally characterized by a strong syllable followed by, if any, one or more weak syllables. Acoustically, it has the normal duration pattern of a long syllable followed by, if any, one or more markedly short syllables. We did a statistical analysis of all the PWs contained in the five-minute speech data. Table 5 shows the mean duration of each syllable in different types of PW.

Table 5. *Duration of syllables in different types of PW*

type of PW	N	1	2	3	4	5
monosyllabic	102	***220***				
disyllabic	292	***205***	***182***			
trisyllabic	100	***209***	***159***	***153***		
quadrisyllabic	24	***247***	***175***	***129***	***149***	
pentasyllabic	4	***212***	***154***	***156***	***121***	***165***

Clearly, the initial syllable in all four types of PW is longer than M. Moreover, it is markedly longer than those after it, which is shorter than M. Thus, the canonical duration pattern of PW can be formalized as LSn ($0\leqslant n\leqslant 4$), where L stands for a long syllable and S a short syllable. That the final syllable in quadrisyllabic and pentasyllabic PWs is longer than some PW-medial syllable is due to the fact that many of them are followed by a B2 or B3, both of which are characteristic of a strong lengthening effect on the preceding syllable. A close look at the data reveals two important factors playing an important role in determining the internal duration pattern of a PW. Firstly, the word class of the initial part of a PW is crucial in determining the internal duration pattern of the latter. If it begins with a lexical word, namely, noun, verb and adjective, it normally has the duration pattern of a long syllable followed by, if any, one or more short syllables. In contrast, a PW beginning with a function word, i.e. a pronoun, quantifier, preposition, conjunction, adverb, numeral, interjections, particle, etc, normally has the internal duration pattern of a sequence of one or more reduced syllables, which are shorter than M. Detailed statistics for the duration of each syllable in a PW with an initial lexical and function word are shown in Table 6.

A few comments are in order. First, except for the disyllabic and quadrisyllabic PW with an initial function word, the mean duration of the initial syllable is longer than any other, if any, subsequent syllable in any other type of PW. The mean duration difference (0.016) between the two syllables in disyllabic PW with an initial functional word is not significant ($p=0.178> 0.1$). Nor is that between the first two syllables in quadrisyllabic PWs with an initial functional word ($p=1> 0.1$). Second, in PW with an initial lexical word the initial syllable is markedly longer than both M and any other, if any, subsequent

syllable, indicating that the normal duration pattern of a PW beginning with a lexical word is LS_n ($0 \leqslant n \leqslant 4$). Third, with the exception of a monosyllabic PW with an initial function word, the first syllable in all other types of PW with an initial function word is shorter than M. A monosyllabic PW with an initial function word is longer than M, due to being the only member of a PW. Thus, the normal duration pattern of a PW with an initial function word is L or S_n ($2 \leqslant n \leqslant 5$). Fourth, for every type of PW having the same number of syllables, the initial syllable belonging to a lexical word is noticeably longer than that belonging to a function word, constituting evidence for the phonological distinction between a functional and lexical word, as seen in many other languages, e.g. English, Japanese and Serbo-Croatian [8]. Finally, the mean duration of the initial and final syllables in each type of PW might well be affected by the number of syllables of the relevant PW and the ratio of different kinds of boundary, which is not discussed here for limitation of space.

Table 6. *Mean duration of syllables in PW with an initial lexical and function word*

type of PW	N	1	2	3	4	5
lexical	44	***245***				
functional	58	***201***				
lexical	185	***227***	***182***			
functional	107	***168***	***184***			
lexical	70	***221***	***162***	***155***		
functional	30	***174***	***152***	***149***		
lexical	19	***212***	***179***	***130***	***149***	
functional	5	***151***	***161***	***128***	***152***	
lexical	3	***254***	***112***	***115***	***125***	***159***
functional	1	***197***	***167***	***169***	***119***	***167***

4. Discussions

The influence of prosodic boundaries over the duration of the neighboring syllables is different between Beijing mandarin and CD. In Beijing, except for syllables before B1, which are significantly reduced, syllables before all other levels of prosodic boundaries (B2, B3 and B4) are lengthened to a large extent [3]. CD is different from Beijing in having syllables, which are not noticeably lengthened before B4. Moreover, prosodic boundaries have no obvious effect on the following syllable in Beijing [3], which is definitely not true in CD.

