The International Journal Of Engineering And Science (IJES)
|| Volume || 5 || Issue || 4 || Pages || PP -08-14 || 2016 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

A study on the realization of pitch in Mandarin Chinese intonation

Maolin Wang
College of Chinese Language and Culture, Jinan University, Guangzhou, China

----------**ABSTRACT**----------
This study explores the pitch realization of utterances in spontaneous Chinese speech, and it is found that there is a great variation of the pitch ranges of intonation phrases. Most of them are about half of the speaker's maximum pitch range, and the pitch register is usually in the lower part, which is due to the principle of economy. In an intonation phrase, owing to the mechanism of declination, tones always drop to a lower scale compared to the previous ones. Pitch tends to be high for intonation phrases at the initial position of a new topic, and low for those at the end position. Final lowering is common for intonation phrases, leading to increased pitch range for the final prosodic word.

Keywords *- Intonation, tone, pitch range*

Date of Submission: 22 March 2016 Date of Accepted: 05 April 2016

I. INTRODUCTION

It is a well-known that, in all languages, no utterance is strictly produced in a monotone, and every utterance shows pitch variation in some way. Such variation in pitch, impressionistically described as rising or falling, are caused by changes in fundamental frequency, which is the acoustic feature of speech signal determined by the rate of vibration of the vocal folds and is recognized as pitch. In languages like English or Italian, the entire pitch contour is specified at the intonation phrase level by means of interplay between metrical structure, syntax and pragmatics. These factors control where pitch changes will occur and of what changes there will be. In tone languages, such as Mandarin or Thai, syllables are lexically specified with tone and tonal changes which affect lexical meaning, and in pitch accent languages, such as Japanese or Swedish, tone operates in a similar way, except that only one syllable in each word is lexically specified with tone. In tone languages, the naturalness of speech is controlled by, including timing, tone and intonation. The acoustic manifestations of duration, tone and intonation, as well as the manner how they interact with each other, depend highly on the language.

Intonation is the global pitch contour of an utterance, and it is shown that, for a given semantic interpretation, the shape of the intonation contour is possible to vary substantially owing to the segmental material with which the semantic meaning is uttered. Such variations are not random, but controlled by the overall prosodic structure of the sentence with which the pitch contour is associated, including the number of words and the position of stresses. The pitch contour does not have a fixed meaning either within a language or across languages. Within a language, the interpretation of the intonation may depend on lexical information and other choices accompanying the application of the melody. Across languages, the differences can be arbitrary. A study on intonation should be able to obtain these properties: it should be capable of explaining the relationship between intonation contour and the meaning, making generalizations from surface pitch contour with enough predictive power in order to generate new pitch contours of the same basic melody to fit new utterances of different lengths and structures.

There is substantial variation in the ways researchers treat the basic properties shown in intonation contour. Many researchers have taken intonation pitch contours as gestalts or configurations, that is, as holistic pitch movements. The pitch movements encompass entire utterances and there is a uniform meaning with it. In other models, melodies are treated as being consisted of primitives of some sort or other. The primitives are taken to be either local configurations or dynamic tones, such as local rising and falling, or level tones, such as high, mid or low. They propose analyses based on the decomposition of intonation events into smaller elements. Some use dynamic tones as the primitives of intonation structure, and some advocate the use of level tones.

In recent years, there have been a lot of studies on Mandarin Chinese intonation and prosody, which include the classification and labeling of intonation[1], the relationship between lexical tone and sentential intonation[2-3], the pattern of statement and question and the related boundary tones[4], the acoustic realization of prosodic boundaries in Chinese[5-6], the prosodic structure of utterance and the construction of the top and bottom pitch lines[7], the prosodic feature and the acoustic realization of dialogue[8], the prediction of Chinese prosodic structure on the basis of syntactic information[9], the property of large information units such as sentences and

paragraphs in text and the acoustic clue at the boundary[10], the predicting model for generating pitch contour of connected speech[11], the analysis of the prosodic feature of multilingual emotion and the recognition[12], the automatic prediction of stress of natural style sentence[13], the comparative study on the pitch and duration of tones in citation form and in sentences in Mandarin and Taiwan Chinese[14], and the study on the downstep of Chinese intonation on the basis of designed sentences with different tone combinations[15]. These studies have greatly deepened our understanding on Mandarin Chinese intonation and prosodic structure, however, there is not much research work on the pitch of natural speech, as the above-mentioned studies are all based on read speech. Wang and Lin[16] give a preliminary analysis on the pitch of natural Chinese dialogue, with a discussion on the maximum and minimum pitch and the pitch range of common speakers, and the pitch range and pitch register of intonation phrase (IP). A preliminary understanding on the pitch of dialogue is obtained from Wang and Lin's work, however, there is still much work to do on the study of pitch of spontaneous speech.

