The Control of Tones on Pitch Movement in Mandarin

Ziyu Xiong

Institute of linguistics
Chinese Academy of Social Science
xiongziyu@163.com

Abstract

This paper studied the control of tones on pitch movement in syllables. Five types of corpora are used here: corpus of monosyllables, corpus of disyllables, corpus of read sentences, corpus of read paragraphs, and corpus of spontaneous speech. The results show that the ability of tones to control the pitch movement is descending in the above listed sequence. With the more intervention of other factors in connected speech, tones play a less important role in controlling pitch movement. It proves that pitch movement in utterance is controlled by various factors and the type of tones is just one of them.

1. Introduction

Chinese is a tone language, which uses pitch to produce distinctive word-meanings. Syllables with different tones have different pitch representation. Therefore, the type of tones inevitably plays a very important role in the pitch. There are four types of tones in mandarin: YINPING, YANGPING, SHANGSHENG and QUSHENG. In the citation form, they can be represented as [55], [35], [214], [51] respectively according to Five-Letter Scale [1].

But the pitch contour and pitch value may be changed under the influence of other factors in connected speech [2, 3, 4, 5]. Many studies have show that the pitch of a syllable is controlled not only by the type of tones, but also by other factors in the utterance [6, 7, 8, 9]. So it can be deduced that the type of tones might has a less power to control the pitch of a syllable with the more intervention of the other factors.

This paper will examine the control of the type of tones on pitch value based on different corpus. The corpora under study include corpus of monosyllables, corpus of disyllables, corpus of read sentences, corpus of read paragraphs, and corpus of spontaneous speech. The difference among those corpora mainly is the naturalness, which is increasing in the above sequence.

2. Description on the Corpora

All the corpora used here were selected from the databases made by Phonetic Lab in Institute of Linguistic of CASS (Chinese Academy of Social Science). The following are the concise explanation for those corpora.

2.1 Corpus of monosyllables

The design of this corpus aimed to serve for the phonetic synthesis. It took into account the phonetic balance for monosyllables by considering the initials, the finals, and the tones. There are altogether 1275 monosyllables spoken by 15 males between the ages of 20 to 45, who speak Standard Chinese well. This study used the sound materials of three speakers, noting as M01, M02 and M03.

2..2 Corpus of disyllables

This corpus also designed for the phonetic synthesis. It considered the juncture of intra-syllables and inter-syllables and the combination of tones. This corpus covered 3471 disyllables, which were read by a male in mid-age who speaks standard Chinese well.

2.3 Corpus of read sentences

It is a sub-corpus of NOKIA-CASS-Ⅰ which includes 1000 independent sentences of 15 to 25 syllables. Ten speakers between the ages of 15 to 25, including 5 males and 5 females, read 100 sentences everyone. They also speak Standard Chinese well.

2.4 Corpus of read paragraphs

This corpus consists of 18 short paragraphs of average 450 syllables. The materials were selected from newspapers, and the contents were fit for reading. Ten speakers between the ages of 20 to 45 read these materials. The speakers are 5 males and 5 females who speak Standard Chinese well. This study used the sound materials of two speakers, including one male and one female, noting as M001 and F002.

2.5 Corpus of spontaneous speech

This corpus covered a speech database of telephone-dialogue in which there is two-and-a -half-hour dialogue. The dialogue took place between clients and operators in hotels. There are altogether 94 independent dialogues, but some are excluded because of the bad quality of recording and the accent in the speech.

3. Annotation and Data Distraction

Five skillful workers in Phonetic Lab annotated the segmental and prosodic information based on SAMPA-C and C-ToBi (Li, A.J. 2002). The annotation for segment covered the segmentation of syllables, the segmentation of initials and finals, the tone type etc. The annotation for prosody includes labeling for prosodic boundary and stress, etc. In this study, we distracted ten pitch values in the pitch contour of each syllable equidistantly according to the information provided by the boundary of the syllable and its initials and finals.

The corpora analyzed were selected from five data banks and the materials were spoken by different speakers. The pitch values of different speakers were normalized to remove the individual diversity as much as possible. Two steps were used here to deal with the pitch value:

Step 1: The pitch values of every speaker were categorized based on the four tones and the mean pitch values of each

point of each tone were calculated. Then the maximum of mean pitch value (F_0max) was regarded as the upper limit of pitch range for this speaker and the minimum (F_0min) as the lower limit. Generally speaking, F_0max is the maximum pitch value in the falling tone and F_0min is the minimum pitch value in the low tone.

Step 2: The pitch values of a speaker were normalized according to F_0max and F_0min. The formula, X=5*(F_0i-F_0min)/(F_0max–F_0min), were used here to normalize the pitch values into the range of 0 to 5. F_0i in this formula refers to the actual value under investigation. X may be positive or negative after normalization. If X is larger than 5, F_0i is above the upper limit of pitch range for this speaker, and if X is smaller than 0, F_0i is below the lower limit of pitch range for this speaker.

