

Effects of topic structure and syntax on boundary pitch variations in Standard Chinese

Xiaohong Yang, Yufang Yang¹

¹Institute of Psychology, Chinese Academy of Sciences, Beijing, China, 100101

yangxh@psych.ac.cn, yangyf@psych.ac.cn

Abstract

This paper reported an acoustic experiment examining the effect of topic structure and syntax on boundary pitch parameters. Simple discourses containing three sentences in Standard Chinese were constructed and the boundary between the first sentence and the second sentence of each discourse varied along the following parameters: (a) topic internal, clause, (b) topic shift, clause, (c) topic internal, sentence, (d) topic shift, sentence. The results showed that relative to topic internal condition, topic shift condition induced higher pitch register following the boundary. Syntactic structure as index by clause or sentence had no effect on boundary pitch parameters.

Index Terms: speech prosody, rhythm, melody, intonation

1. Introduction

Text-to-speech synthesis requires the synthesis of isolated phonemic segments into fluent and human-like speech flow. To sound truly natural, artificial speech must contain at least prosodic features such as duration and pitch variation [1]. A major challenge for synthesizing long stretches of sentences is the reinforcement of appropriate and satisfactory prosody. In this paper, we attempted to explore how topic structure and syntactic structure influences boundary prosodic features. Findings of the present research may shed further light on the prosody modeling of discourse.

Previous research in the area of prosodic study has converged on a set of prosodic correlates necessary for synthesis. It has been shown that longer pauses are likely to appear at larger boundaries or topic transitions [2-3], at the beginning of large discourse segments [4-5], or are significantly affected by the depth of embedding [6-7]. Also, it is found that expanded pitch range frequently occurs at the beginning of paragraphs [8-10], or initial segments of discourse [4-5]. A related finding is that higher position in a discourse hierarchy also induces greater f_0 max [7]. Other prosodic markers of interest include boundary tone [2, 11], speaking rate [12, 7], RMS amplitude [4-5], and final lengthening [12].

With regard to variables determining and influencing prosodic features, it is found that prosodic features are determined by a few factors: prosodic structure [13], phrase length [14], syntax [9-10], discourse [2-7], speaker [15], speech rate [16]. In this paper we mainly focus in two of them and their possible interaction: discourse and syntax. With regard to the factor of discourse structure, it has been shown that topic structure has a significant impact on prosodic features in discourse: the type of transition in topic between two successive sentences had a significant effect on the amount of sentence-final lengthening, the duration of the pause between sentences, and the speech rate at the end of a sentence and the beginning of the following sentence [12].

Furthermore, topic structure was also found to affect pitch parameters: the onset f_0 value tends to be higher when the topic is changed at the utterance boundary and the final f_0 value is much higher (on average) at speech-act continuation boundaries than those at other boundaries [17]. However, it is worth to note that previous researches studying topic structure all relied on natural discourses which are long and complex, thus, it remains to be seen whether the effect of topic structure on prosody could be found in much simpler discourse.

As regards syntax, a number of studies have shown that boundary prosodic parameters are affected by syntactic structure of the sentence. Syntactic boundaries are often marked by prosodic means [18]. Pauses within sentences have been found to be shorter than pauses at sentence boundaries [19] and the length of pause duration to be influenced by the syntactic complexity of the preceding and upcoming sentence [14]. Besides time parameter, frequency parameters were also found to be affected by syntactic structures. Thorsen [9-10] have conducted studies in which clauses or short sentences were combined to constitute long sentences or paragraphs. It was found that declination slopes were steeper in the case of the complete sentences and less steep in the case of the clauses separated by commas. Ladd [8] also found that boundary strength indexed by syntactic difference was reflected in the way declination was reset following the boundary.