The study on intonation and prosody is of great significance in the field of linguistics and speech technology. Intonation, in a narrow sense, is the tonal pattern of an utterance[2], i.e. the pitch pattern. Therefore, a thorough and comprehensive understanding to the pitch of spontaneous speech is quite useful to both phonetic study and to the improvement of the naturalness of speech synthesis, and it is also helpful in improving the correctness of speech recognition. For the purpose of improving the naturalness of synthesized speech so as to represent the cadence of speech is a major task in nowadays engineering project, and the fundamental work for this is to have a comprehensive understanding on the pitch pattern of spontaneous speech. In this paper, based on a relatively large corpus by multi-speakers from a natural conversation corpus, an analysis is done on the pitch range, pitch register and the classification of intonation phrase, and the cause for the variability of pitch range and pitch register of intonation phrase is also discussed.

II. STUDYING MATERIALS

The studying materials in this experiment are taken from a conversation corpus, in which there are about 80 hours' recordings. Most of the recordings are free dialogue between two speakers, and there is a great variety of topics, including everyday life, entertainment, working, social life, food, sports, etc. The dialogues have been transcribed in Chinese. Examples are shown in (1) and (2) below.

Syllable and sentence type are annotated on the corpus in the phonetics laboratory of Jinan University, and pitch is extracted by Praat (www.praat.org), with correction done manually by referring to the narrow band spectrogram. Since the speech material is spontaneous dialogue, there are some repetition, hesitation, interruption and fillers such as 'en', 'zhege' (this) and 'nage' (that) in the speech. These kinds of materials are discarded in this study, as it is aimed to analyze the pitch representation in fluent speech at normal condition.

III. RESULTS

For the purpose of imitating the pitch variability in natural speech, the pitch is usually adjusted in speech synthesis, and the basis of the adjustment is the reliable study of the pitch representation in actual speech. Wang and Lin[16] report that in spontaneous speech, the average pitch ranges from 85 to 210 Hz for male and 145 to 365 Hz for female, so there is a great difference between the pitch ranges of male and female, with male 90 Hz narrower than female. However, when converted to semitone, the average pitch ranges for male and female are quite similar to each other, both about 16 semitones, so semitone will be used for measuring the pitch of intonation phrase in this study.

3.1 Pitch range and pitch register of intonation phrase

There has been much research work done on the prosodic structure in Mandarin Chinese in recent years, and it is found that syllables in an utterance can be grouped into prosodic units, which will further form hierarchical prosodic structure[5-8]. Based on these studies and the characteristic of spontaneous speech, prosodic annotation is done on the corpus. Three levels of prosodic unit, including prosodic word, prosodic phrase and intonation phrase, are annotated, and 1358 intonation phrases are used in this study. Intonation phrase is an important prosodic unit in an utterance, and its pitch register and pitch range vary a lot in actual speech, so its pitch representation need to be analyzed. The pitch pattern of intonation phrase will be discussed, and that of the lower level prosodic units will not be analyzed in this study.

Sinclair and Coulthard[18], based on a systematic analysis on the classroom speech, put forward a set of hierarchical structure for discourse analysis, which includes lesson, transaction, exchange, move and act. An exchange is a go-around in interaction, including initiation, response and follow-up. A move is a continuous utterance of one speaker, and an act roughly corresponds to a clause. The speech material of this study is face to face conversation, which, similar to classroom speech, also includes exchanges consisting of initiation, response and follow-up. The following example is taken from the corpus.

(1) A: Ni shi Xinjiang de ma? (Are you from Xinjiang?)
B: Ao, dui. (Oh, yes)
A: Xinjiang you shenme haoyan de defang ma? (Are there any places of interest in Xinjiang?)

