The analysis on the duration and pitch of Lhasa Tibetan prosodic unit

JinyongAi/s per 1st
Tibet University for Nationalities
Xianyang, China
ajy0529@126.com

XiaodanGuo/s per 3rd
Tibet University for Nationalities
Xianyang, China
ajy0529@126.com

XiaoyingChen*/s per 2nd
Tibet University for Nationalities
Xianyang, China
ajy0529@126.com

Abstract—**Based on the small speech database of Tibetan single sentences, we get three layers of prosodic structure units by prosodic marking, syllables, prosodic words and prosodic phrases, to study the internal and boundary syllable duration and fundamental frequency data in different prosodic units. The statistical results show that: (a) the syllables duration before and in the prosodic boundary are lengthened, the larger the prosodic units are, the larger the degree of lengthening duration is, but there is no rules of syllables systematic variation after the boundary. (b) It is possible that the f0 reset in the boundary of prosodic hierarchy is positive or negative reset, and the degree of f0 reset is positively correlated with prosodic hierarchy.**

Keywords- Tibetan; prosodic words; prosodic phrases; duration

I. INTRODUCTION

When we use language in communication, we will try to improve the distinctive contrast of pronunciation in some key parts in order to increase significance of perception and allow the listeners to understand the utterances effectively. This kind of significance not only finds expression in segment information such as co-articulation between phonemes and phonetic change rules in the speech flow, but also in the phenomenon of suprasegmental aspects, that is, the characteristics of speech prosody. It has an important role in the naturalness of language performance, enhancing the understanding of speech, eliminating the ambiguity of sentences[1] and improving the naturalness of speech synthesis, etc.

The researches of Chinese prosody have shown that the enhancement or intensification of articulation is usually at the edge of the prosodic borderline, so it reflects the hierarchical structure of prosody with functions of instructing the prosodic structures. And this has important application value in natural language processing. Existing researches agreed that the acoustic performance of the hierarchical structure of prosodic boundary mainly finds expression in the pitch reset, syllable lengthening before the boundary, adding the unvoiced segments and so on. In Chinese researches, the pitch and

duration changes have also been proved to be an important indicator of prosodic boundary. Cao Jianfen[2] put forward it has been found that there is a significant segmental lengthening and pitch highlight in the statements of prosodic boundary and stressed position of sentences, and Feng Yongqiang[3] did the statistical analysis of Chinese syllable duration based on the Microsoft's Chinese corpus and proposed that the lengthening phenomenon of syllable duration before different prosodic boundaries is obvious. Xiong Ziyu[4][5] also proposed that there is a degree of correspondence between the pitch changes and prosodic structure of utterances in Chinese speech flow. Ni Chongjia and some others [6] also have proved that the prosodic phrase boundary has an obvious role in lengthening syllable duration, and the phenomenon of pitch reset in the prosodic phrase boundary is very obvious.

However, now there is not yet in-depth exploration in the aspects of Tibetan prosodic characteristics. In this paper, based on the three kinds of Tibetan prosodic structure units we studied the syllable duration and pitch changes in the interior and boundary of unit mainly from the level of single sentence of prosodic structure, such as syllable, prosodic word and prosodic phrase.

II. PROSODIC MARKING OF SPEECH DATABASE

This recording material in this paper is from the Tibet Daily newspaper in 2007 and we selected in total of 1000 sentences that is fit for standard with about 30 words long in a sentence from short news. The corpus selection took into account the clang association and the position of syllables in different levels of sentences prosodic structure in a continuous speech flow at all levels. And the speaker is a professional male announcer.

We finished the prosodic information marking of Tibetan sentences mainly in Praat software. According to the rules of semantics and the break level of perception we divided the prosodic structure of Tibetan sentences into three layers: syllables, being divided mainly based on the three-dimensional spectrogram of speech signals and voice signals; prosodic words, reflecting the syllable group of Tibetan prosody, no pause in the prosodic words and not necessarily having pause but it can in the boundaries; prosodic phrases,

* Corresponding author: XiaoyingChen

being mainly composed of one or more prosodic words, and usually there are obvious pauses between prosodic phrases and it is the inclusion relation between several prosodic units. The boundary of prosodic phrases must be the boundary of prosodic words, and the boundary of prosodic words can only fall on syllable boundary. So we get three kinds of prosodic boundaries: syllable boundary, prosodic word and prosodic phrase boundary.