To test the validity of the normalization, the data of monosyllables read by M01, M02 and M03 in database of monosyllables were selected to be normalized according to the above method. The data without normalization were first investigated according to the different tones. Figure 1 is the diagram for mean frequencies of different tones.

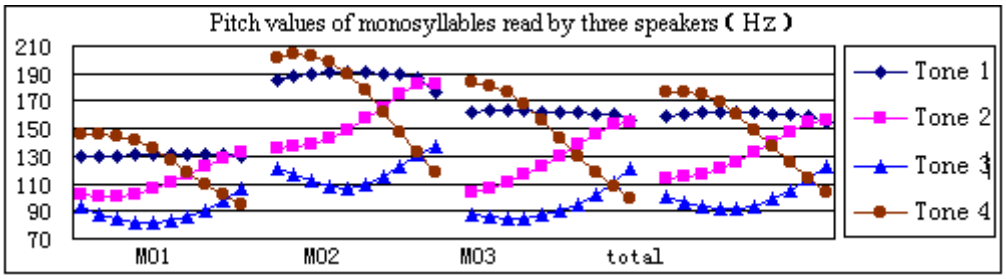

Figure 1: *The diagram for mean frequencies of different tones read by three speakers*

Figure 1 tells us that there is significant difference in the pitch range and the pitch register among different speakers: M01 has a lower pitch register and a narrower pitch range, M02 has a higher pitch register and a wider pitch range, and M03 has a lower pitch register and a wider pitch range. We then got the F_0max and F_0min first and normalized the pitch values into the range of 0 to 5 according to the formula X=5*(F_0i- F_0min)/(F_0max–F_0min). Figure 2 was made by the normalized values.

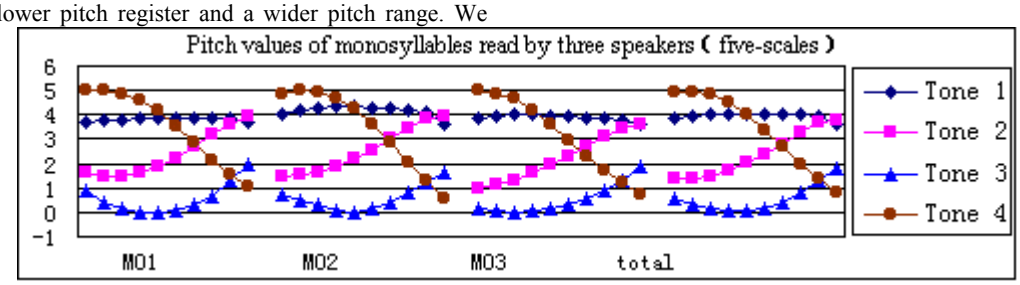

Figure 1: *The diagram for mean five-scales of different tones read by three speakers*

Figure 2 represents that the pitch values are consistent for the three speakers after normalization. The difference in pitch register and pitch range among different speakers can almost be eliminated. It proves that the normalization is efficient and can be used to remove the individual diversity as much as possible in pitch among speakers. We can study the control of tones on pitch movement by analyzing the normalized pitch values.

4. Data Analysis

In the citation form, there are four tones in Mandarin, i.e. YINPING, YANGPING, SHANGSHENG and QUSHENG. Therefore, there is no neutral tone in the database of monosyllables, but much more in the database of continuous speech. In order to make a contrastive study, we didn't consider the pitch movement of the neutral tones. Table 1 shows the number of syllables appeared in the five databases based on the type of tones.

Table 1: *The number of syllables with different tones in the five databases.*

Type of database	Monosyllables	Disyllables	Read sentence	Read paragraph	Spontaneous speech
Tone 1	1008	1654	3942	3328	2672
Tone 2	783	1770	4766	3810	2910
Tone 3	972	1158	3220	2846	3174
Tone 4	1062	2353	6612	5722	5594
Total	3825	6935	18540	15706	14350

The control of tones on pitch movement can be weighed using the following statistical scales:

(1) The difference of pitch among different tones, which is represented by F value in statistics. The bigger the F value is, the more powerful the control of tones on pitch is.

(2) Eta Squared value in statistics, which refers to the ratio of the variance on pitch value caused by tones to the total variance. The bigger the Eta Squared value is, the more powerful the control of tones on pitch is.

(3) The success rate, which is the rate of judging the tones by the pitch values of syllables. The higher the success rate is, the more powerful the control of tones on pitch.

The three scales are included in Table 2 to analyze the control of tones on pitch movement. All the statistical work was done by using SPSS 10.0. We took the type of tones as the independent variables and the ten pitch values as dependent variables to do the One-Way ANOVA analysis. The F values and Eta Squared values are listed in Table 2.