Example (1) is an exchange, which consists of initiation, response and follow-up. However, there are cases where there is no follow-up, only initiation, response. For example,

(2) A: Ni ruguo qu waiguo de hua xiang qu nage guojia? (Which country do you prefer if you go abroad?)
B: Bu zhidao. Wo hai mei kaolu guo zhege wenti. Ni shuo guojia de hua, qishi youxie shi shihe luyou de (I don't know. I have not thought about this. Regarding countries, some of them are actually suitable to travel there).

It is demonstrated that the pitch of an intonation phrase is affected by its position in an exchange, so intonation phrases are annotated with exchange position in the text, with those at the beginning labeled 'start' (S), those in the middle labeled 'intermediate' (I), and those at the end labeled 'end' (E). For example, in (2) the utterance of 'A' is labeled 'start', 'some of them are actually suitable to travel there' labeled 'end', and those in the middle labeled 'intermediate'. Examination will be done on the pitch range and pitch register of intonation phrase.

In speech synthesis, the pitch range and pitch register of an intonation phrase may need be adjusted to improve naturalness, so research work is necessary on it. According to Wang and Lin[16], when pitch is converted to semitone, the pitch ranges of male and female are quite similar to each other, so the data of male and female are computed together here. F_0 is converted to semitone by the following formula,

$$St = 12 \times \log_2\left(\frac{F_0}{F_{0\,\mathrm{ref}}}\right) \qquad (1)$$

F_0 is the fundamental frequency at any point, $F_{0\mathrm{ref}}$ the reference F_0, which is 27.5 Hz here, and St is the semitone got from the computation. The pitch range of an intonation phrase is obtained by minusing its minimum semitone by its maximum semitone.

There are four tones in Mandarin Chinese, among which Tone 1, whose pitch contour is flat, is a high level tone, and it is hard to calculate the pitch range of an intonation phrase if all the syllables in it are Tone 1. There are 38 such cases in the corpus, which are excluded when computing pitch range of intonation phrase. Fig. 1 shows the frequencies of the intonation phrases of various pitch ranges, and it is displayed that there is great variability in the pitch range of intonation phrase, with the maximum more than 20 semitones, about two octaves, and the minimum less than five semitones.

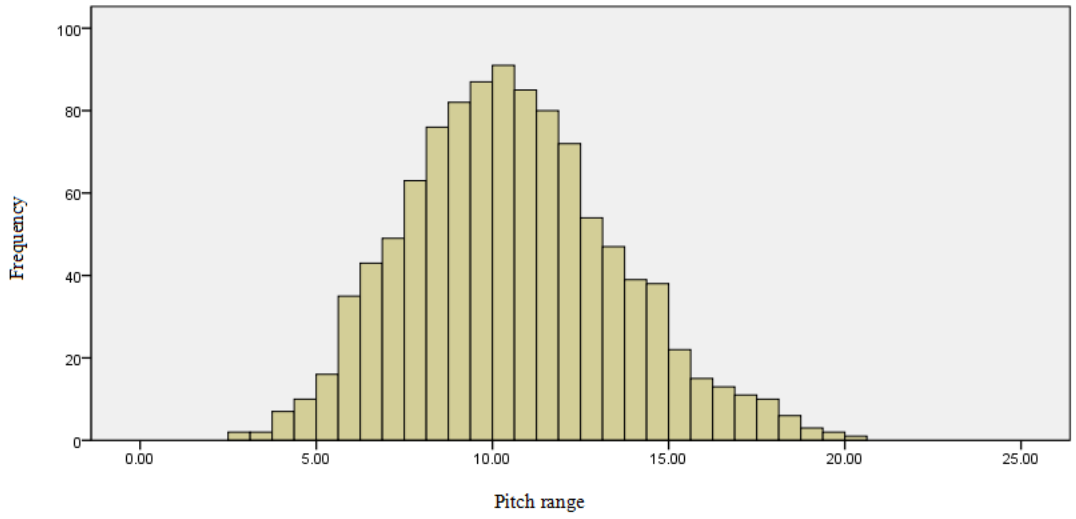

Fig. 1 Distribution of pitch ranges of intonation phrases

As for pitch register, to put it in simple words, it is the position of pitch, and the pitch register of an intonation phrase can be represented by different parameters, such as the top point, bottom point, median and the average. Here it is presented in a different way from Wang and Lin[16], that is, it is represented by the bottom point. The bottom point of pitch is comparatively stable in speech, so it can be used to stand for the pitch register of an intonation phrase. That is,

$$Rg_i = St_{min(i)} - St_{min(s)} \quad (2)$$

$St_{min(i)}$ and $St_{min(s)}$ are the pitch minimums for the intonation phrase and the speaker respectively, with Rg_i the pitch register of the intonation phrase, and the units for the three parameters are all semitones.

