A Global Perspective on Chinese Intonation

Wu Lihan[1], *Ma Qiuwu* [2]

[1]Department of Foreign Languages, Nankai University
[2]Department of Foreign Languages, Tongji University
wulihan8228@126.com, tjnkmqw@126.com

Abstract

An analysis procedure of Chinese intonation is tentatively proposed from a global perspective. This procedure starts from an observation of the interaction among three prosodic parameters and the hierarchical systems of production by looking at three continuums of utterances, followed by developing a system of acoustic and perceptual based features which can be further applied to the analysis of various types of intonation. This procedure is demonstrated by investigating into the five-syllable utterances produced by a male speaker. This pilot study suggests that the hierarchies of both duration and intensity might go up in a non-linear way in a speaker's mind, while the degree of pitch measured in semitone increases in a linear way. The selection and combination of global features indicate the restriction coming from both extrinsic (the communicative principle of explicitness) and intrinsic factors (the constraint of physical realization).

Index Terms: Chinese intonation, global feature, continuum, prosodic parameters

1. Introduction

Studies on Chinese intonation since the 1960s have been mainly focus on the relation between citation tones and sentential intonation (Chao 1968), the acoustic realization of boundary tones (Lin 2004), focuses (Chen 2006), intonational phrases and prosodic words (Li & Wang, 2003), all of which are conducted at the local level. Analyses which targets at the global effect are confined to either an impressional description that lacks data support, or a comparison between functional intonations in which the sample utterances are insufficient to reveal the fact that intonation can actually be produced at any degree within a wide range of register, tempo and intensity.

Since the previous elaborate acoustic descriptions on the specific positions of an utterance may have neglected the overall features of an utterance as a whole, we here argue that the sentential intonation can also be dealt with from a macro point of view, i.e. using simply one feature in each prosodic dimension to describe the whole utterance, and propose an analysis procedure which may offer help in the comparative study among all types of intonation, especially those concerning emotional speech.

As one of the foremost goals in scientific research is to figure out the relation between different parameters that constitute the phenomenon, we firstly delve into the interaction among the three prosodic elements (i.e. duration, pitch, intensity) in Chinese intonation. To see how the other two elements alter along with certain element, one of the possible ways is to increase the value of a parameter without controlling the alternation of the other two, and this shall be done in a speaker's natural speech instead of using speech synthesis. Hence, three continuums of utterances can be produced, each of which focuses on the increase of one parameter, leaving the rest two changing simultaneously (See 2.1. for the detail.)

On the basis of the three continuums, a further observation can be made to see whether a speaker follows any pattern when increasing the value of each parameter. This may shed light on our understanding of the mechanism of intonation production.

These continuums can also be used as materials for a perceptional experiment, which aims to find out the acoustic boundaries of perceptional categories (high, mid and low) of each parameter, based on which the acoustic and perceptional based global features might be set up and applied to various types of intonation. In this way, three aspects of intonology (productory, acoustic and perceptual) are well connected, and so are phonetics and phonology.

The following sections illustrate how this method may be conducted by using a 24-year-old male Mandarin speaker's utterance samples as a case study, which foreshadows some findings that needs more data support in future studies.

2. Methods

2.1. Materials and procedure

The speaker is to produce four declarative sentences without laying stress (focus) on any particular syllable. Each sentence contains five syllables, all of which are of the same tone so as to reduce minimally the interaction between different tones.

Tone 1 (high) sentence: *Jintian chi xican.*
'Eat western food today.'
Tone 2 (rise) sentence: *Mingnian hui Nanning.*
'Go back to Nanning next year.'
Tone 3 (low) sentence: *Wudian mai yusan.*
'Buy an umbrella at five o'clock.'
Tone 4 (fall) sentence: *Xiaji kan Aoyun.*
'Watch the Olympic Games in summer.'

For each sentence, three sets of utterances are produced to form respectively the continuums of duration, pitch and intensity. Specifically, the duration continuum is produced by uttering each sentence from an extremely fast speed to an extremely slow one, with one utterance at a certain speed. The speaker is free to produce any number of utterances as long as he feels that an utterance is slower than the previous one. Similarly, to produce the pitch continuum, the speaker utters firstly in his lowest register, then at a slightly higher one which he considers as being different from the previous one. This goes on gradually until his highest register, with one utterance occurs at a certain pitch level. Again, no limitation is made on the number of levels within his register. Lastly, the intensity continuum is produced by speaking from an extremely soft voice to an extremely loud one.

