Prosodic encoding and perception of focus in Tibetan (Anduo Dialect)

Ling WANG[1,2] *Bei WANG*[1*] *Yi XU*[3]

[1] Minzu University of China
[2] Guizhou University for Nationalities
[3] University College London, UK.
wangzhenling2118@163.com, bjwangbei@gmail.com, yi.xu@ucl.ac.uk

Abstract

The prosodic realization of focus and its perception in Tibetan (Anduo dialect) were experimentally investigated. Using the question-and-answer paradigm, the speakers were asked to read aloud two target sentences in different focus conditions. Systematic acoustic analysis and statistical tests showed that, [1] On-focus F_0 was raised sharply in medial and final focus conditions, but not much in initial focus. In addition, post-focus compression (PFC) occurred in initial and medial focus conditions. [2] Duration lengthening was found (about 11%) in focused words, but not in pre-focus or post-focus words. [3] Intensity was increased significantly (about 1.2 dB) in on-focus words, and decreased in post-focus words (about 0.5 dB). [4] In perception, correct focus identification was near 80% for medial focus, 63.3% for final focus, but only about 40% for initial focus. Overall, except for initial focus, the production and perception of focus in Tibetan were similar to those in Mandarin and English.

Key words: Tibetan, focus, intonation, perception

1. Introduction

In speech communication, intonation conveys meanings with ups and downs of F_0 curves. Generally, there are two types of models on intonation as summarized by Hirst [1] and Prom-on et al. [2]. One type starts with form by exploring the linguistic significance of conspicuous ups and downs in intonation; the other type starts from function, searching for encoding mechanisms of various speech functions. Intonation models belonging to the first category include the AM theory [3, 4] and the Tilt Model [5]. Models belonging to the latter include the Fujisaki Model [6, 7], the Stem-ML Model [8], and the PENTA Model [2, 9]. Xu and his colleagues [2, 9] argued that there are two major problems in the models based on form. Firstly, due to physiological constraints of the articulatory system, the realization of underlying intonational targets could only be approached with a process of Target Approximation. Secondly, the movement of intonation usually is the result of encoding multiple communicative functions. Therefore, intonation contours usually do not reflect underlying pitch targets directly, and so it is difficult to assign linguistic meanings to conspicuous intonational patterns. Instead, the investigation of intonation from the perspectives of communicative functions can explain the F_0 variations in a more explicit and direct way[10]. To investigate intonation of Tibetan, we will start with a commonly used communicative function, namely, focus.

Focus is to highlight certain information against the rest of the sentence as motivated by a particular discourse situation[11-13]. It has been found that focused word typically has higher F_0, longer duration and greater amplitude compared to its unfocused counterpart. Focus also suppresses the pitch range of post-focus words, while leaving that of pre-focus words largely intact [12, 14, 15]. This pattern has been found in many languages, such as English [16, 17], German [18], Greek [19], Japanese [20], Swedish [21] and Uygur[10, 22, 33], etc.

It has also been found that a focused word is usually lengthened[12, 23-25]. In Mandarin, the average lengthening of a focused syllable is 4.6% - 17% [12, 14, 17].

As for the perception of focus in English, Herment-Dujardin and Hirst [27] have reported that duration lengthening is not sufficient for focus recognition, whereas pitch raising and pitch range expansion are also required. For languages with on-focus F_0 raising and post-focus F_0 compression, such as Beijng Mandarin [28] and Uygur [22, 33], the recognition rate is above 90%. As Taiwanese lacks post-focus compression, the recognition rate of focus is less than 60%[28]. Xu et al. [29] therefore concluded that post-focus compression is important for focus recognition.

To our knowledge, there has not been much experimental research on the intonation of Tibetan. Tibetan belongs to the Tibetan-Burma branch of Sino-Tibetan language family, and it includes three major dialects, Wei Tibet, West Kang and Anduo. The first two are tonal while Anduo is non-tonal[30]. The general goal of this paper is to investigate the production and perception of focus in Anduo Tibetan.

