PROSODIC REALIZATION OF DISCOURSE TOPIC IN MANDARIN CHINESE: COMPARING PROFESSIONAL WITH NON-PROFESSIONAL SPEAKERS

Bei Wang[a], *Yi Xu*[b] *& Jiang Xu*[a]

[a]Minzu University of China, Beijing, China;
[b]University College London, London, UK
bjwangbei@gmail.com; yi.xu@ucl.ac.uk; bolexujiang@126.com

ABSTRACT

This study investigated prosodic realization of discourse topic by comparing professional and non-professional speakers' production of four discourses. In each discourse, the target sentence, as the discourse topic, was either at the discourse initial position or at a non-initial position. Six professional and 20 non-professional speakers read aloud all the discourses with two repetitions. Extensive acoustic analysis revealed that: (1) Sentences in the discourse initial position started with higher maximum F_0 and longer duration, mostly in the first prosodic phrases, with the biggest effect on the first word; (2) Professional speakers raised discourse-initial F_0 by a much larger amount and in a more consistent way than non-professional speakers; (3) The two groups of speakers both lengthened discourse-initial duration, and there was no difference between the two groups in terms of duration lengthening.

Keywords: discourse topic, prosody, intonation

1. INTRODUCTION

When several sentences are strung into a larger unit, they form a discourse with a shared common topic [8]. A large body of literature suggests that the structure of such a discourse is reflected in prosody. A well-known effect is that the beginning phrase starts with higher F_0 and larger pitch range than the following phrases [7, 10] and it is louder and slower than other phrases [3]. As the discourse goes on, there is a tendency for F_0 to either decrease gradually or remain the same over the course of the paragraph [2, 4, 6, 7, 9].

Lehiste [5] and Thorsen [9] have shown that F_0 is higher in paragraph-initial position than in medial and final positions and in isolated sentences.

In [11], it has been found that topic raises pitch range, and lets F_0 decline gradually afterward. However, in that study, the F_0 raising in a discourse initial sentence is just about 1 st, which is much smaller than the effect reported by Umeda [10]. From that study and several pilot experiments, it occurred to us that Umeda may be right in her suggestion that some speakers are better than others in making their reading vivid by raising discourse-initial F_0 [10]. If this is the case, speakers with more training in reading aloud should show greater topic-related initial F_0 raising. To our knowledge, however, there has been no direct experimental evidence supporting such an idea. There is nevertheless emerging evidence that an important feature of charismatic speech is having larger F_0 excursions, which presumably could include having a topic-related initial F_0 raising. The present study was therefore designed to test the hypothesis that speakers with professional training raise discourse-initial F_0 more than speakers with no professional training.

2. METHOD

2.1. Materials

Four discourses were constructed, each containing two paragraphs with about 300 syllables. For each discourse, two versions were constructed. In version A, the target sentence was the title sentence. In version B, the target sentence was embedded in the middle of the second paragraph of the discourse. The four target sentences all start with a four-syllable noun phrase with varied tone combinations. The sentences are 12 -14 syllables long.

The four target sentences in Pinyin are shown below with English translations.

(1) jia1rong2 guo1zhuang1 shi4 min2zhu2 yi4shu4 de0 qi2pa1
Jiarong guozhuang be national culture DE exotic art
'The Guozhuang dance of Jiarong is an exotic art of the national culture.'

(2) ping2le4 gu3zhen4 bao3liu2 zhe0 you1ju3 de0 li4 shi3 wen2 hua4
Pingle old town preserve PFV long DE history culture
'The oldl town of Pingle preserves the long history and culture.'

(3) ling3nan2 yuan2lin2 ju4you3 feng1fu4 de0 wen2hua5 nei4han2
Lingnan garden hold rich DE culture connotation
'The gardens of Lingnan hold rich culture connotation.'

(4) re4 gong4 yi4shu0 hui4 ju4 le0 zhong4duo1 yi4shu0 men2lei4
Regong art gather PFV many art category
'The art of Regong gathers many different kinds of art.'

2.2. Participants

Twenty non-professional and six professional speakers participated in the experiment. The non-professional speakers were university students from Minzu University of China, 11 male and 9 female, aged 19-37. They were all from northern China and spoke Mandarin without noticeable accent, but had no professional training in news reading. The professional speakers were 3rd year students at the Communication University of China, majored in broadcasting. They had taken extensive training in professional reading.

