An Analysis of Pitch in Chinese Spontaneous Speech

Maolin Wang

College of Chinese Language and Culture, Jinan University, Guangzhou
wangmaolinp@sohu.com

Maocan Lin

Institute of Linguistics, CASS
linmaocan@263.net

Abstract

In this study, based on spontaneous speech corpus, the F0 values and the F0 ranges of 79 speakers are analyzed. It is found that when F0 value is converted to semitone, there is no difference between the ranges of the two genders. Intonation phrases can be classified according to their pitch range and pitch register.

1.   Introduction

Prosodic study is significant both for phonetics and for speech engineering. To get high quality synthesis and high accuracy recognition, the prosody of spontaneous speech need to be fully investigated. By understanding the prosody of human speech, especially the pitch pattern of spontaneous speech, the naturalness of the synthesized speech can be improved. In this paper we will discuss the pitch pattern in spontaneous speech.

2.   Corpus

The speech materials used in this study are from a telephone dialogue corpus (Zong et al 1999), in which there are mainly recordings of people booking hotels or railway tickets. The syllable annotation, syllable onset and rhyme annotation and sentence type annotation have been made by the Phonetics Laboratory, Institute of Linguistics, CASS. The F0 is first extracted by Praat (www.praat.org), then manually checked through viewing the narrow-band spectrogram.

As the speech material is spontaneous speech, there are a lot of repetitions and hesitations, and there are many fillers (Watanabe & Ishi 2000), such as "en", "this" and "that", which have no actual meaning in dialogue. As this paper is to study the pitch pattern of continuous speech in normal conversation, sentences with repetitions and hesitation are not included in this study. After selection, we used 69 dialogues, in which 79 speakers are involved (37 males and 42 females), whose ages range from 20 to 40.

3.   Results and discussion

3.1. Speakers' maximum and minimum F0

At first, we investigated the maximum and minimum F0 of male and female speakers respectively. The result is shown in Table 1.

Table 1. *Maximum and minimum F0 of male and female speakers*

	Male		Female	
	Maxi F0	Mini F0	Maxi F0	Mini F0
Average F0_Hz_	211.6	85.9	364.2	146.3
Standard deviation	33.5	18.82	39.8	33.4
Number of speakers	37	37	42	42

Table 1 shows that there are great differences between the fundamental frequencies of male and female speakers. Compared to the female speakers, the average maximum F0 of the male speakers is about 152.6 Hz lower, and the minimum F0 is 60.4 Hz lower.

3.2. Speakers' F0 range and semitone range

The speakers' F0 ranges are calculated by the following formula,

$$Rngf0sp = F0\max - F0\min \quad (1)$$

Rngf0sp stands for a speaker's F0 range, and *F*0max and *F*0min stands for his maximum and minimum F0 respectively. The result is as follows:

Table 2. *Speakers' F0 range*

	Male	Female
Average F0_Hz_	125.7	217.9
Standard deviation	34.5	55.4
Number of speakers	37	42

Table 2 shows that the female speakers' average F0 range is 92.2 Hz wider than that of the male speakers'.

Not only is the difference between the F0 values of different genders of speakers quite great, but the difference among the F0 values of different speakers of the same gender is also great. Therefore, when different speakers are involved, the F0 values should be normalized. Here we converted the F0 value to semitone, and the speakers' pitch range is calculated by the following formula,

$$Rngstsp = \frac{12 \times \lg \dfrac{F0\max}{F0\min}}{\lg 2} \quad (2)$$

Rngstsp stands for a speaker's semitone range, and $F0\max$ and $F0\min$ stands for his maximum and minimum F0 respectively. The result is as follows:

Table 3. *Speakers' semitone range*

	All speakers	Male	Female
Average semitone range	16.3	16.4	16.2
Standard deviation	4.41	4.75	5.07
Number of speakers	79	37	42

There is no difference between the semitone ranges of male and female speakers (p < 0.01). The corpus we used here is mainly conversations about people booking hotels and ticket, without highly emotional speech, so their pitch stands for the normal conversational range.

3.3. The pitch range and pitch register of intonation phrase

In recent years there have been a lot of studies about prosody in China, and results show that an utterance is hierachically structured (Lin 2000, 2002; Cao 2001; Li 2002). Based on these studies, we made prosodic labeling for the corpus. Three levels of prosodic units are labeled: prosodic word, intermediate phrase and intonation phrase. For this study we used 1109 intonation phrases.

In section 3.2 it is found that there is no difference between the semitone ranges for male and female speakers. Therefore, when calculating the pitch ranges and pitch register, male and female speakers are not differentiate, the result is as follows:

Table 4. *The pitch range of intonation phrase*

Average pitch range (St)	sd	n	Maximum range (St)	Minimum range (St)
10.2	3.56	1083	26	1.38

Table 4 shows that the average pitch range of intonation phrases is 10.2 semitones, but the differences among intonation phrases are quite great, with the maximum 26 semitones and the minimum 1.38 semitones.