2. Production experiment

2.1. Method

2.1.1. Stimuli

Two target sentences were constructed, one is short (3 words) and the other is long (5 words). To minimize perturbation and interruption of the continuity of F_0 contours, most of the syllables had sonorant onsets. The sentences are as follows.

Short:

Tibetan: ཝ་མོས་ ར་མ་ བསད།

Chinese: 狐狸 山羊 杀死。(狐狸杀死了山羊)

IPA: wami rama se.

English: Fox goat kill. (The fox killed the goat.)

Long:

Tibetan: ཨ་མས་ ནུ་མོ་ ལ་ ལྭ་བ་ ཉོས།

Chinese: 妈妈 妹妹 给 衣服 买。(妈妈给妹妹买衣服)

IPA: ami nəmu la lawa ni。

English: Mom sisiter for clothes buy.
(Mom bought clothes for my sister.)

For the short sentence, four focus conditions were elicited by *wh*-questions, which were initial, middle, final and neutral focus. Since Tibetan is a verb-final language, the realization of

*Corresponding author.

final focus is not so clear. The last verb is usually a weak element in the sentence. To solve the problem, we added one more condition for the long sentence, for which a focus was put on the penultimate word. Thus there were 1 (short) × 4 (foci) × 3 (repetitions) + 1 (long) × 5 (foci) × 3 (repetitions) = 27 unique stimulus sentences for each speaker.

2.1.2. Participants

Eight native speakers participated in the experiment, five females and three males, aged 19-23, all from Guide County, Qinghai province. They were all college students at Minzu University of China with Tibetan as their first language. None of them reported any speaking or hearing disorders. They were paid with a small amount of money for their participation.

2.1.3. Recording procedure

All the speakers were recorded individually in the speech lab at Minzu University of China. The questions were pre-recorded by a 19-year-old female native speaker. The experimental sentences were repeated three times in a random order, and for each speaker and different random order was used. Before the recording, the speakers read the sentences silently. During the recording, when the experimenter (a Tibetan native speaker) determined that a particular sentence was not uttered properly, the question was played again, and the subject was asked to repeat the target sentence.

The speech signals were directly digitized onto the hard disk of a DELL computer (with built-in 16 bit sound) by a 24Bit/96K Firewire Recording System (PreSonus Firebox) using a condenser microphone (Rode NT1-A) at a sampling rate of 22 kHz.

2.1.4. Acoustic measurement

The target sentences were extracted and saved as separate wav files. To extract continuous F_0 contours, the vocal cycles were firstly marked by Praat and then hand-checked for errors. Segmentation labels were also added to mark syllable boundaries. A Praat script[31] was used to compute maximum F_0, minimum F_0, duration and mean intensity of each syllable. For each word, because the maximum F_0 is mostly at the edge of the first word, which is actually the ending F_0 of the preceding syllable. We used a method similar to Chen and Gussenhoven[25] and Wang and Xu[14], that is, extracting the maximum F_0 from all the non-initial syllables of a word. Because minimum F_0 is not affected by the preceding syllable, we extracted it from the entire word.

The F_0 values were converted from Hz to semitones (st) by the following formula.

$$f_{st} = 12 \times \log_2 (f_0 / 50) \qquad (1)$$

2.2. Results

2.2.1. F_0

The time-normalized F_0 contours of the two target sentences in the four/five focus conditions are presented in Fig. 1 and Fig. 2, averaged across 3 repetitions by 8 speakers.

From these two figures, we can see that focus causes on-focus F_0 raising and post-focus F_0 lowering, while leaving pre-focus F_0 mostly intact. An exception is that initial focus does not show large F_0 raising.