2.3. Recording procedure

All the speakers were recorded individually in a sound-proof recording room. The speakers were asked to get familiar with the texts before the recording. They were instructed to read the texts in a manner suitable for radio broadcasting, with expressiveness and emotional engagement. When a mistake was made, such as reading a word wrongly or having an obvious hesitation, the speaker was asked to read the whole text again. The recording lasted about one hour for non-professional speakers and about 40 minutes for professional speakers.

2.4. Acoustic measurement

The target sentences were extracted and saved as separate wav files. The acoustic analysis procedures were similar to those in [14]. ProsodyPro, a Praat script [13] was used to take F_0 and duration measurements from the sentences. To extract continuous F_0 contours, the vocal cycles were first marked by Praat [1] and then hand checked by the first and the third author for errors. While checking for vocal pulse markings, segmentation labels were also added to mark the syllable boundaries. The vocal periods were converted into F_0 values by ProsodyPro, which then removes spikes from the resulting F_0 contours using a trimming algorithm [12, 13]. The vocal pulse marking, segment labels, and F_0 values for each utterance were saved in text files. ProsodyPro also obtained the highest and lowest F_0 (in semitone) and duration of each syllable.

3. RESULTS

3.1. General descriptive analysis

For each discourse, the F_0 contours of the target sentences in discourse initial and non-initial positions are displayed together in each plot in Fig. 1 for professional speakers and Fig. 2 for non-professional speakers.

Figure 1: Intonational contours of the target sentences in discourse initial and non-initial position, averaged by 6 professional speakers.

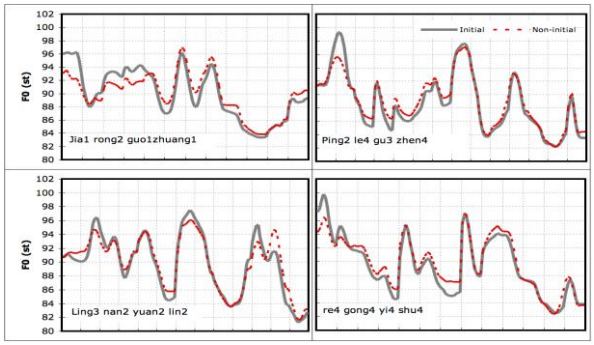
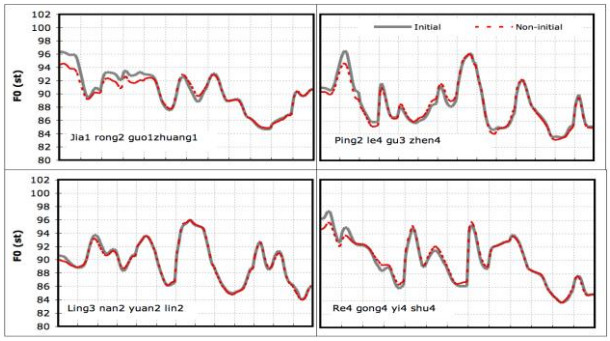


Figure 2: Intonational contours of the target sentences in discourse initial and non-initial position, averaged by 20 non-professional speakers.



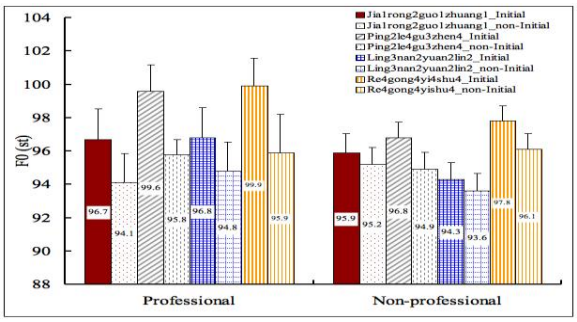
As can be seen in Fig. 1, professional speakers use higher F_0 in discourse-initial sentences than in the same sentences in non-initial positions. However, the F_0 difference is much smaller in the speech of non-professional speakers as seen in Fig. 2.

With a closer look at Figs. 1-2, we can make three observations. First, the F_0 raising in discourse initial sentences occurs mostly in the first prosodic phrase, with the largest effect on the first word. Second, the amount of F_0 raising in discourse initial position varies with tone combinations. It seems that the effect is the largest on tone 4 (falling), smaller in tone 1 (high) and tone 2 (rising), and the smallest in tone 3 (low-dipping). Third, the F_0 raising in discourse initial position is shown mostly in maximum F_0, with little effect on minimum F_0.