Table 2: *Result of F values and Eta Squared values.*

Type of Database	**Monosyllables**		**Disyllables**		**Read sentence**		**Read paragraph**		**Spontaneous speech**	
	F	*Eta-S*	*F*	*Eta-S*	*F*	*Eta-S*	*F*	*Eta-S*	*F*	*Eta-S*
X1	5552	81.4%	2036	46.8%	1861	23.1%	823	24.2%	522	11.2%
X2	9091	87.7%	3694	61.5%	2708	30.5%	967	27.2%	592	12.5%
X3	11071	89.7%	5844	71.7%	3526	36.3%	1035	28.6%	645	13.5%
X4	10531	89.2%	7752	77.0%	3706	37.5%	1012	28.2%	678	14.1%
X5	8505	87.0%	6865	74.8%	3486	36.1%	937	26.6%	687	14.3%
X6	6134	82.8%	4731	67.2%	3221	34.3%	880	25.4%	678	14.1%
X7	4518	78.0%	4050	63.7%	3131	33.6%	899	25.8%	658	13.8%
X8	4130	76.4%	5595	70.8%	3248	34.5%	1005	28.0%	637	13.4%
X9	4281	77.1%	8962	79.5%	3220	34.3%	1134	30.5%	621	13.1%
X10	2638	67.5%	7958	77.5%	2896	31.9%	1212	32.0%	605	12.8%
Total	6645	81.7%	5749	69.1%	3100	33.2%	990	27.7%	632	13.3%

Table 2 shows that the more natural the speech is, the less powerful role the type of tones plays in controlling pitch movement, and the weaker the interpretability of the total variance of pitch movement is. In the database of monosyllables, the interpretability of the total variance of tones on pitch movement is about 82%, which shows the type of tones is the major factor to control pitch movement in citation form. However, the interpretability in spontaneous speech is just about 13%, which tells that the type of tones is not the major factor any longer. It also shows that more factors come to affect pitch movement with the increase of the naturalness.

Generally speaking, the type of tones can be judged according to the representation of pitch in a syllable because syllables with different tones have different pitch contours in theory. The more significant the difference in pitch among different tones is, the higher the success rate of judging tones on pitch is. Accordingly, the control of tones on pitch movement can be judged by the success rate. That is to say, the higher the success rate is, the more powerful the control of tones on pitch movement is. Table 3 shows the judgment of tones on pitch by using SPSS 10.0.

Table 3: *The success rate of the judgment of tone.*

Type of database	**Monosyllables**	**Disyllables**	**Read sentence**	**Read paragraph**	**Spontaneous speech**
Tone 1	99.6%	96.6%	77.4%	69.1%	61.8%
Tone 2	98.3%	91.1%	68.0%	58.7%	47.5%
Tone 3	98.1%	96.6%	77.2%	70.3%	54.9%
Tone 4	93.6%	98.5%	61.0%	59.0%	45.5%
Total	97.3%	95.8%	69.1%	63.1%	51.0%

Table 3 shows that the more natural the speech is, the lower the success rate is. The success rate in database of monosyllables is about 97%, and that in database of spontaneous speech is about 51% because about a half

syllables need to be judged on syntax, semantics, and prosody apart from their pitch feature. It further proves that the higher the naturalness of a speech is, the less powerful the control of tones on pitch movement is.

5. Conclusion

This paper studied the control of tones on pitch movement in Mandarin by using the database of monosyllables, of disyllables, of read sentences, of read paragraphs and of spontaneous speech. The result shows that the control of tones on the pitch movement is descending in the above listed sequence. In conclusion, the type of tones plays a less important role in controlling pitch movement with the increase of the naturalness of speech. That is to say, the type of tones is just one factor of controlling pitch movement because there may be other factors as syntax, semantics, and prosody, etc.

6. References

[1] Chao, Y.R., 1968. *A Grammar of Spoken Chinese*. University of California Press.

[2] Cao, Jianfen, 2002. The relationship between tone and intonation. *Zhongguo Yuwen* (Chinese linguistics), 3.

[3] Lin, M., Lin, L., Xia, G. & Cao, Y., 1980. Putonghua erzici biandiao de shiyan yanjiu [An experimental study of tonal variation in disyllabic words in Standard Chinese]. *Zhongguo Yuwen* (Chinese linguistics), 74-79.

[4] Lin, M. & Yan, J., 1991. Tonal coarticulation patterns in quadrisyllabic words and phrases of Mandarin. *Proceedings of The 12th International Congress of Phonetic Sciences*, Aix-en-Provence, France, 242-245.

[5] Wu, Z.J., 1982. Putonghua yuju zhong de shengdiao bianhua [Tonal variations in Mandarin sentences]. *Zhongguo Yuwen* (Chinese linguistics), 439-450.

[6] Xu, Y., 1996. Factors affecting the surface tonal contours of Mandarin. Proceedings of The 3rd National Conference on Chinese Phonetics, Beijing, 35-36.

[7] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61-83.

[8] Xu, Y., 1999. Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics,* 27, 55-105.

[9] Xu, Y., 2001. Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, monograph series, No.17, 1-31.

[10] Li, Aijun, 2002. Chinese prosody and prosodic labeling of spontaneous speech. In Bel B. and Marlien I. (eds.), *Proceedings of the Speech Prosody 2002 Conference.* Aix-en-Provence, France, 39-44.