The Design of Corpus for Interrogative Speech Synthesis

Yanhong Wu (1)     Jianhua Tao (1)       Jilun Lu (3)

(1)(2)Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100080
(1)(3)Tianjin Normal University, Tianjin, China
{yanhwu (1), jhtao (2)} nlpr.ia.ac.cn

Abstract

In recent years, the study on more expressive speech synthesis has been a hot topic. The paper will focus on how to establish an interrogative corpus to get the fluent and natural synthetic interrogative sentences. The main characteristics of the corpus are based on the natural situational conversations and dialogues. The corpus consists of both interrogative and declarative utterances. Interrogative utterances are divided into five types: the Yes-No questions, Wh-questions, tag questions, choice questions, and Positive-Negative questions. We select different kinds of sentence structures with different interrogative intonation types, which are meaningful in syntax and semantics. When choosing in terms of these different types, we take the information of prosody and rhythm into consideration.

The paper consists of three parts. In the first part, we analyze the original corpus based on the various sentence structures, which has been tagged into prosodic phrase and intonation phrase. From here, the information on pitch contour and duration of interrogative utterance is got. Thus the tone-carrier unit of the interrogative ones can be determined, and then we focus on the boundary tone’s influence on the intonation, and the prosodic words’ influence on the boundary tone in terms of these five types. The second part is about how to design. In order to synthesize natural and fluent interrogative utterances, the corpus selected in this paper is oriented to the real situation. Thus it mainly derives from natural situational conversations and dialogues. Moreover, we design the relative declarative sentences for the comparison with the interrogative sentences, which are read by the same speakers. The topic chosen in the corpus covers many aspects in our daily life such as the politics, culture, education and recreation etc. Then we tag the text into prosodic words, prosodic phrases and intonation phrases. Later the comparison of interrogative and declarative sentences is made, and we compute pitch contour and duration of these two kinds of intonation and get the result that the last prosodic word carries the interrogative information, and this word influences the adjacent words and phrases in terms of such a rule that the more near it, the higher its pitch contour. Lastly, we choose the text with prosodic information. During the process of choosing, about the segment, we mainly take the syllable into consideration; about the prosody, pitch contour, duration, and the rhythm of the sentence are carefully studied. In the last part, we focus on the results of the analysis on the chosen text, especially on pitch contour and duration, and its influence on the adjacent units. The results are used in the synthesis of the interrogative sentences, and we can get the fluent and natural synthetic speech.
I Introduction

This paper introduces the construction of interrogative utterance corpus and tries to synthesize the fluent and natural interrogative speech. At first, we explain why the corpus is being established, and the analysis results of the original interrogative corpus. After doing this, we describe how to collect the corpus, a few characteristics of the corpus, the procedure developed for the prosodic transcription of the corpus, and the process of analyzing the corpus. Lastly, we discuss the experimental result, and then a conclusion about the corpus is made.

II The reason for the construction of corpus

The construction of corpus is born out of an interest in Chinese intonation especially in interrogative one, the tone-carrier unit of interrogative information, and their performance under the influences of tone. On making sure all these points, we can synthesize natural and fluent interrogative utterances.

Chinese is known as a typical tone language. Its intonation is extremely complex. The study of Chinese intonation is of great interest to researchers throughout the linguistic circle. Not only the linguists but also the speech engineers pay a great attention to it. Linguists have classified intonation into several types such as declarative, interrogative, and commanding one. This paper focuses on the construction of interrogative corpus, the tone-carrier unit of interrogative ones and that unit’s influences on the adjacent units or the rhythm.

Shen (1985, 1994) made an experiment of intonation, and generalized that pitch high line which is concern with semantic stress and pitch low line which is concerned with rhythm, are the two characteristics of intonation. The interrogative utterances are observed as gradual falling of pitch high line and moving up of pitch low line. Lin(2003) made research on pitch accent and boundary tone in Chinese intonation through acoustical analysis and listening test of echo questions in read speech and yes-no questions in spontaneous speech. He found out that the register of the ending-point (or the slope) of the F0 curve in boundary tone plays a more important role than the register of its starting-point in differentiating between question and statement. He pointed out that information about question and statement is mainly carried by the last one or two syllables without neutral tone in the final prosodic word of the intonation phrase. Shi( 2002) made a study on the differences between declarative and interrogative intonation in Chinese, which can be accounted for by two mechanisms: an overall higher phrase curve for interrogative, and higher strength values of sentence final tones for interrogative. Moreover, the phrase curves of the two intonation types tend to be parallel and boundary tones are not necessary for modeling the difference between them in Chinese.