Although topic structure and syntactic structure are both found to be influential on boundary pitch parameters, it is worthy to point out that previous studies are almost exclusively based on Germanic languages. Taken into consideration that Chinese is a tone language, which employs f_0 variation substantially to signal tone identification and differs greatly from non-tonal languages [20-21], it is worth exploring whether topic structure and syntactic structure are reflected similarly by prosodic features in standard Chinese as have been found in other non-tonal languages. Also, given that both topic structure and syntactic structure have been demonstrated to affect pitch parameters, it is reasonable to suspect that there may be some interactions between these two factors. The current experiment is thus designed to examine the effect of topic structure and syntactic structure on boundary pitch features in Standard Chinese and their possible interaction. To this aim, we manipulated two variables: topic structure (topic internal vs. topic shift) and syntactic structure (clause vs. sentence). Boundary pitch features of interest to us were: mean f_0 , f_0 max, and f_0 min of pre-boundary syllable and post-boundary syllable.

2. Method

2.1. Experimental material

In this experiment, clauses or short sentences were conjoined to form longer domains (long sentences or short

paragraphs) which differed along four conditions. For example, see the following.

Topic internal, clause

这条/裙子/非常/不错,/ 看起来/款式/是/最流行的, / 价钱/也是/很公道的。/

This / skirt /pretty nice, /it seems/ the style/ is /the hottest, /the price / also is / very reasonable./

This skirt is pretty nice, it seems that the style is the hottest, and the price is also very reasonable.

Topic internal, sentence

这条/裙子/非常/不错。/ 看起来/款式/是/最流行的, / 价钱/也是/很公道的。/

This / skirt /pretty nice. /it seems/ the style/ is /the hottest, /the price / also is / very reasonable./

This skirt is pretty nice. It seems that the style is the hottest, and the price is also very reasonable.

Topic shift, clause

这条/裙子/非常/不错,/ 看起来/小静的/眼光/很好, /买衣服时/会很/挑选。/

This / skirt /pretty nice, / it seems that /Xiaojing's/eye/is good, /when buying clothes/good at/choices.

This skirt is pretty nice, it seems that Xiaojing has a good taste, since she makes good choices when buying clothes.

Topic shift, sentence

这条/裙子/非常/不错。看起来/小静的/眼光/很好, /买衣服时/会很/挑选。/

This /skirt /pretty nice. / it seems that /Xiaojing's/eye/is good, /when buying clothes/good at/choices.

This skirt is pretty nice. It seems that Xiaojing has a good taste, since she makes good choices when buying clothes.

Note: We used “/” to segment each word, while the English translations for isolated words and whole sentences were presented below each Chinese sentence.

In the “Topic internal, clause” condition as shown above, the topic of the first segment is about “the skirt” and the topic of the following segment is about the “the style”, which is also part of the skirt. Since the first and the second segment were both talking about the skirt, and the first segment was a clause marked by comma, they constituted the “topic internal, clause” condition. The “topic internal, sentence” condition was constructed simply by changing the comma of the first segment into a period, such that the first segment was changed from a clause into a sentence. The two topic shift conditions below were manipulated by changing the topic of the second segment: the topic of the second segment was changed from “the style” to “Xiaojing”. To sum up, syntactic structure (clause vs. sentence) was manipulated by using comma or period while topic structures (topic internal vs. topic shift) was manipulated by altering the topic of the second segment so that it was either part of the topic of the first segment or irrelevant to it.

Ninety-six short paragraphs constituting of three segments were constructed, twenty-four for each condition. To ensure that the prosodic features investigated did not vary due to segmental structure, for each short paragraph, the first

segment and the first three syllables of the second segment were kept constant across conditions so that boundary segments of interest were always the same in the four conditions. To illustrate, the final word of the first segment was always “不错” and the first word of the second segment was always “看起来” across the four conditions in the example material. In this way, the confounded factor of segmental differences was removed.

2.2. Subjects and recording procedure

Seven native speakers of Standard Chinese read the ten paragraphs aloud. Five of them were female, and the others male. All of them were college students recruited from universities in Beijing. They all spoke standard Chinese without any noticeable accent and were quite experienced in reading. All speakers were aware that it was an experiment on prosody, but were naïve as to the specific purpose of the present study. They were paid for their participation.