Given an ideally constant speech rate, the listener can distinguish B1 and B2 from B3 and B4 by comparing the duration of the syllable before them with M. In other words, when a syllable is noticeably shorter than M, the listener may reasonably take it to be followed by B1; when a syllable is more or less the same long as M, B2 is expected to occur after it; when a syllable is markedly longer than M, B3 or B4 may well go after it. Similarly, the listener may distinguish B3 and B4 from B1 and B2 by comparing the duration of the syllable after boundaries. A syllable longer than M suggests a preceding B1 or B2, one markedly shorter than M, a preceding B4, and a syllable more or less of the same length as M a preceding B3. Note the speech rate of spontaneous speech is rarely uniform but keeps changing, in which case the listener obtains M instantly by averaging the duration of syllables within a very short span, perhaps within each utterance.

Overall, the influence of different levels of prosodic boundaries over the duration of the syllables before them is significantly different [$F(3,584)=30.953$, $p=0.000<0.001$]. Specifically, syllables before B1 are significantly shorter than those before B2 ($p=0.000<0.001$) and those before B3 ($p=0.000<0.001$) but not significantly shorter than those before B4 ($p=0.219>0.05$); the duration of syllables before B2 is not significantly different from that of syllables before B3 ($p=0.929>0.1$) and that of syllables before B4 ($p=0.055>0.05$); the mean duration difference between syllables before B3 and B4 is not significant either($p=0.18>0.05$). Thus it is not possible to distinguish the four levels of boundary by considering the duration of the syllable before them alone.

Overall, different levels of boundary exert significantly varying influence over syllables after them [$F(3,584)=7.121$, $p=0.000<0.001$]. Specifically, the mean duration of syllables after B1 is not significantly different than those after B2 ($p=0.169>0.1$) and those after B3 ($p=0.996>0.1$); the mean duration of syllables after B2 is not significantly longer than those after B3 ($p=0.265>0.1$); syllables after B4 are significantly shorter those after B1 ($p=0.000<0.001$), those after B2 ($p=0.000<0.001$), and those after B3 ($p=0.038<0.05$). Thus it is possible to distinguish B4 from B1, B2 and B3 by considering the duration of the syllable after them alone, but not possible to differentiate the latter three in the same way.

Clearly, it seems impossible to tell precisely the strength of a prosodic boundary by considering the duration of the syllable after or before it alone. Note that listeners may have one other way of distinguishing different levels of prosodic boundary in terms of duration of syllables, i.e., the change of duration among neighboring syllables. For instance, given that syllables tend to be shortened before B1 but lengthened after it, a syllable longer than M following one shorter than M normally suggests a PW boundary, whereas a marked decrease in duration between two successive syllables indicates a possible higher level of boundary, i.e. IP or U boundary, which can be further distinguished by looking at whether the first syllable of the disyllabic sequence is markedly lengthened compared with M, as in the case of the former, and/or whether the second of the disyllabic sequence is noticeably reduced, as in the case of the latter. Normally, little or no obvious change of duration between two neighboring markedly lengthened syllables indicates a PPh boundary, since both syllables immediately close to a PPh boundary tend to be lengthened to more or less the same degree, considering that the small difference in mean duration between them (235-229=6) is not significant at all ($p=0.723>0.1$).

Thus listeners might be able to distinguish the four different levels of prosodic boundary by considering the duration of the syllables before and after boundary. However, it does not mean

this is the only means listeners can employ to distinguish different levels of prosodic boundaries. In addition, since prosodic boundaries differ not only in whether they are accompanied by a pause but also in how long the concomitant pause is, pause may be employed by listeners to perceive the different levels of prosodic boundary. Table 7 gives statistics for the ratio of boundaries with a pause for each level of the prosodic boundaries and their corresponding mean duration.

While all B3s and B4s are accompanied by a pause, only a small proportion (16.96%) of B1s and a large proportion (78.65%) of B2s are followed by a pause. Moreover, the pause accompanying a B3 or B4 is normally far longer than that following a B1 or B2. ANOVA shows that overall, the effect of different levels of prosodic boundary on the duration of pause is significant [$F(3, 236)=118.64$, $p=0.000<0.001$], providing a favorable condition for the listener to distinguish the different levels of prosodic boundary involved. Specifically, with the exception of B3 and B4, which do not show a significantly different effect on the duration of pause ($p=0.221>0.1$), the mean duration differences of pauses accompanying different levels of prosodic boundary are all quite significant ($p=0.000<0.001$). This suggests that normally the listener can distinguish B1 and B2 from B3 and B4 by attending to the duration of the relevant pauses accompanying them.