All of them try to find pitch contour difference between declarative and
interrogative utterances, and find the characteristics of interrogatives. Among them, some try to use this result and make models to synthesize natural and fluent utterances. But the result is still not satisfying as what we expected. Our corpus is created with consideration of all these, to find the differences on pitch contour and duration between declarative and interrogative utterances, and to serve a wide range of needs in the speech sciences community, it is somewhat unique.

III Corpus construction

3.1 The original corpus

We choose 590 interrogative sentences from the conversion with a total of 7200 syllables. Among them, there are 256 types of common interrogative sentence structures. Four talkers (two men and two women) are from the Northern area, and all speak standard mandarin. We labeled the part of speech, and tagged them into prosodic word, prosodic phrase, and intonation phrase. All are done automatically and with artificial modification. In it, we have 140 yes-no questions (with interrogative mark), 294 wh-questions, 40 choice questions, 100 yes-no questions (without interrogative mark), and 16 positive-negative questions. Pitch contour and duration of these interrogative sentences are analyzed with the PRATT software. We get that pitch contour of the last syllable in most yes-no questions with interrogative modal word is 10-40Hz higher than declaratives for the influence of the modal word. On the other hand, duration is about 40-200ms shorter than declaratives. The mean pitch value of questions is 6Hz higher than statement. Moreover in the yes-no questions without interrogative mark, the last prosodic word bears the interrogative information, and its pitch value is 20-50Hz higher. Duration is 10-100ms shorter. The mean pitch value is 34Hz higher. It is shown in figure1.

![Figure 1](image)

We design 200 yes-no questions and categorize them into five types: noun + adjective (na), noun + noun + verb (nnv), noun + verb + adverb (nvd), noun + verb + noun (nvn), noun + verb + noun + adverb (nvnd). Each type is 40. The corresponding statement is 200. These 200 questions are simple sentences. It is designed according to different syntactic structures as we pointed out above. Two men and two women, who speak standard Chinese, are asked to record the sentences. We analyze it with the same method used above, the result is the same as the former one. Thus it is observed that in the yes-no questions the last prosodic word carries the interrogative information. However pitch contour of the adjacent prosodic words (the first three
words before it) is raised under the influence of the last word. In questions with interrogative mark, the mark carries the interrogative information. Thus pitch contour of the last two syllables is influenced by interrogative mark, and is raised.

3.2 The construction of the new corpus

Most of interrogative sentences in the original corpus are experimental sentences. Talkers are asked to read it with strong emotion. Thus the greatest shortage of the corpus is lack of naturalness.

The new corpus is got from the real situational dialogues and conversions, which cover every aspect in our daily life such as: politics, education, amusement and sports. During choosing, about the segment, we mainly take the syllable into consideration; in case of the prosody, pitch contour, duration, and the rhythm of the sentence are carefully studied. In order to make comparison with declaratives, we design the same number of corresponding declarative utterances and interrogative ones without interrogative mark. The same speakers are asked to record the corpus in a speech laboratory. Cubase is used to record. One sample is the man, and the other is the woman, who all speak standard Chinese and read it in a natural, fluent and smooth speed. The phonetic software Cool Edit Pro is used firstly to cut the needed interrogative sentences out of the wave file, phonetic software Praat is used to deal with the wave files including phonetic annotation, phonetic drawing and data abstracting. Before phonetic drawing and data extraction, some pitch contour problems are modified artificially at first.

Corpus is tagged into prosodic word, prosodic phrase, and intonation phrase, and part of speech of each word is labeled. Four graduate students, who are from Tianjin Normal University and Chinese Academy of Chinese, have no hearing handicap and know nothing about the author’s intent to compute the rhythm of the interrogative and declarative utterances, are asked to label the pause among prosodic word, prosodic phrase, and intonation phrase.

3.3 The analysis of corpus

Interrogative utterances we get from real situation conversions and dialogues have 4750 syllables. Among them, there are 20 yes-no questions, 20 wh-questions, 20 choice questions, 20 tag questions and 20 positive-negative questions. The purpose of our analysis is to make sure the tone carrier unit, and find its influences on adjacent units and the differences between interrogative and declarative utterances through the comparison between them.