3.2. Analysis of F_0

First, the maximum F_0 of the first prosodic phrases in the target sentences was calculated (see Fig. 3).

Figure 3: Maximum F_0 of the first prosodic phrase in discourse-initial and non-initial target sentences.



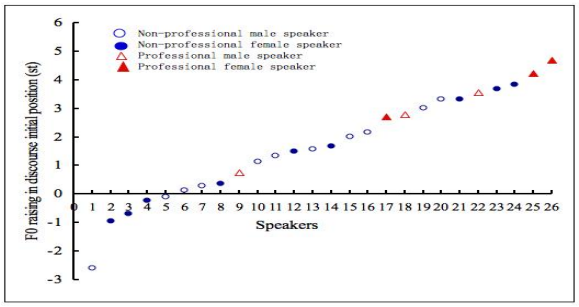
As can be seen in Fig. 3, the maximum F_0 of the first prosodic phrase is higher in discourse-initial sentences than in non-initial sentences, which is true for all texts and both groups of speakers. It is supported by a two-way mixed repeated measures ANOVA with text and topic level as two within-subject factors and training as a between subject factor (see Table 1). What can also be seen in Fig. 3 is that the effect is much bigger in professional speakers than in non-professional speakers (3.1 vs 1.25 st on average), which is shown as the significant interaction between topic level and training in Table. 1.

Table 1: The F and *p* values of 2-factor repeated measures ANOVAs with training as a between subject factor on the maximum F_0 of the first prosodic phrase.

	F	*p*
Text	F(3,72)=21.497	<0.001
Text*Training	F(3, 72)=19.876	<0.001
Topic	F(1, 24)=265.13	<0.001
Topic*Training	F(1, 24)=249.869	<0.001
Topic*Text	F(3,72)=5.006	0.01
Topic*Text*Training	F(3, 72)=5.102	0.009

For all the speakers, the mean differences (across the four texts and two repetitions) in maximum F_0 of the first prosodic phrase between discourse initial and non-initial conditions is calculated (see Fig. 4). Five professional speakers (83.3%) and 7 non-professional speakers (35%) raised discourse initial sentences more than 2 st relative to non-initial sentences. Among the 7 non-professional speakers, four are male and three are female, which indicates that gender is not a factor once the semitone scale is used.

Figure 4: Mean maximum F_0 diference of the first prosodic phrase of the target sentences in discourse-initial and non-initial conditions.

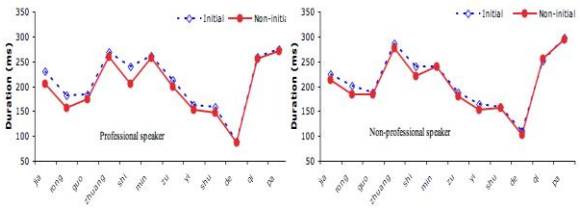


Due to space limitation, results on minimum F_0 and the acoustic measurements in the other words are not presented here.

3.3. Analysis of syllable duration

For professional and non-professional speakers, durational values of all syllables in the target sentences of the text "jiarong guozhuang" are presented in Fig. 5, with discourse initial and non-initial conditions overlaid in one figure. Due to space limitation, similar figures of the other texts are not presented here.

Figure 5: Duration of all the syllables in the target sentences of discourse initial and non-initial conditions.



In Fig. 5, we can see that the lengthening effect in discourse initial sentences is applied mostly in the first prosodic phrase. This holds for both professional and non-professional speakers. And it is the same in the other three texts as well.

The mean duration of the first prosodic phrase in each text for the two groups of speakers are presented in Fig. 6, which shows that duration at discourse initial position is longer than at non-initial positions in all the texts. Table 2 presents a similar two-way mixed repeated measures ANOVA on the duration of the first prosodic phrase. The topic level has a significant effect on the durational lengthening. On average, the discourse-initial lengthening is 11% for professional speakers and 7.9% for non-professional speakers. As shown in Table 2, training does not have a significant interaction with

topic level. It indicates that the two groups of speakers use durational lengthening to roughly the same degree in discourse initial sentences.

Figure 6: Mean duration of the first prosodic phrase in discourse initial and non-initial target sentences.

