A Special Prosodic Phrasing in Broadcasting News Programs

Yu Zou, Wei He, Yuqiang Zhang, Min Hou
Broadcast Media Language Research Center
Communication University of China
Beijing 100024, P. R. China
zouiy@cuc.edu.cn

Weibin Zhu
Institute of Information Science
Beijing Jiaotong University
Beijing 100044, P. R. China

Abstract

In order to improve and convey the semantic information perfectly and vividly through synthesized speech, a special prosodic phrasing in broadcasting news programs, namely complex prosodic phrase (C-PP), is proposed in this study. Empirical results show that the C-PP, which there is the pitch declination trend within it, a stable syntactic structure and whole semantic chunk, has a very important role for creation of announcers. Modeling on the special prosodic phrase in broadcast news can effectively develop and improve the speech synthesis in Chinese.

1. Introduction

Generally, prosodic structure and its correlations with semantic information in the spontaneous speech are very important for speech synthesis. But a speech corpus for speech synthesis demands speaker read some isolated sentence with limited speed and intonation [1]. It can hardly reflect the true of the correlation between prosody and semantics in context.

Prosodic phrase boundaries can be helpful to solve above problem because there is a strong correlation between prosodic phrase boundaries and syntactic phrase boundaries [2]. That is to say, prosodic phrase boundaries can play an important role in understanding utterance as punctuation marks do in written language. What’s more, prosodic phrase boundaries also have the ability of disambiguation.

But there is some difference of prosodic structure in broadcasting speech, our earlier investigations of presenter’s spoken language revealed that the prosodic chunk of the talking style uses the bigger unit and a more diversified syllable density change [3].

After further study, we found that there is a special phrase structure in read speech of broadcasting news programs. Its structure is similar to the bigger prosodic unit of presenter’s spoken language. This paper is intended to analyze this special one, namely complex prosodic phrase.

The paper is organized as follows: Section 2 of this paper describes the selection and pre-processing of speech data, the prosodic annotation methods. Section 3 is dedicated to analysis and discussion of the special prosodic phrase. Finally, some conclusions and outlines of our future work are given in section 4.

2. Corpus

A large-scale dynamic and diachronic broadcast media language monitoring corpus, which includes different periods of radio and TV multimodal database, is available at Broadcast Media Language Branch of National Language Resources Monitoring & Research Center, Communication University of China.

In this study, some representative speech materials were chosen for prosodic labeling. The detailed data processing and annotation method is presented as follows.

2.1. Selection of speech data

Xinwen he Baozhi Zhaiyao (News and Newspapers Summary) is a very famous broadcasting news program of China National Radio (CNR). It is a flagship news program at CNR in the past 50 years, with extensive, long-term and fixed audiences. It has a typical read speech style and high level quality of voice.

News and Newspapers Summary contains more semantic and prosodic information, speaking styles and high quality voice in real context. It is the best speech materials for our study. Therefore, 908 programs, which contain 26020 minutes speech data from January 1, 2006 to June 30, 2008, were selected for pre-processing.
2.2. Pre-processing of speech data

This program has four regular male announcers and four fixed female announcers. At the pre-processing step, the programs are segmented in accordance with the discourse of news reports. There are different length wav files because of different styles of news reports. Each wav file is aligned with its text file.

Table 1. Distribution of labeled speech data and its aligned texts

<table>
<thead>
<tr>
<th></th>
<th>Sum of files</th>
<th>Texts size (byte)</th>
<th>Length (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>main data</td>
<td>1540</td>
<td>428680</td>
<td>41349</td>
</tr>
<tr>
<td>comparable data</td>
<td>110</td>
<td>59220</td>
<td>6136</td>
</tr>
<tr>
<td>total</td>
<td>1650</td>
<td>487900</td>
<td>47485</td>
</tr>
</tbody>
</table>

According to the characteristics of their voice, we selected two female announcer’s speech materials (one announcer’s material forms the main data, and another one’s is supplemented for comparable data) as a sub corpus for annotation and analysis (the method of annotation will be presented in section 2.3). The distribution of labeled speech data is shown in table 1.

2.3. Method of prosodic labeling

After the pre-processing step, all the selected data, which segmentation, transcription & prosodic labeling (including break index, stress index and summary of emotional tendentiousness etc) was made by six transcribers, is formed an about 13 hours speech corpus to analyze for speech synthesis.

As for labeling prosodic phrase boundaries, we strictly dependent on the prosodic criteria and perception by using the wave files and their transcriptions, which use many prosodic features such as F0 contour, energy contour etc.