Before starting the marking, in order to check the consistency of different persons' marking we asked two Tibetan teachers and two postgraduates of Han nationality who study Tibetan prosody to train the same 50 sentences from both meaning and perception by definition respectively. In this step, they need to discuss the results of the marking, summarize rules, and eventually reach a consensus. Then, we selected another 50 sentences to ask them to mark. The concordance rate of the four marking results for the prosodic word boundary reached to 90.9% and the concordance rate of three marking results for the prosodic phrases boundary reached to 98.9%. This high concordance rate indicates that after training and learning the judgments on prosodic boundary of different people are similar. Finally the corpus is marked by two Han postgraduates and corrected by Tibetan teachers to complete.

III. THE DURATION CHARACTERISTICS OF PROSODIC UNIT OF DIFFERENT HIERARCHIES

A. Number statistics of syllables within the prosodic units

We did the statistics on the number of syllables contained in different prosodic units, and the specific distribution is shown in Table Ⅰ.

TABLE I. THE NUMBER STATISTICS OF SYLLABLES IN DIFFERENT PROSODIC UNIT

Number of syllables	Number of prosodic words	Number of prosodic phrases
1	80	0
2	980	40
3	1320	200
4	1580	470
5	400	410
6	190	420
7	140	330
8	40	180
>8	0	180

Table Ⅰ shows the basic characteristics of Tibetan prosodic units: （ⅰ）the prosodic word contains 2 to 7 syllables, with an average syllable number of 3.5. The conclusion is consistent with the basic features that words are usually composed of the monosyllabic simple words and polysyllabic compound words in most Sino-Tibetan languages.

And Tibetan words is mainly formed by the compounding method, being composed of more than two roots with several changes in morpheme and phoneme, but also some compound words are formed by the three roots and adding up to four syllables. In this sense, the distribution of Tibetan prosodic words is basically consistent with the formation of function words. However, the prepositions are not separated as a division in the scope of prosodic words, so in the prosodic words there are most four-syllable words, while most more than five-syllable prosodic words are some proper nouns such as some foreigner or institution names. From the table we can clearly see that the total number of two-syllable, three-syllable, and four-syllable prosodic words is accounted for 20.7%, 27.9%, 33.4%, being largely accounted for 80% of the prosodic words. （ⅱ）The size of prosodic phrases is from 2 to 13 syllables, the average number of syllables is 6.1, but most of them are in the range from 3 to 8 syllables. The distribution of prosodic phrases is not only physical pause and ventilation needs, but also corresponds to that the span of people's block memory is usually 7±2 syllables in some abroad researches and in line with the rules of speech prosody. Therefore, the length scope of prosodic phrases may correspond to the mechanism of speech production and natural time-domain control of perception.

B. The internal syllable duration statistics in prosodic units

Table 2 shows the statistics of average duration in syllables, prosodic words and prosodic phrases, which are respectively 0.178s, 0.179s, and 0.181s.

TABLE II. SYLLABLE DURATION STATISTICS OF PROSODIC UNITS

prosodic units	average duration （s）	The total number of syllables	Std.
syllable	0.178	33520	0.0391
prosodic words	0.179	16760	0.0436
prosodic phrases	0.181	16760	0.0294

From the above table, it can be seen that the syllable duration of prosodic words is basically consistent with the average duration, and the syllable duration of the prosodic phrases is longer than the average, which may be due to that there are silent pauses between some prosodic words of prosodic phrases. But in general, it can be seen that there is no obvious changes of average syllable duration in different prosodic hierarchies.

C. The statistics of the first syllable duration before prosodic unit boundary

In order to better describe the duration changes of different prosodic units before the boundary of prosodic hierarchies, we did statistics on the syllable duration before the boundary of prosodic words and phrases and compared it with the average syllable duration. The specific result is shown in Figure 1.
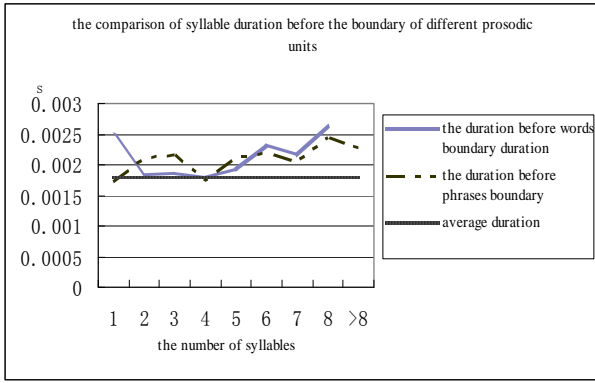
Figure 1.   the comparison of syllable duration before the boundary of different prosodic units