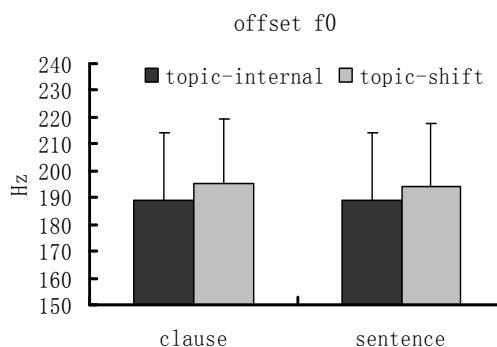
The recordings were made in a sound-proof chamber using the Multi-Speech software at a 22,050 sampling rate. The materials were randomized and presented in 3 pages of A4 paper. The speakers were told to read the sentences aloud at a normal speaking rate. Before the recording started, the speakers were instructed to prepare the materials carefully. In case of errors, they were asked to read the sentences again. The recording session took about an hour and they were encouraged to take at least one break during the session.

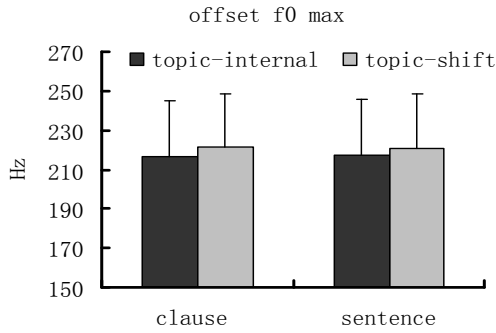
2.3. Acoustic measurement and statistical analysis

For the speech material, the onset and offset, the pinyin, and the tone of each syllable were manually labeled in Praat, and durations were obtained (as illustrated in Fig.4). The f0 curves were visually inspected to remove outliers and f0 max occurring at final rises. Other erroneous pitch-measurements were also corrected. Then, F0 max and f0 min of each word were measured. The data of the seven speakers were pooled together into statistical analysis. Repeated Measures Analyses by subjects was conducted in SPSS 16.0.

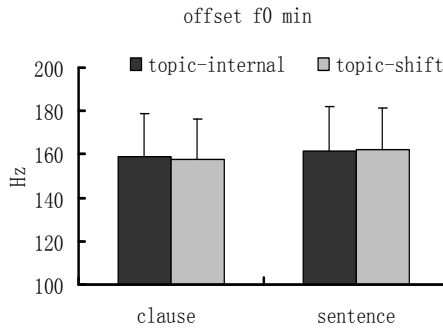
3. Results

3.1. The offset f0, f0 max and f0 min





1B

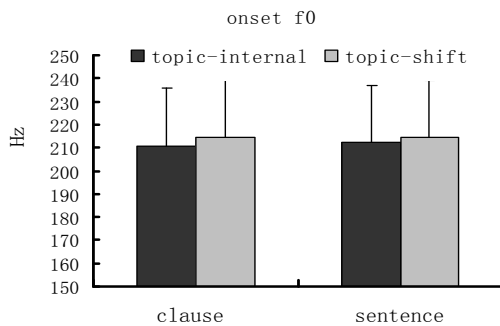


1C

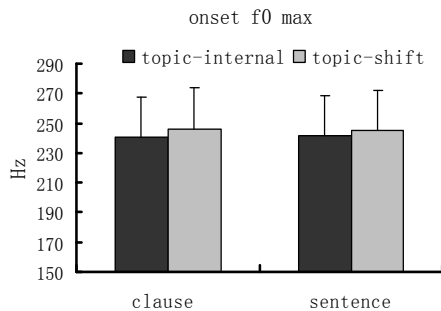
Figure 1 Offset f0 mean, f0 max, and f0 min of pre-boundary segment as a function of topic structure and syntactic structure.

With regard to offset f0, f0 max, and f0 min of the preboundary segment as shown in figure 1A, 1B and 1C, our results revealed no main effects of syntactic structure or topic structure ($p > 0.05$). Also, there was no interaction effect ($p > 0.05$).