According to the characteristic of broadcasting news read speech, a new prosodic hierarchical structure (see Figure 1) and two different types of prosodic phrase boundaries were defined and used in our data labeling (Figure 2 shows a labeled example):

- /1 for prosodic word (PW) boundary
- /2 for normal prosodic phrase (PP) boundary
- /2* for complex prosodic phrase (C-PP) boundary, and ^ for its left boundary
- /3 for intonation phrase (IP) boundary
- /4 for intonation group (IG) boundary
- /5 for meaning expression cluster (MEC) boundary.

Figure 1. Prosodic hierarchical structure of read speech in broadcasting news programs

Figure 2. An example of label result

From figure 1, the complex prosodic phrase (C-PP), which contains nesting structure, is defined. It is a special prosodic phrase. Why is there C-PP? Because the prosodic phrase, which contains $7 \pm 2$ syllables, includes two or more prosodic words [4]. The boundaries between prosodic words must be different if it contains three or more prosodic words. There must be a nesting structure, such as A + (B + C), (A + B) + C, A + (B + C) + D, (A + B + C) + D, A + (B + C + D), A + B + (C + D), (A + B) + (C + D) and so on. Otherwise, it is just a normal prosodic phrase. The C-PP is labeled as follows:

- ^A/1+ B/1 C/2*
- ^A/1 B/2 C/2*
- ^A/1+ B/1 C/2 D/2*
- ^A/1 B/1 C/2 D/2*
- ^A/1+ B/1 C/1 D/2*
- ^A/1+ B/1+ C/1 D/2*
- ^A/1 B/2 C/1 D/2*
- ......

(The marker of “/1+” shows that the boundary is a little different to the normal PW boundary.)

3. Results analysis and discussion

In our corpus, the number of each boundary type is not equal. Among them, PW boundary accounts for
53% of the total boundaries, PP is 35%, IP is 8.2%, IG is 2.1% and MEC is 1.5%. Table 2 shows the details.

<table>
<thead>
<tr>
<th>Type</th>
<th>Times</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW</td>
<td>54062</td>
<td>53.2</td>
</tr>
<tr>
<td>PP</td>
<td>35542</td>
<td>35.0</td>
</tr>
<tr>
<td>IP</td>
<td>8296</td>
<td>8.2</td>
</tr>
<tr>
<td>IG</td>
<td>2091</td>
<td>2.1</td>
</tr>
<tr>
<td>MEC</td>
<td>1561</td>
<td>1.5</td>
</tr>
</tbody>
</table>

According to table 2, we find that the PP boundary is the largest one except PW. By analyzing the composition of the PP boundaries, the C-PP boundary is 16.8% of the total PP boundaries. This reveals that the C-PP has a very important role for creation of announcers in broadcasting news program.

In this case, what acoustic-phonetic features does the C-PP have? Cao pointed out that prosodic phrase has a relatively stable pattern of prosody, that is, the pitch declination trend within prosodic phrase. It also has a relatively stable pattern of phrase accent, which is associated with syntactic structure [4]. Another study argues that there is the pitch declination of bottom lines in the prosodic phrases of broadcasting style [5].

Figure 3. An illumination of C-PP

Figure 3 shows there are two C-PP in the sentence, for example:

中国/1 全国/1 人大/2*
zhong1guo2 quan2guo2 ren2 da4, means “Chinese National People’s Congress”

同/1+ 希腊/1 议会的/2 关系/2*
tong2 xi1 la4 yi4 hui4 de0 guan1 xi0, mean “the relationship with Parliament of Greece”

From above two examples of C-PP, the pitch declination trend within prosodic phrase is obvious despite small resetting at position of marker “/1+”. On the other hand, it has a stable syntactic structure and whole semantic chunk. This reveals that semantic information is the important factor to announce news by announcers.

4. Conclusions

This paper proposes the special prosodic phrase, namely complex prosodic phrase (C-PP), which contains a nesting structure in broadcasting news. From the analysis of the result, we find that the C-PP, which there is the pitch declination trend within it, a stable syntactic structure and whole semantic chunk, has a very important role on creation of announcing news by announcers. Modeling on the special prosodic phrasing in broadcast news can effectively develop and improve the speech synthesis based on semantic computing in Chinese.

Acknowledgements

This work is supported by 863 Hi-Tech Research and Development Program of China (No. 2007AA01Z198), the Department of Science and Technology at Ministry of Education (No. 107118), and the National Language Resources Monitoring and Research Center (YZYS08-07, YZYS08-09).

References