Table 7. *The ratio of boundaries with a pause and their mean duration for each level of prosodic boundary*

Break index	ratio	Mean	No	Std. D
B1	16.96%	27	67	1.4
B2	78.65%	57	70	5.6
B3	100%	463	63	29.9
B4	100%	601	40	58.9

The cano duration pattern of PW (LSn) is closely related to tone sandhi pattern in CD. CD has four tones, whose citation forms are 45, 31, 53 and 213 in Chao's letters, respectively. We assume the following underlying representation for the four tones in CD T1(MH), T2(ML), T3(HL) and T4(LM). CD has the following three tone sandhi rules.

Rule one (R1): MH→H/X__

Rule two (R2): HL→ HH/__ X

Rule three (R3): LM→ L/X__

That T1 and T4 following another tone undergo tone sandhi may be due to the short duration of its host syllable, which is normally reduced and shorter than M, as can be seen from our discussions above. A natural question could be raised as to why T2 and T3 after another tone fail to undergo any tone sandhi. Note that in CD the tones undergoing tone sandhi (T1, T4) are rising, whereas those without any tone sandhi (T2, T3) are falling. The reason for the different tone sandhi behavior between T1 and T4 on the hand and T2 and T3 on the other in CD, we suggest, is that it takes more time to realize the former two than the latter two, since pitch lowering is faster than pitch elevation [9, 10, 11]. Then why is T3 changed to a high level tone on the initial syllable of a PW, which should be long enough for it to fully realize its underlying high falling tone? We assume that this is caused by a constraint in CD forbidding a high falling tone PW-initially. Within Optimality Theory, it is a top ranked constraint, which is never violated in CD.

5. Conclusions

This paper has presented a study on the prosodic structure of CD. The acoustic analysis suggests that both syllable duration and duration of pause between syllables are important cues to the strength of prosodic boundaries. Speakers manipulate them to create different levels of perceptual break to express various kinds of functions, linguistic or extra-linguistic, and listeners try to reach speakers' intension by attending to them. This paper shows that CD and Beijing behave differently in the realization of their prosodic structure. Besides, it shows a close relation between the tone sandhi patterns and the duration pattern of PW in CD.

6. References

[1] Wightman. C. W. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal Acoustic of Society of America* 91(3). 1707—1717, 1992.

[2] Lin, Maocan. Putonghua yujuzhong jianduan he yuju yunlv duanyu [Break and prosodic phrase in standard Chinese]. *Dangdai Yuyanxue* 4: 210-217, 2000.

[3] Xiong, Ziyu & Maocan Lin. Yuliu jianduanchu de yunlü biaoxian [Prosodic manifestation of break in connected speech]. *Proceedings 6th National Conference on Man-Machine Speech Communication,* 2001.

[4] Yang, Yufang. Jüfa Bianjie De Yunlü Biaoxian [acoustic correlates at syntactic boundaries]. *Shengxue Xuebao* 22(5):414-421, 1997.

[5] Cao, Jianfen. Rhythm of spoken Chinese: linguistic and paralinguistic evidence. *Proceedings of ICSLP'2000*, Beijing, 16-20, 2000.

[6] Cao, Jianfen. The Rhythm of Mandarin Chinese. *Journal of Chinese Linguistics*, Monograph Series No. 17, University of California, Berkeley, USA, 2001.

[7] Wang, Bei, Yufang Yang, and Shinan Lü. Hanyu yunlü cengji jiegou bianjie de shengxue xiangguanwu [acoustic correlates of prosodic boundaries in Chinese]. *Proceedings of 5th National conference of Modern Phonetics*, 161-165, 2001.

[8] Selkirk, Elisabeth. O. The prosodic structure of function words. In Papers in Optimality Theory, eds. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 439-70. Amherst, MA: GLSA Publications, 1995.

[9] Ohala, J. J., and Ewan, W. G. ''Speed of pitch change,'' *J. Acoust. Soc. Am*. 53, 345(A), 1973.

[10] Sundberg, J. ''Maximum speed of pitch changes in singers and untrained subjects,'' *J. Phonetics* 7, 71–79, 1979.

[11] Xu, Y. and Sun X. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399-1413, 2002.