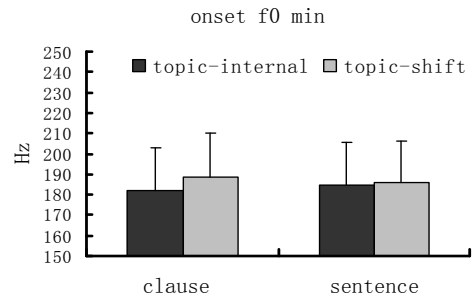
3.2. The onset f0, f0 max and f0 min



2A



2B



2C

Figure 2. Onset f0 mean, f0 max, and f0 min of post-boundary segment as a function of topic structure and syntactic structure.

With respect to onset f0 of post-boundary segment as shown in figure 2A, the results showed that there was a main effect of topic structure [$f(1, 6) = 16.92, p = 0.006$]. Onset f0 were significantly higher following topic shift boundary than following topic internal boundary. The main effect of syntactic structure and the interaction effect were not significant.

For onset f0 max as shown in figure 2B, there was a main effect of topic structure [$f(1, 6) = 15.50, p = 0.008$]. Topic shift induces higher onset f0 max than topic internal condition. The main effect of syntactic structure and the interaction effect were not significant. For onset f0 min as shown in figure 2C, there was a marginally significant effect of topic structure [$f(1,6) = 5.71, p = 0.054$], f0 min following topic shift boundary were higher than f0 min following topic internal boundary. No significant effect of syntactic structure or interaction was found.

Thus, it was found that topic shift induces higher onset f0 following the boundary. Furthermore, the results showed that the higher onset f0 was induced by the raising of pitch register rather than the expansion of pitch range, since both f0 max and f0 min were found to raise after topic shift boundary.

4. Discussion

The experiment reported in the current study aimed to study the effect of syntactic structure and topic structure on boundary prosodic feature and their possible interaction in Standard Chinese. It was found that relative to topic internal condition, topic shift condition induces higher onset f0 following the boundary. Moreover, this higher onset f0 was caused by the raising of f0 max and f0 min, indicating pitch register raising rather than pitch range expansion following the boundary. No effect of syntactic structure or interaction was found in the current study.

With regard to topic structure, we found that topic shift induces higher onset f0, which compares favorably with previous study showing that the onset f0 value tends to be higher when the topic is changed at the utterance boundary [17]. Also, we further demonstrated that the raising of f0 was in fact a result of pitch register raising, which is in line with previous study showing that while focus is encoded by pitch range expansion, topic is manifested by higher pitch register [22]. Furthermore, our results also provided evidence that topic structure did not affect offset f0 as found in the current study. This finding is in complete agreement with Nakajima & Allen [17], which showed that final f0 values at topic shift

boundaries were almost identical to topic continuation boundaries. It is worth to note that instead of long and complex speech materials as used in previous studies [12, 17], we used much simpler discourse and we have thus provided evidence that the effect of topic structure on prosody was as robust in simpler discourses as in complex ones..

As for syntactic structure, while previous studies have found that syntactic structure as indexed by clause or sentence (which was also only marked by orthography such as comma or period) was an influential factor on pitch parameters [9-10], we found no effect of syntactic structure or interaction with topic structure in the current study. Comparing the materials used in the previous study [9-10] with those used in the current study, we found that the semantic connection between the sentences in the text materials was much stronger in the current study than in previous ones. Thus, it is possible that the strong semantic connection between the sentences surpassed the effect of syntactic marker so that no effect of syntactic structure could be found. Also, it is possible that the contradictory findings may result from the difference of language, since the current study was based on Standard Chinese while the previous studies was based on Germanic languages [9-10]. Nonetheless, given the inconsistent results between the current study and the previous one, it remains to be testified whether syntactic structure affects prosodic realization and on what condition the effect could be found.

5. Conclusion

This research attempted to study how prosodic features are influenced by topic structure and syntactic structure. Our results showed that topic structure was realized in higher pitch register following the boundary while syntactic structure was found to have no effect on boundary pitch parameters. Further studies are needed to explore on what conditions these two factors interact with each other.

6. Acknowledgement

This research was supported by the Scientific Research Foundation for Young Scientists, Institute of Psychology, Chinese Academic of Sciences (Y1CX322005).