There is a normal pitch range for people's utterances, and it varies among different speakers. A speaker's pitch range is the difference between his maximum and minimum pitches in a dialogue, and the variability of the speaker's pitch range may lead to the difference in the pitch ranges of intonation phrase, that is, if a speaker's pitch range is large, the pitch ranges of his intonation phrases will also be large, and if his pitch range is small, the pitch ranges of his intonation phrase will also be small. In this study, in order to eliminate the pitch range difference among speakers, an approach of normalization is applied when computing the pitch range and pitch register of intonation phrase. The formulas are as follow,

$$Ra_n = \frac{Ra_i}{Ra_s} \quad (3) \qquad Rg_n = \frac{Rg_i}{Ra_s} \quad (4)$$

Ra_i and Rg_i are the pitch range and pitch register of intonation phrase respectively, Ra_s being the speaker's pitch rang, the unit of Ra_i, Rg_i and Ra_s being semitone, and the results of Ra_n and Rg_n got from the formulas are the normalized pitch range and pitch register of the intonation phrase. Ra_s is the speaker's pitch range represented in a whole dialogue, and Ra_i and Rg_i are only the pitch range and pitch register of any intonation phrase in the dialogue. There are from a dozen of to scores of intonation phrases in a dialogue, so Ra_i and Rg_i are certainly smaller than Ra_s, therefore the values of Ra_n and Rg_n will range from 0 to 1. The result is shown in Fig. 2 and Fig. 3.