Table 1 presents the values of maximum F_0 and minimum F_0 of each word (initial, medial or final) under different focus conditions, averaged across short and long sentences of 3 repetitions by 8 speakers. For instance, the initial word in the on-focus condition was calculated with F_0 values of the first word in the initial-focus condition of both short and long sentences. And, the initial word of pre-focus condition was calculated with the F_0 values of the first word in the medial focus condition. The two medial focus conditions were averaged for the long sentence.

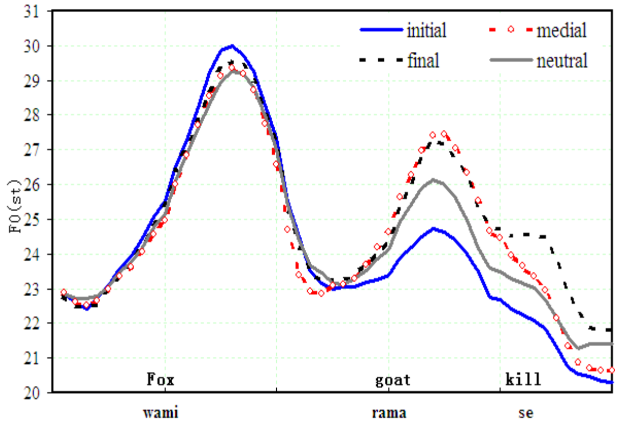

Figure 1. *Time-normalized F_0 contours of the short sentence in four focus conditions.*

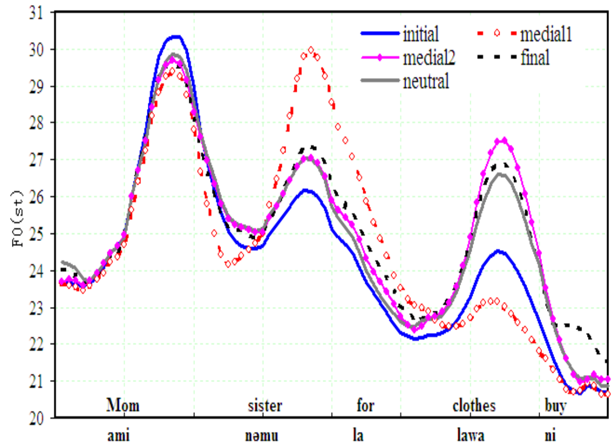

Figure 2. *Time-normalized F_0 contours of the long sentence in five focus conditions.*

Table 1. *Maximum and minimum F_0 of the three target words in different focus conditions*

		Initial	Medial	Final
MaxF0	Neutral	29.7	26.7	24.1
	On-Focus	30.4	28.3	25.0
	Post-Focus	–	25.0	23.5
	Pre-Focus	29.7	27.3	–
MinF0	Neutral	25.2	23.9	20.5
	On-Focus	25.2	24.3	21.1
	Post-Focus	–	22.9	20.1
	Pre-Focus	25.0	24.1	–

Two-way repeated measures ANOVAs with word position and focus condition as independent variables were carried out for short and long sentences separately. The results are presented in Table 2.

Table 2 shows clearly that focus has an effect on both maximum and minimum F_0. The interaction between focus condition and word position mostly comes from the fact that on-focus F_0 raising does not apply in initial focus condition. A post-Hoc test (S-N-K) shows that post-focus F_0 goes lower than its neutral-focus counterpart.

Table 2. *Results of two-way repeated measures ANOVAs on maximum and minimum F_0 of short and long sentences.*

	F_0	Focus	Word	Interaction
		Short: F(3,21)= Long: F(4,28)=	Short: F(2,14)= Long: F(3,21)=	Short: F(6,42)= Long: F(12,84)=
Short	max	8.3**	104.9***	12.6***
	min	4.8*	111.2***	3.8**
Long	max	14.7***	102.9***	26.2***
	min	14.0***	72.0***	8.0***

Note: *stands for p<.05, ** stands for p<.01， *** stands for p<.001.

2.2.2. *Duration*

Fig. 3 presents word duration in different focus conditions, averaged across the short and long sentences. We can see that duration is lengthened in all the focused words.