7. References

- [1] Shriberg, E., L. Ferrer, et al. "Modeling prosodic feature sequences for speaker recognition." *Speech Communication* 46(3-4): 455-472, 2005.
- [2] Swerts, M. "Prosodic features at discourse boundaries of different strength." *J. Acoust. Soc. Am.* 101(1), 514-521, 1997.
- [3] Swerts, M. and R. Geluykens. "The prosody of information units in spontaneous monologue." *Phonetica* 50(3): 189-196, 1993.
- [4] Grosz, B., and Hirschberg, J. "Some intonational characteristics of discourse structure," *Proc. International Conference on Spoken Language Processing, Banff, Canada, October 1992*, pp. 492-432.
- [5] Hirschberg J., Nakatani CH. "A prosodic analysis of discourse segments in direction-giving monologues," *In: Proceedings of the 34th annual meeting on Association for Computational Linguistics; 1996. Association for Computational Linguistics Morristown, NJ, USA.* pp. 286-293.
- [6] Noordman, L., Dassen, I., Swerts, M., Terken, J. "Prosodic markers of text structure," *In: van Hoek, K., Kibrik, A., Noordman, L. (Eds.), Discourse Studies in Cognitive Linguistics, Selected Papers 5th International Cognitive Linguistics Conference. John Benjamins, Amsterdam*, pp. 133-148, 1999.
- [7] den Ouden, H., L. Noordman, et al. "Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports." *Speech Communication*. 51(2), 116-129, 2009.
- [8] Ladd, D.R. "Declination reset and the hierarchical organization of utterance," *J. Acoust. Soc. Am.* 84 (2), 530-544, 1988.
- [9] Thorsen, N.G. "Intonation and text in Standard Danish," *J. Acoust. Soc. Am.* 77, 1205-1216, 1985.
- [10] Thorsen, N.G. "Sentence intonation in textual context—supplementary data," *J. Acoust. Soc. Am.* 80, 1041-1047, 1986.
- [11] Herman R. "Phonetic markers of global discourse structures in English." *Journal of Phonetics.*, 28(4), 466-493, 2000.
- [12] Smith, C. L. "Topic transitions and durational prosody in reading aloud: production and modeling," *Speech Communication*, 42(3-4), 247-270, 2004.
- [13] F. Ferreira, "Creation of prosody during sentence production," *Psychological Review*, vol. 100, pp. 233-253, 1993.
- [14] J. Krivokapi, "Prosodic planning: Effects of phrasal length and complexity on pause duration," *Journal of phonetics*, vol. 35, pp. 162-179, 2007.
- [15] Fant, G., Kruckenberg, A., & Ferreira, J. B. "Individual variations in pausing. A study of read speech." *In Proceedings of Fonetik 2003, Reports in Phonetics, Umea University, PHONUM, Vol. 9*, pp. 193-196.
- [16] Fletcher, J. "Some micro and macro effects of tempo change on timing in French". *Linguistics*, 25, 951-967, 1987.
- [17] S. Nakajima and J. F. Allen, "A study on prosody and discourse structure in cooperative dialogues," *Phonetica*, vol. 50, pp. 197-210, 1993.
- [18] A. Batliner, et al., "M = Syntax + Prosody: A syntactic-prosodic labelling scheme for large spontaneous speech databases," *Speech Communication*, vol. 25, pp. 193-222, 1998.
- [19] Sanderman, A. A., & Collier, R.. "Prosodic phrasing at the sentence level". *In F. Bell-Berti, & L. J. Raphael (Eds.), Producing speech: Contemporary issues. For Katherine Safford Harris (pp. 321-332). New York: American Institute of Physics.* 1995.
- [20] Tseng, C. Y. and Z. Y. Su. "What's in the F0 of Standard Speech -Tones, Intonation and beyond." *2008 6th International Symposium on Chinese Spoken Language Processing, Proceedings: 45-48*, 2008.
- [21] Chen, Y. and C. Gussenhoven. "Emphasis and tonal implementation in Standard Chinese." *Journal of phonetics* 36(4): 724-746, 2008.
- [22] Wang, B. and Xu, Y. "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese". *Journal of Phonetics* 37: 502-520, 2011.