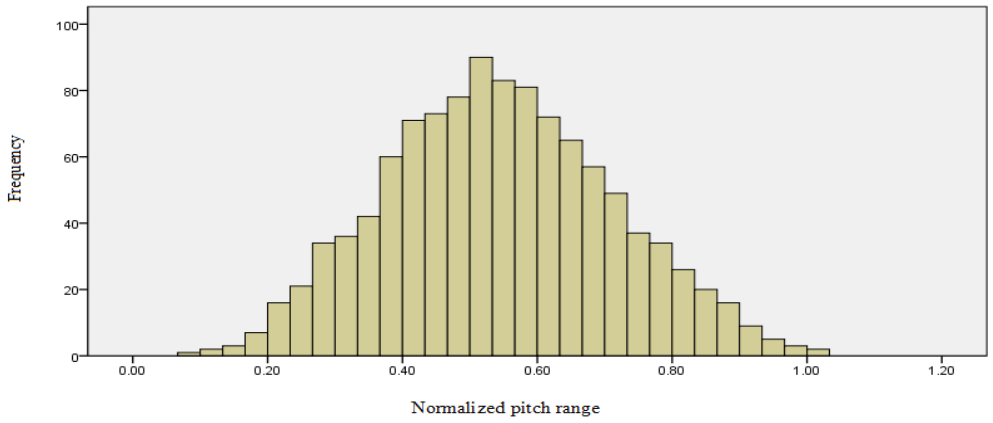

Fig. 2 Distribution of normalized pitch ranges of intonation phrases

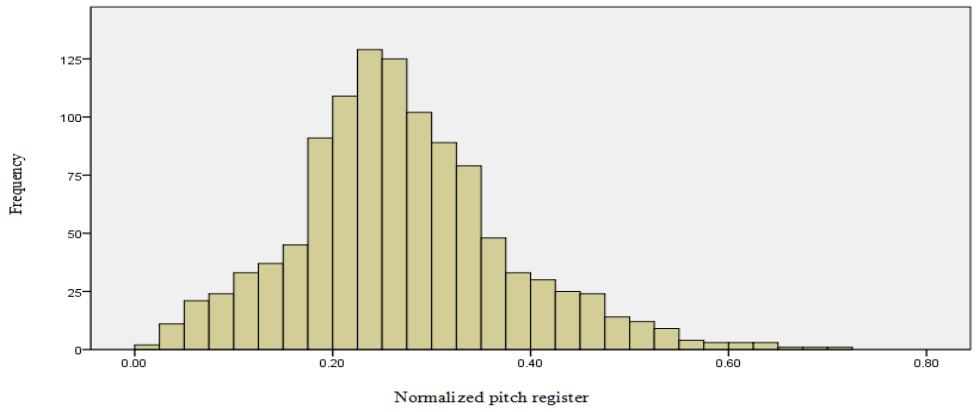

Fig. 3 Distribution of normalized pitch registers of intonation phrases

The pitch range of a intonation phrase is the difference between its top and bottom pitch points, and the pitch register is its bottom pitch point, therefore, they are negatively correlated to each other, with the coefficient being -0.62. For some of the intonation phrases, their bottom pitch points reach the bottom points of the speaker, resulting in normalized pitch range as 0, while for some others, their top pitch points reach the top point of the speaker, resulting in $Rg_n = 1 - Ra_n$, with the normalized pitch registers of other IPs range for 0 to $1 - Ra_n$.

3.2 Pitch realization within intonation phrase

3.2.1. Intonation phrase with 2 prosodic words

In this section, the top, bottom pitch and the range of prosodic word (PW) will be analyzed. Fig. 4 presents the three parameters of pitch with 2 prosodic words, with PW1 as the first prosodic word, PW2 as the second one. One-way ANOVA analysis shows that there are significant difference between the first and the second prosodic words, top pitch: $F(1, 395) = 10.5$, $p < 0.01$; bottom pitch: $F(1, 395) = 56.9$, $p < 0.001$; range: $F(1, 395) = 26.8$, $p < 0.001$. For top and bottom pitch, the value of the first prosodic word is larger than the second, while for pitch range, the value of the second prosodic word is larger than the first.

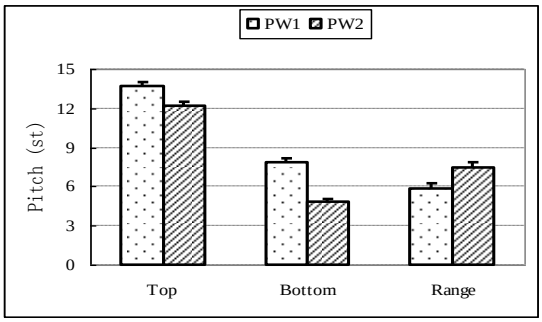

Fig. 4 The top, bottom and range of pitch of intonation phrase with 2 prosodic words

3.2.2 Intonation phrase with 3 prosodic words

Fig. 5 shows the top, bottom and range of pitch with 3 prosodic words, with PW3 as the third prosodic word. It is shown from one-way ANOVA analysis that there are significant difference among the three prosodic words, top pitch: $F(2, 502) = 15.3$, $p < 0.001$; bottom pitch: $F(2, 502) = 53.7$, $p < 0.001$; range: $F(2, 502) = 12.5$, $p < 0.001$. For top and bottom pitch, the value of the first prosodic word is larger than the second, and the second prosodic word is larger than the third. For pitch range, the value of the last prosodic word is larger than the earlier two, while there is no significant difference between the first and the second.

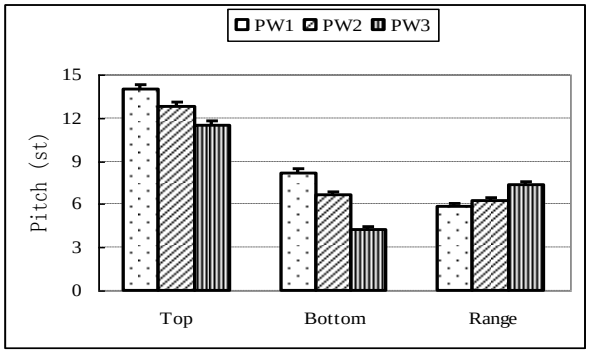

Fig. 5 The top, bottom and range of pitch of intonation phrase with 3 prosodic words

From Fig. 4 and Fig. 5 it is can be also seen that, whether the utterances are long or short, there is an overall downward trend for the pitch of the utterances, with only minor exceptions at the first part of longer ones. The declination is less dramatic in the earlier part of the sentences. The pitch range of the last prosodic word is usually large, and the drop of the minimum pitch is also large.