Since this study intends to explore both the psychological and the corresponding physical aspects of intonation production, the psychological term of "pitch" and the physical "duration" and "intensity" are used at the same time, so as the later psychoacoustic unit of "semitone".

2.2. Recordings

Recordings are made with a ThinkPad T61 laptop computer and a Sony microphone (ECM-MS907). Audio files are saved with 16 bits in a mono-channel PCM digital form. Praat, SPSS and Excel are then used in data analysis.

2.3. Calculations

2.3.1. Duration

For each sample utterance, calculate the average syllable duration (ASD) of the five syllables by adopting this formula: ASD(ms)=overall length of an utterance with silences subtracted /5

2.3.2. Pitch and intensity

1) For each syllable, obtain the F0 (Hz) and intensity (dB) of 10 dots which distribute evenly along its pitch contour.
2) Calculate the average F0 and intensity for each syllable.
3) Calculate the average F0 and intensity of the five syllables.
4) Transfer the Hz value into Semitone (St) by the following formula, adopting 64Hz as the reference frequency:

St=12*lg(F0 frequency Hz/ref frequency) /lg2 (Shi *et al*, 2009)

3. Results and Discussion

3.1. Correlation among three prosodic parameters

Wu (2004:13) proposes the compensatory relations between the three prosodic features as follows: 1) High/Low of tone→ ←Long/Short of duration (one way or two ways).2) High/Low of tone → Strong/Weak of intensity (one way).3) Long/Short of duration→←Strong/Weak of intensity (one way or two ways). However, Wu does not present his methods or data in his paper. Thus, a further correlation analysis is conducted here, since the speaker focuses on the increase of one single parameter when producing one continuum, making sure that the other two parameters are only a simultaneous alternation.

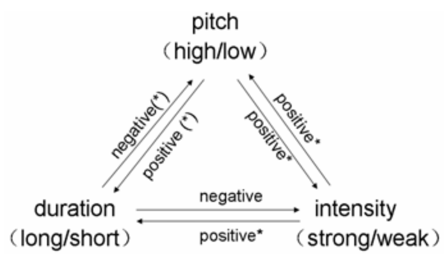

Figure 1: *Correlation triangle of prosodic parameters.*

The results are summarized in Fig. 4, where the asterisk "*" stands for being significant, and the brackets "( )"suggests that being significant is possible. Being different from Wu's claim that pitch and intensity are in one way positively correlated, the two are in significantly positive correlation in two ways, i.e. in the pitch continuum, when pitch elevates, the intensity increases (Sig.=0.000), while in the intensity continuum, when intensity increases, so does the pitch (Sig.=0.000). For duration and intensity, when a syllable is lengthened in the duration continuum, its gets weaker---a demonstration of negative correlated; when it gets stronger in the intensity continuum, its duration is significantly lengthened (Sig.=0.001), showing a one-way negative or positive correlation, being different from Wu's claim that when duration increases, so does the intensity. Duration and pitch also exhibits the similar one-way correlation---when pitch gets higher in the pitch continuum, it is also lengthened significantly in the area of low duration (Sig.=0.009), but insignificantly in the area of high duration; when it is lengthened in the duration continuum, the pitch goes down significantly in the higher register (Sig.=0.032) and insignificantly in the lower register, being different from Wu's claim that when duration increases, so does the pitch.

3.2. Hierarchical systems of prosodic continuums

By looking into M1's three continuums, we are to reveal the underlying hierarchical system of each prosodic parameter in his mind. In the duration continuum, M1 produces 9,7,8,7 utterances respectively for the 4 sentences. Figure 1 shows how the average pitch duration (y-axis) of the utterance increases as the degree of duration (x-axis) gets higher (from the fastest to the slowest) for Tone 1-4 sentences (4 curves).