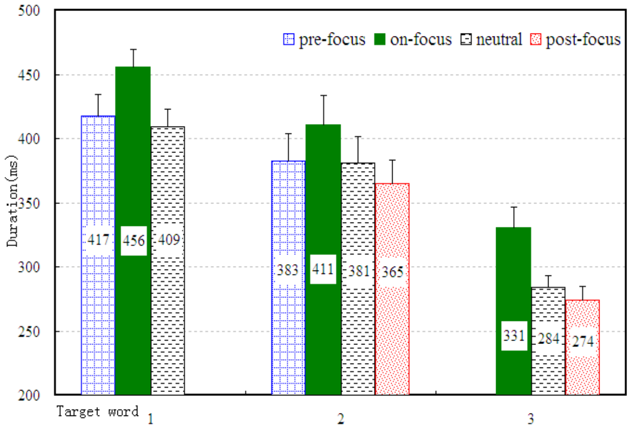

Figure 3. *The average word duration in four focus conditions.*

Two-way repeated measures ANOVAs, with word position and focus condition as independent variables, were carried out for short and long sentences separately (see Table 3).

Table 3. *Results of two-way repeated measures ANOVAs on word duration of the short and long sentences.*

Duration	Focus	Word	Interaction
	Short: F(3,21)= Long: F(4,28)=	Short: F(2,14)= Long: F(3,21)=	Short: F(6,42)= Long: F(12,84)=
Short	2.13, *n.s.*	23.7***	19.3***
Long	3.6*	21.9***	12.0***

As can be seen in Fig. 3 and Table 3, focus has an effect on word lengthening in the long sentence, but not in the short sentence. There is also significant interaction between focus condition and word position.

To summarize, focus is realized with raised F_0, lengthened duration, and post-focus F_0 compression, while leaving F_0 and duration of pre-focus words mostly intact. An exception is that initial focus does not show much F_0 raising.

3. Perception experiment

3.1. Method

3.1.1. *Stimuli*

The number of focus conditions is not the same in the short and long sentences, however the pattern of prosodic realization of focus is the same in the two sentences (see Fig. 1 and 2). To make the perception experiment simple and comparable to similar experiments in Mandarin and Taiwanese[28], we only used the long sentences as the stimuli, and tested initial, medial-1, final and neutral focus conditions. In total, 96 sentences (4 focus conditions × 3 repetitions × 8 speakers) were used as stimuli.

3.1.2. *Participants*

Eleven native speakers, five females and six males, participated in this experiment, and five of them participated in the production experiment as well.

3.1.3. *Listening procedure*

The task was to identify focused word (initial, medial, final, or none). All the 96 sentences were played in Praat using the script by Liu and Xu[32] with a random order for each listener. Each participant listened to the sentences once. During the test, the subjects sat comfortably in front of a computer screen in a quiet room, wearing a headphone set. The whole process took less than an hour.

3.2. Results

Table 4 shows the confusion matrix of focus perception. It can be seen that, the recognition rate of medial focus is the highest (78.4%), followed by final focus (63.3%), with initial and neutral focuses being the lowest (less than 50%).

Table 4. *Confusion matrix of focus perception (%).*

Original	Heard as: Initial	Medial	Final	Neutral
Initial	***37.5***	20.5	10.2	31.8
Medial	6.8	***78.4***	3.4	11.4
Final	8.7	9.5	***63.3***	18.6
Neutral	13.3	22.3	15.2	***49.2***

Overall, the recognition rate of focus in Tibetan is similar to that in Beijing Mandarin[28] and Uygur[22, 33], except for initial focus. The hit rate of initial focus is about 90% in those two languages, but only 37.5% in Tibetan.