First, from the Fig.1 we can see that all the syllable duration before the boundary of prosodic hierarchies is obvious larger than the average syllable duration. Second, in the prosodic words and phrases, the number of syllables contained in the units has a great influence on the last syllable duration. And Fig.1 also illustrates the lengthening degree of syllables before boundary of the different prosodic hierarchies is not the same, larger prosodic units with larger degree of syllable extension before boundary. The specific performance is that the average syllable duration before the boundary of prosodic words is 0.188s and prosodic phrases is 0.21s, and both are larger than the average duration of all syllables, which is 0.178s.

D.   The comparative analysis of syllables duration in different positions prosodic units

This paper also takes into account the syllable duration in the left and right boundary positions of prosodic units and compares it with the average syllable duration before the boundary, where the left position represents the position next to the left boundary of prosodic unit and the right one represents next to the right boundary of prosodic unit.
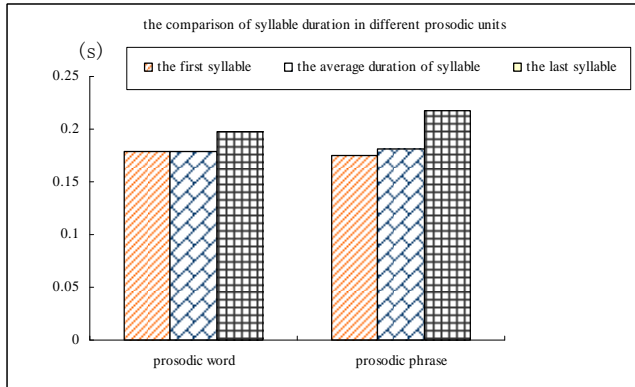

Figure 2.   the comparison of syllable duration in different prosodic units

From Fig. 2 we can see that in the prosodic unit the duration of first syllable is essentially flat with the average syllable duration, while the duration of last syllable is obviously larger than the average one. This is consistent with

what Stares verified, that[7] the syllable lengthening in and before the Chinese prosodic boundary, indicating that there is also the phenomenon of syllable lengthening in the last syllable of Tibetan prosodic units, but there is no systematic change rules in the syllables after the boundary. These may be due to the decrease of sound velocity between prosodic units. Therefore, we can take segmental lengthening of the last syllable in Tibetan prosodic units as important information of prosodic structures.

IV.   THE PITCH CHARACTERISTICS OF THE PROSODIC BOUNDARIES

A.   The pitch prominence of the prosodic boundaries

Table 3 lists the change range calculated by the treble and bass point of each syllable in the prosodic boundaries and the average pitch before and after that to examine the actual pitch and change range of syllables in different hierarchies of prosodic boundaries.

TABLE III.   THE PITCH AND CHANGE RANGE OF SYLLABLES BEFORE AND AFTER THE BOUNDARY IN DIFFERENT HIERARCHIES OF PROSODIC UNITS

units value	prosodic words		prosodic phases	
	before boundary	after boundary	before boundary	after boundary
average pitch (hz)	186.7	179.1	229.6	186.9
pitch range （hz）	52.5	85.1	22.9	79.3

According to Table Ⅲ we can find the distribution overview of the pitch variables: （ⅰ）in every prosodic hierarchy, the pitch of syllable after boundary is lower than that before boundary.（ⅱ）And the change range of pitch after boundary is larger than that before the boundary.（ⅲ）There is a larger gap between the pitch and the pitch change range in the boundary of larger prosodic units and a smaller one between that of smaller prosodic units. Where, the conclusion （ⅰ）and（ⅱ）may be due to the role of "pitch declination", causing the pitch ceiling of the next components of prosodic units in time axis to reduce and the range change. Conclusion （ⅲ）illustrates that the range between prosodic units increases as the prosodic hierarchy does.