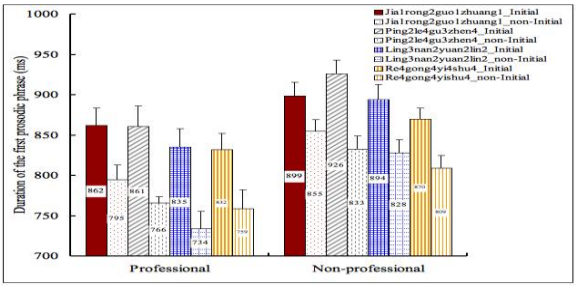
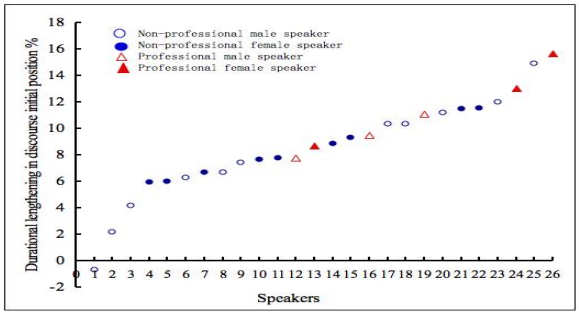


Table 2: The F and *p* values of 2-factor repeated measures ANOVAs with training as a between subject factor on the duration of the first prosodic phrase.

	F	*p*
Text	F(3,72)=3.371	0.033
Text*Training	F(3, 72)=0.644	*n.s.*
Topic	F(1, 24)=125.93	<0.001
Topic*Training	F(1, 24)=1.796	*n.s.*
Topic*Text	F(3,72)=3.828	0.017
Topic*Text*Training	F(3, 72)=0.683	*n.s.*

Fig. 7 shows mean percentages of durational lengthening of the first prosodic phrases from non-initial to discourse-initial conditions.

Figure 7: Percentages of durational lengthing in the first prosodic phrase comparing discourse initial with non-initial condition for all the speakers.



In Fig. 7, we can see that three professional speakers (50%) and seven non-professional speakers (35%) increased phrase duration up to 10% in the discourse initial condition than the non-initial counterpart. Five of these seven non-professional speakers raised F_0 smaller than 1.5st (0.73 st on average) in the discourse initial condition, whereas all the three professional speakers raised discourse initial F0 greater than 2.7st (3.5st on average). It seems that non-professional speakers only lengthened duration when initiatig a discourse, whereas professional speakers increased both duration and F_0.

4. GENERAL DUSCUSSION AND CONCLUSIONS

Prosodic variation in discourse initial position has drawn lots of attention in the past. However, this study for the first time showed clear evidence that professionally trained speakers increase discourse-initial F0 relative to non-initial sentences more than non-professional speakers. However, there is no clear difference between these two groups of speakers in durational lengthening at a discourse initial position. What this seems to suggest is that extensive F0 raising of discourse-initial F0 is a desirable but not obligatory feature in reading aloud, and it is a skill that can be trained.

5. ACKNOWLEDGEMENTS

The research was supported by National Natural Science Foundation of China (Grant No. 60905062) to the first author.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2005. Praat. *http://www.fon.hum.uva.nl/praat/*

[2] Bruce, G. 1982. Textual aspects of prosody in Swedish. *Phonetica* 39, 274-287.

[3] Hirschberg, J. 2002. Communication and prosody: Functional aspects of prosody. *Speech Communication* 36, 31-43.

[4] Ladd, D.R. 1988. Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* 84, 530-544.

[5] Lehiste, I. 1975. The phonetic structure of paragraphs. In Cohen, A., Nooteboom, S.G. (eds.), *Structure and Process in Speech Perception*. Berlin: Springer-Verlag, 195-206.

[6] Nakajima S., Allen, J.F. 1993. A study on prosody and discourse structure in cooperative dialogues. *Phonetica* 50, 197-210.

[7] Sluijter, A., Terken, J. 1993. Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica* 50, 180-188.

[8] Smith, C.L. 2004. Topic transitions and durational prosody in reading aloud: production and modeling. *Speech Communication* 42, 247-270.

[9] Thorsen, N.G. 1985. Intonation and text in standard Danish. *Journal of the Acoustical Society of America* 77(3), 1205-1216.

[10] Umeda, N. 1982. F_0 declination is situation dependent. *Journal of Phonetics* 10, 279-290.

[11] Wang, B., Xu, Y. accepted. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics.*

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

[13] Xu, Y. 2005-2010. TimeNormalizeF0.praat. *http://www.phon.ucl.ac.uk/home/yi/tools.html*

[14] Xu, Y., Xu, C.X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-197.