IV. DISCUSSION

In this study, the pitch realization of spontaneous Chinese speech is analyzed, and it is found that there is a great variation of the pitch ranges of intonation phrases. The pitch range of an intonation phrase can be as wide as more than twenty semitones, and can also be as narrow as less than five semitones. Most of them are around ten semitones. When compared to the pitch range of speaker, most of the intonation phrase is about half of the maximum range, which is due to the principle of economy. In face to face conversation, it is not necessary for the speakers to use their full pitch range. In this way, they may save their energy, and leaving space for special occasion. Under special occasion, say, when they want to express very excited emotion, they will raise their pitch, using their full pitch range, to show their affective meaning. Therefore, in usual case, they will only use half their pitch range.

The pitch register of intonation phrase is also analyzed, and it is shown that, in some cases, the pitch register reaches the upper part of the speaker's full pitch range. These usually happen when the speaker is talking about something exciting. The studying materials are natural, free talk, and sometimes the speaker will have different emotions. In affective expression, there will be much variation in pitch. In other cases, the pitch register may reach the bottom of the speaker's pitch range. This is also due to the need of affective expression, when the speaker is talking about some less exciting. For most of the intonation phrases, the pitch register is around the lower part of speaker's full range. This is also owing to the principle of economy. In this way, the speaker will save energy.

There is a correlation between pitch realization and discourse position, i.e., the pitch representation of intonation phrase varies at different exchange positions. It is observed that, pitch tends to be high for intonation phrases at the start position, and low for those at the end position. There is functional reason for this. When one initiates a new topic, the information is new. In this case, he wants to arouse the listener's attention, and usual way is to raise the pitch. At the end of a topic, there is nothing new, the speaker usually lower his pitch.

The pitch realization of prosodic words within intonation phrase is investigated, and it is found that there is declination for pitch. In an intonation phrase, tones always drop to a lower scale compared to the previous ones. In regard to the overall pitch range of the intonation phrase, it is found that compared to intonation phrase with 2 prosodic words, the overall pitch ranges of longer intonation phrase is large. As for the pitch ranges of prosodic words within the intonation phrases, it is shown that the pitch range of the last prosodic word is the largest. For intonation phrase with three prosodic words, there is no significant difference between the first and the second one.

Liberman and Pierrehumbert [20] put forward the Gradient model of declination, which defines dropping patterns as a gradual lowering toward an abstract reference line, or asymptote. Their approach of pitch assignment displays an exponentially declining curve in which each step down is proportionally the same as the forgoing one in terms of its distance from the reference line. Later drop intervals are gradually smaller than earlier ones, and tend to get vanishingly small as the reference line is approached. This approach can be called a 'soft-landing' model of declination implementation as it demonstrates a curve similar to that of a plane gliding smoothly down to a landing strip.

The soft-landing model is proposed on the basis of the top line of the pitch contour. In this study, the pitch range, which is the difference between the top and bottom point, is also investigated, and it is found that the pitch range of the earlier prosodic words are small, with that of the last the largest, which is due to the 'final lowering' effect. Final lowering, the lowering of pitch at the end of an intonation phrase, has been observed in many languages, like Spanish and Yoruba.

It is found that the pitch ranges of prosodic word near the middle of the intonation phrase tend to be the smallest. This is because words in this part are less prominent in communication. Words near the end of the intonation phrase may have 'final lowing' effect, which have demarcative function, i.e., marking the end of a sentence. As a result, the pitch range of the final prosodic word is the greatest. The first prosodic word is at the beginning of a sentence, with the function of attracting the listener's attention, so it is usually more prominent, and its top pitch usually high. If the words near the middle of an intonation phrase are not under focused condition, they will be the least prominent in communication, and their pitch ranges will be the least large.