B.   The pitch reset in the boundary of different prosodic units

The pitch characteristics of prosodic boundary are mainly expressed in the reset phenomenon, and in order to get the more scientific and more intuitive analysis of the pitch phenomenon of prosodic boundary, in this paper we converted the f0 value in all of the time points that we measured actually into the chromatic value and analyzed it by using f0 minimum as the base, the formula is:

$$B (i) = 12 * \lg (F0 (i) / F0min) / \lg 2 \qquad (1)$$

According to this formula, we can put f0 of each point into a chromatic value B (i) in order. Where, F0 (i) is f0 value of

each measured point and F0min is the minimum of f0 in the measured words. Then we figured out the f0 reset in the boundary of different prosodic hierarchies. TableⅣ, TableⅤ respectively lists the f0 reset of prosodic words and phrases.

TABLE IV.　　F0 RESET BEFORE AND AFTER THE BOUNDARY OF PROSODIC WORDS

	f0 reset(positive)			f0reset(negative)			f0 reset(all)		
	reset	times	std	reset	times	std	reset	times	std
f0 (max）	7.22	3390	7.01	-8.719	1340	6.822	2.621	4730	10.8
f0 (min）	6.31	3390	4.3	-5.901	1340	4.993	2.966	4730	7.1

TABLE V.　　F0 RESET BEFORE AND AFTER THE BOUNDARY OF PROSODIC PHRASES

	f0 reset(positive)			f0reset(negative)			f0 reset(all)		
	reset	times	std	reset	times	std	reset	times	std
f0 (max）	9.87	2180	6.75	-9.37	50	6.59	5.38	2230	10.6
f0 (min）	7.63	2180	4.58	-4.45	50	5.17	6.5	2230	5.8

TableⅣ and TableⅤ show that（ⅰ）it possible the f0 reset in the boundary of prosodic units is positive or negative, but the frequency of positive reset is higher than negative.（ⅱ） With the hierarchy of prosodic unit increased the occurrence possibility of positive reset is higher and the negative is lower. （ⅲ）The f0 reset is closely related to the prosodic hierarchy, the higher the hierarchy is, the higher the f0 reset is, vice versa. （ⅳ）The changes of f0 floor better reflects the structure of prosodic hierarchy, and in the condition of no prosodic boundary, f0 floor gradually decreased.

V.　CONCLUSIONS

Based on the existing conditions of speech technology and Chinese prosodic researches, we did the initial statistical analysis of syllables duration and f0 data of different prosodic units in Tibetan news corpus and obtained some meaningful structure characteristics of Tibetan prosodic units.

However, due to current few prosodic researches on Tibetan language, it is difficult to find some research results to use for reference. This paper is just slight exploration and there are many areas needing for improvement. For example, there is a lack of universality in the selection of corpus, and the number need to expand; besides, there is still room for improvement in the acquisition of experimental data; there is also a lack of deep understanding of prosodic characteristics on the level of stress and intonation, such as the influence of stress on duration and f0, characteristics of the tone phrases; these will need further studies of refinement and depth in the future.

ACKNOWLEDGMENT

REFERENCES

[1] Feng Shengli. Chinese prosodic syntax [M]. Shanghai: Shanghai Education Press, 2000

[2] Cao Jianfen. The acoustic and phonetics characteristics of Mandarin rhythm [A]. Proceedings of modern phonetics[C]. Beijing: Jincheng Press, 1999, 155-159

[3] Feng Yongqiang, Chu Min, He Lin. Statistical analysis of Chinese syllable duration [A]. Modern Phonetics of a New Era [C]. Beijing: Tsinghua University Press, 2001, 66-69

[4] Xiong Ziyu, Lin Maocan. The rhythm performance in the break of speech flow [A]. The proceedings of the Sixth National Conference on Human-machine Communication [C]. Shenzhen: the Sixth National Conference on Human-machine Communication, 2001,179 -182

[5] Xiong Ziyu. F0 reset and break of speech flow [A]. The proceedings of the Fifth National Conference on modern linguistics [C]. Beijing: the Fifth National Conference on Modern Linguistics ,2001,189-193

[6] Ni Chongjia, Liu Wenju, Xu Bo. Duration and pitch of Chinese prosodic phrases [J]. Journal of Chinese Information Processing, 2009,23 (4) ,82-87

[7] Yang Yufang. Prosody performance of syntactic boundary [J]. Acta Acustica, 1997,5:414-421