We take the top point of the intonation phrase as its pitch register. The reason is that there are four tones in Standard Chinese, HH, MH, LL and HL, with only the LL tone lack H feature. However, when a LL tone precedes another LL tone, it will change to MH by tone sandhi rule. Therefore, generally speaking, there will always be H feature in an intonation phrase, so we can take top point as the pitch register.

The pitch ranges and pitch registers of intonation phrases are normalize by the speaker's pitch range by using the following formula,

$$Rngnom = \frac{Rngip}{Rngstsp} \quad (3)$$

$$Rgnom = \frac{Sgip}{Rngstsp} \quad (4)$$

Rngnom and *Rgnom* stand for the normalized pitch range and pitch register of the intonation phrase respectively, *Rngip* and *Sgip* for its range and register in semitone respectively, and *Rngstsp* for the pitch range of the speaker in semitone. The result is as follows:

Table 5. *Normalized pitch range and pitch register of the intonation phrase*

	Average (normalized)	Standard deviation	Number
Pitch range	0.59	0.19	1083
Pitch register	0.8	0.13	1108

The pitch range and pitch register of the intonation phrase are correlated with each other (r = 0.62).

Intonation phrases (IP) can be classified according to its pitch range and pitch register (Shen 1998). According to the pitch range, they can be classified to three types, Narrow, Mean, and Wide, and according to the pitch register, Narrow IP can be further classified to High, Mid and Low, and Mean IP can be further classified to High and Low. Altogether there are six types intonation phrases, namely, High narrow, Mid narrow, Low narrow, High mean, Low mean and wide.

The pitch representations of intonation phrases are affected by three factors, the length of the intonation phrase, the interactive function of the utterance and its position in the discourse. We find that when an intonation phrase is long, its pitch register tends to be high, and its pitch range also tends to be wide. On the contrary, when an intonation phrase is short, its pitch register tends to be low, and its pitch range also tends to be narrow.

Another factor for the pitch representations of the intonation phrases is the interactive function of the utterance. The corpus we used here is material of telephone conversations, in which are mainly short utterances, and most of the utterances includes only one intonation phrase. An utterance can functionally be either informative or interrogative, and we find that there are degrees for informativeness and interrogativeness. When an utterance conveys much information to the listener, it is of high informative degree. On the other hand, when an utterance conveys little information to the listener, it is of low informative degree. The same is true for interrogative degree.

Still anther factor for the pitch representations of the intonation phrases is the position of the utterance in the discourse. When an utterance is at the beginning of a new topic, its pitch register tends to be high, and its range also tends to be wide. On the other hand, when an utterance is at the end of a topic, its pitch register tends to be low, and its range also tends to be narrow. This is in line with the findings in English (Hirschberg & Pierrehumbert 1986; Fisher & Tokura 1996).

4. Conclusion

F0 values and F0 ranges of male and female speakers are quite different, but there is no difference between the semitone ranges of the two genders. Intonation phrases can be classified according to their pitch ranges and pitch registers. In speech engineering, the overall pitch range and pitch register can be constructed by analyzing the length of the intonation phrase, the interactive function of the utterance and its position in the discourse.

5. References

[1] Cao, Jianfen, 2001. The phonetic and linguistic clue for Chinese prosodic segmentation, *Modern phonetics in the new century,* Beijing: Tsinghua University Press.

[2] Fisher, Cynthia and Tokura, Hisayo, 1996. Prosody in speech to infants: Direct and indirect acoustic aues to syntactic structure. In Morgan, James L., Katherine Demuth, ed, *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

[3] Hirschberg, Julia and Pierrehumbert, Janet, 1986. The intonational structuring of discourse. *Proceeding of the 24th annual meeting of the ACL.*

[4] Li, Aijun, 2002. The acoustic representation of the prosodic cues in Putonghua dialogue, *ZhongguoYuwen,* 2002, 6.

[5] Lin, Maocan, 2000. Breaks and sentence prosodic structure in Putonghua sentences, *Contemporary Linguistics*, 2000, 4.

[6] Lin, Maocan, 2002. Prosodic structure and the construction of F0 top and bottom lines of Putonghua sentences, *Contemporary Linguistics*, 2002, 4.

[7] Shen, Jiong, 1998. The classification of Chinese intonation and its annotation, *Yuyan Wenzi Yingyong,* 1998, 1.

[8] Watanabe, Michiko and Ishi, Carlos Toshinori, 2000. The Distribution of Fillers in Lectures in the Japanese Language. Proceedings in ICSLP 2000, Beijing.

[9] Zong, Chengqing, Wu, Hua, Huang, Taiyi and Xu, Bo, 1999. An analysis of Chinese dialogue corpus for a specific field. *Proceeding of the Calculational Linguistics.* Beijing: Tsinghua University Press.