V. CONCLUSION

There is a great variation for the pitch range and pitch register of intonation phrases in spontaneous speech, with the range of some as small as less than five semitones and some as great as more than twenty semitones. In face to face conversation, it is not necessary for the speakers to use their full pitch range. In this way, they may save their energy, and leaving space for special occasion. The pitch range for intonation phrases at the start position of a new topic tends to be large, while that at the end position tends to be small. The pitch range and pitch register of intonation phrases in for spontaneous speech is analyzed in this study, and this is supposed to be of some help to the construction of pitch in speech synthesis. At the first stage in synthesis, the pitch range and pitch

register of intonation phrases can be set according to the average values, and then they can be adjusted based on their position. With much influence from linguistic and paralinguistic factors, the pitch realization is quite complicated for spontaneous speech.

ACKNOWLEDGEMENTS

The research reported here was carried out in the Laboratory of Phonetics of Jinan University. This work was partially supported by the Humanity and Social Sciences Research Fund of China Educational Department, Grant No. 14YJA740037.

REFERENCES

[1] Shen, J. The classification of Chinese intonation and a preliminary discussion on annotation. *Application of Chinese Linguistics and Characters*, 1998; 25(1): 102-104
[2] Cao, J. The relationship between tone and intonation in Chinese. *Chinese Linguistics*, 2002; 3: 195-202
[3] Lin, M. Chinese intonation and tone. *Application of Chinese Linguistics and Characters*, 2004; 3: 57-67
[4] Lin, M. The mood of statement and question and boundary tone. *Chinese Linguistics*, 2006; 4: 364-376
[5] Hu, W., Xu, B., Huang, T. An experimental study on prosodic boundary in Chinese Mandarin. *Journal of Chinese information Processing*, 2002; 16 (1): 43-48
[6] Wang, B., Yang Y., Lu, S. Acoustic analysis on prosodic hierachical boundaries of Chinese. *Acta Acustica*, 2004; 29 (1): 29-36
[7] Lin, M. Prosodic structure and lines of F_0 top and bottom of utterances in Chinese. *Contemporary Linguistics*, 2002; 4 (4): 254-265
[8] Li, A. The acoustic representation of the prosodic feature in Chinese dialogue. *Chinese Linguistics*, 2002 (6): 525-535
[9] Cao, J. Prediction of prosodic organization based on the grammatical information. *Journal of Chinese information Processing*, 2003; 17 (3): 41-46
[10] Wang, B., Yang, Y., Lu, S. The acoustic characteristics of large information units' boundaries in monologue discourse. *Acta Acustica*, 2005; 30 (2): 177-183
[11] Hu, W., Zu, Y., Wang, Z. Predict F0 contours prediction on sentence level. *Acta Acustica*, 2006; 31(1): 19-27
[12] Jiang, X., Tian, L., Cui, G. Statistical analysis of prosodic parameters and emotion recognition of multilingual speech. *Acta Acustica*, 2006; 31(3): 217-221
[13] Shao, Y., Han, J., Liu, T., Zhao, Y. Study on automatic prediction of sentential stress with natural style in Chinese. *Acta Acustica*, 2006; 31(3): 203-210
[14] Deng, D., Shi, F., Lu, S. The contrast on tone between Putonghua and Taiwan Mandarin. *Acta Acustica*, 2006; 31(6): 536-541
[15] Huang, X., Yang, Y., Lu, S. Experimental studies on downstep in Chinese intonation. *Acta Acustica*, 2007; 32(1): 56-61
[16] Wang, M., Lin, M. An Analysis of Pitch in Chinese Spontaneous Speech. *Proceedings of TAL 2004 Beijing，International Symposium on Tonal Aspects of Languages, with Emphasis on Tone Languages*, 203-205
[17] Zong, C., Wu, H., Huang, T., Xu, B. An analysis on the corpus of Chinese dialogue of a definite field. *Proceedings of Fifth National Allied Conference on Computational Linguistics* . Beijing: Tsinghua University Press, 1999.
[18] Sinclair, J., *Coulthard, M. Towards an Analysis of Discourse: The English used by teachers and pupils.* Oxford: Oxford University Press, 1975
[19] Center of Putonghua Training and Testing of National Working Committee of Language of China. *An Practical Outline for Chinese Level Test* . Beijing: The Commercial Press, 2004
[20] M. Liberman and J. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language and sound structure*, M. Aronoff and R. T. Oehrle, Eds. Cambridge, MA: MIT Press, 1984, pp. 157–233.