4. Discussion

The prosodic encoding of focus in Tibetan (Anduo dialect) is similar to that of Beijing Mandarin[28], Uygur[22, 33] and English[17], in that the focused word has higher F_0 and longer duration compared to its unfocused counterpart. Moreover, there are sharp F_0 lowering and pitch range compression in post-focus words. The analysis of intensity also shows on-focus increase (about 1.2 dB) and post-focus lowering (about 0.5 dB).

It is worth mentioning that the initial focus of Tibetan is an exception. The maximum F_0 of initial focus is only raised about 0.5 st and dropped about 1 st. In contrast, in Beijing Mandarin[28], initial focus raised maximum F_0 about 1 st and lowered the maximum F_0 of the following word about 2 st. The recognition rate of initial focus in Beijng Mandarin (about

91%[28]) is also much higher than that in Tibetan (37.5%). It might because the initial word of Tibetan carries a confound of being the topic. Wang and Xu[14] have found that topic raises sentence-initial F_0. It is possible that topic effect may already saturated the normal pitch range. More studies on this issue is needed.

In addition, the recognition rate of neutral focus in Tibetan is only 49.2%, with relatively equal confusion with initial and final focus (about 10%), but much more with medial focus (22.3%). This result is different from Beijing Mandarin[28] and Uygur[22, 33]. In Beijing Mandarin[28], neutral focus was mostly confused with final focus (27.9%), whereas in Uygur[22, 33], initial focus and neutral focus were confused easily (33%). It raises an interesting question. When we compare different languages on focus realization, can we always take neutral focus as the base-line for all the languages? The properties of neutral focus in different languages therefore need more examination.

5. Conclusions

Based on the above results about focus realization in Anduo Tibetan, we can draw the following conclusions:

1. On-focus pitch was raised sharply in medial and final focus, but not much in initial focus. In addition, post-focus compression (PFC) applied in the initial and medial focus conditions.
2. Durational lengthening was also found (about 11%) in focused words, but not in pre-focus or post-focus words.
3. Intensity was also increased significantly (about 1.2 dB) on focused words, and decreased in post-focus words (about 0.5 dB).
4. Word position has an effect on the perception of focus. The correct identification was nearly 80% for medial focus, 63.3% for final focus, but only about 40% for initial focus.

Overall, except for initial focus, the production and perception of focus in Anduo Tibetan were similar to those in Beijing Mandarin[28], Uygur[22, 33] and English[17].

6. Acknowledgements

We are especially grateful to Professor Liu Yan for many constructive suggestions. We are grateful to Zhou Xianji and her classmates for helping with the experiments. This study was funded by “111” Project of Minzu University of China to the second author.

7. References

[1] Hirst, D. J. Form and function in the representation of speech prosody. Speech Communication, 2005; 46: 334–347.

[2] Prom-on, S., Xu, Y. and Thipakorn, B. Modeling tone and intonation in Mandarin and English as a process of target approximation. Journal of the Acoustical Society of America, 2009; 125 (1): 405-424.

[3] Pierrehumbert, J. The phonology and phonetics of English intonation. Massachusetts Institute of Technology. Ph. D. 1980.

[4] Pierrehumbert, J. and Hirschberg, J. The meaning of intonational contours in the interpretation of discourse. P. R. Cohen, J. Morgan and M. E. Pollack (Eds). Intentions in Communicationeds. MIT Press: 1990: 271-311.

[5] Taylor, P. Analysis and synthesis of intonation using the Tilt model. Journal of the Acoustic Society of America, 2000; 107(3): 1697-1714.

[6] Fujisaki, H. Dynamic characteristics of voice fundamental frequency in speech and singing. P. F. M. ┌ (Eds). The Production of Speech. Springer-Verlag, New York┐, : 1983: 39–55.

[7] Fujisaki, H., Wang, C., Ohno, S., & Gu, W. Analysis and synthesis of fundamental frequency contours of standard Chinese using the command-response model. Speech Communication, 2005; 47: 59-70.

[8] Kochanski, G. and Shih, C. Prosody modeling with soft templates. Speech Communication, 2003; 39: 311-352.