3.3.1 The analysis of the yes-no questions

In questions with interrogative modal word, pitch value of the first adjacent three syllables before modal word is raised by 6-15 Hz under its influence. The mean pitch value of the whole sentence is 7 Hz higher than the statement. However, duration of the last prosodic word in the interrogative sentence is 135 ms shorter than the declarative one. In interrogative utterances without interrogative mark, we verify Lin (2003)’s result that the last prosodic word carries the interrogative information except for the neutral tone. Its pitch value is about 31 Hz higher than that it in declarative one. However, pitch contours of the adjacent units are also raised. Moreover, the mean pitch value is about 23 Hz higher, but its duration is 86 ms shorter. The figure is as
3.3.2 The analysis of wh-questions

Wh-questions’ interrogative information is carried by the interrogative words such as *shen me* (what), *zen me* (when), *na li* (where), *shui* (who), and *wei shen me* (why). The following figure is an example of the interrogative word in the middle, and we cut the interrogative word “*duo shao*” (how many) which should lie between the third and fourth syllable. We can see that pitch value of the last three syllables in the red line is much higher than it in the cyan line and yellow line for the influence of the modal word. The second, third and fourth syllable in wh-questions are mostly the same for the interrogative word, but they are much higher than that in the statement. Duration of the last syllable is about 150 ms shorter than that in the statement. The nearer the interrogative word and modal word is, the shorter duration is.

3.3.3 The analysis of choice questions

Through the comparison of choice questions with declarative ones, we use the last three syllables of the first half sentence and of the next half one. We get that pitch contour in the first half sentence in choice questions is quietly different from it in declaratives, and that in the next half one is a little different. Duration of the last prosodic word in questions is about 80ms shorter. Because there are some problems with choice questions with modal word, only questions without it are compared. We can see from the following figure.
3.3.4 The analysis of positive-negative questions

On comparison of this kind of question, we cut the words showing interrogative information such as “shi bu shi” (isn’t it) which should be placed between the first and second syllable. Pitch contour of the questions without modal words have no great distinction from corresponding declaratives; only the former’s duration is 20-40ms shorter than the latter one. By making comparison between positive-negative questions with modal word and that without it, pitch contour of the last prosodic word in these two sentences is totally different, the former one is about 28 Hz than the latter one, and duration is about 65 ms shorter. Moreover, the former’s pitch value is also 36 Hz than that in declaratives. Duration is about 120 ms shorter.

IV The discussion of experimental result

Through the analysis we made above, we can find the characteristics of interrogative intonation, and its difference with declarative one.

1. Interrogative intonation with modal word VS interrogative intonation without it.

Modal word (ma, ni) and interrogative word (why, what etc) carry interrogative information and raise pitch contour value of the first three adjacent syllables in all types of interrogative utterances with such a rule that the nearer it is, the higher the pitch value is. The last prosodic word in yes-no questions without modal word carries the interrogative information except for the neutral tone, and its pitch value is higher
and duration is longer than them in questions with modal word. In addition, its mean pitch value is also higher. The interrogative information of the choice questions is mainly shown in the first half sentences.

2. The interrogative intonation with the corresponding declarative one

Through the comparison made between these two kinds of intonation, it is observed that the mean pitch value of declarative intonation is much lower than that of all types of interrogative one, but duration is longer.

Moreover, the rhythm of prosodic words in questions especially that with modal word or interrogative mark, is much more compact than that in statement under the influence of the modal word (ma, ni) and interrogative mark such as you mei you( isn’t it) in positive and negative questions, shen me(what) in wh-questions. In yes-no questions, rhythm of the last prosodic word is quietly different from that in statement. It is compacter, and its duration is shorter, which is mainly shown in two aspects: the penultimate syllable’s duration in questions is longer than that in statement, duration of the last syllable is shorter.

V Conclusion

If the analysis result is applied in the synthesis of interrogative utterances, the fluent and natural synthetic utterances can be got. However, some shortcomings also exist. For example, we only consider the short interrogative sentences and the complex ones are rarely studied; the stress in it is not concerned about. We should take all these attributes into consideration in our next step.

VI References

Yuan, Jiahong, Shih, Chilin etc. 2002. Comparison of declarative and interrogative intonation in Chinese