[9] Xu, Y. Speech melody as articulatorily implemented communicative functions. Speech Communication, 2005; 46: 220-251.

[10] Wang, B., Lv, Sh. N. Intonation realization of communicative functions. Hanzang Yu Xubao. 2011; 5: 206-217 (in Chinese)

[11] Dwight, B. A Theory of Pitch Accent in English. Word, 1958; 14: 109-149.

[12] Xu, Y. Effects of tone and focus on the formation and alignment of f0 contours. Journal of Phonetics, 1999; 27: 55-105.

[13] Bolinger, D. Intonation and its uses: melody in grammar and discourse. Stanford: Stanford University Press, 1989;

[14] Wang, B. and Xu, Y. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. Journal of Phonetics, 2011; 37: 502-520.

[15] Chen, Y. Y. Post-focus F0 compression-Now you see it, now you don't. Journal of Phonetics, 2010; 38: 517-525.

[16] Eady, S. J. and Cooper, W. E. Speech intonation and focus location in matched statements and questions. Journal of the Acoustical Society of America, 1986; 80: 402-415.

[17] Xu, Y. and Xu, C. X. Phonetic realization of focus in English declarative intonation. Journal of Phonetics, 2005; 33: 159-197.

[18] Féry, C. and Kügler, F. Pitch accent scaling on given, new and focused constituents in German. Journal of Phonetics, 2008; 36: 680–703.

[19] Botinis, A., Fourakis, M., and Prinou, I. Prosodic effects on segmental durations in Greek. Eurospeech98, 1998;

[20] Ishihara, S. Intonation of Wh-questions in Japanese and its Influence on Syntax. In Proceedings of TCP 2002, edited by Yukio Otsu, 165–89, 2002.

[21] Bruce, G. Developing the Swedish intonaiton model. Working Papers of Lund University Dept. of Linguistics, 1982; 22: 51-116.

[22] Wang, B., Qadir, T. and Xu, Y. Prosodic encoding and perception of focus in Uygur. Chinese Journal of Acoustics. in press. (in Chinese)

[23] Wang, B., Lv, Sh. N. and Yang, Y. F. Pitch movement of stressed syllable in Chinese. Chinese Journal of Acoustics, 2002; 27: 234－240. (in Chinese)

[24] Chen, Y.-Y. Durational adjustment under corrective focus in Standard Chinese. Journal of Phonetics 2006; 34: 176–201.

[25] Chen, Y. Y. and Gussenhoven, C. Emphasis and tonal implementation in Standard Chinese. Journal of Phonetics, 2008; 36: 724-746.

[26] Cooper, W. E., Eady, S. J. and Mueller, P. R. Acoustical aspects of contrastive stress in question-answer contexts. Journal of the Acoustical Society of America,, 1985; 77: 2142-2156.

[27] Herment-Dujardin, S. and Hirst, D. Emphasis in English: A perceptual study based on modified synthetic speech.Speech Prosody 2002.Aix-en-Provence, France.2002: 379-382

[28] Chen, S.-W., Wang, B. and Xu, Y. Closely related languages, different ways of realizing focus.Interspeech 2009.Brighton, UK.2009: 1007-1010

[29] Xu, Y., Chen, S.-w. and Wang, B. Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? The Linguistic Review, in press;

[30] Zeng, G. Q. History and Culture of Tibetan. Minzu Publisher: 2004. (in Chinese)

[31] Xu, Y. TimeNormalizeF0.praat. available from. http://www.phon.ucl.ac.uk/home/yi/tools.html, 2005-2010;

[32] Liu, F. and Xu, Y. Parallel encoding of focus and interrogative meaning in Mandarin intonation. Phonetica, 2005; 62: 70-87.

[33] Wang, B., Wang, L. and Qadir, T. Prosodic encoding of focus in six languages/dialects in China. ICPhS. Hong Kong. 2011: 144-147