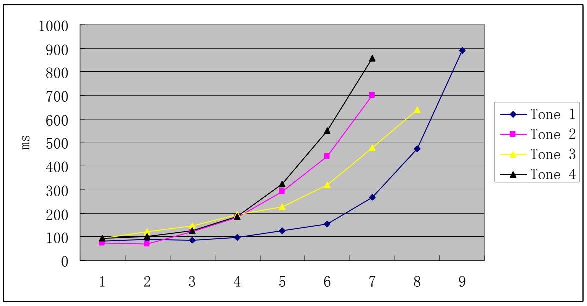

Figure2: *Average syllable duration of the duration continuum.*

Formulas of the trend lines and goodness of fit (R^2):

Tone 1: $y = 4.8372x^3 - 49.162x^2 + 151.81x - 39.346$ ($R^2 = 0.9945$)

Tone 2: $y = 2.4278x^3 - 6.5238x^2 + 18.206x + 55.114$ ($R^2 = 0.9982$)

Tone 3: $y = 1.8657x^3 - 12.015x^2 + 51.11x + 51.3$ ($R^2 = 0.9973$)

Tone 4: $y = 3.1889x^3 - 6.6333x^2 - 1.1651x + 99.686$ ($R^2 = 0.9997$)

In the pitch continuum, M1 produces 16, 12, 12, and 11 utterances respectively for the 4 sentences. Figure 2 shows how the average pitch height (y-axis) of the utterance increases as the degree of pitch (x-axis) gets higher (from the lowest to the highest) for Tone 1-4 sentences (4 lines).

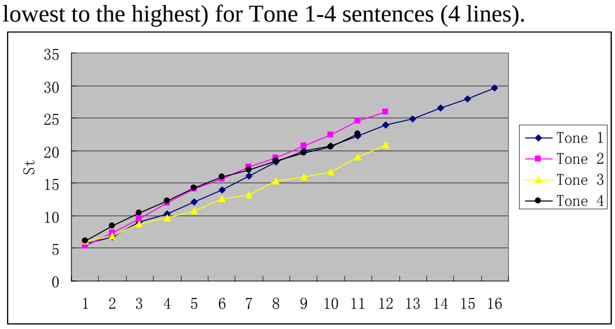

Figure 3: *Average pitch height of the pitch continuum.*

Formulas of the trend lines and goodness of fit:

Tone 1: $y = 1.6151x + 4.2525$ ($R^2 = 0.9947$)

Tone 2: $y = 1.8594x + 4.0803$ ($R^2 = 0.9961$)

Tone 3: $y = 1.3042x + 4.5394$ ($R^2 = 0.9906$)
Tone 4: $y = 1.5727x + 5.6455$ ($R^2 = 0.9872$)

In the intensity continuum, M1 produces 4 utterances for each sentence. Fig. 3 shows how the average intensity (y-axis) increases as the intensity (x-axis) gets louder (from the weakest to the strongest) for Tone 1-4 sentences (4 curves).

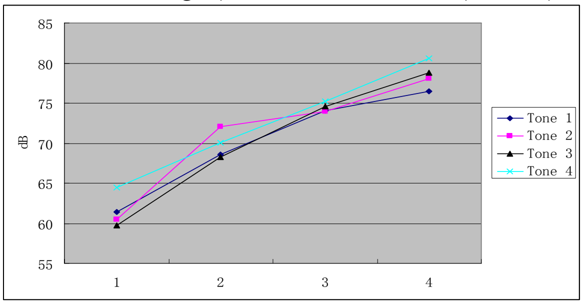

Figure 4: *Average syllable intensity of the intensity continuum.*

Formulas of the trend lines and goodness of fit:
Tone 1: $y = -0.2333x^3 + 0.55x^2 + 7.1833x + 53.9$ ($R^2 = 1$)
Tone 2: $y = 2.0333x^3 - 17.1x^2 + 48.667x + 26.9$ ($R^2 = 1$)
Tone 3: $y = 0.0333x^3 - 1.35x^2 + 12.417x + 48.6$ ($R^2 = 1$)
Tone 4: $y = 0.1333x^3 - 1.05x^2 + 7.8167x + 57.6$ ($R^2 = 1$)

1) Utterances in the pitch continuum increases at a distance of 1.61 semitones on average (STDEV=0.528) in a linear way (Fig.2), whereas the degree of both duration and intensity increases in a non-linear way (Fig. 1, 3). This may suggest that the hierarchy of different prosodic categories in the speaker's mind follow different cognitive modes.

2) Rather than supporting the common argument that sentential intonation has an impact on the syllable tone, the citation tone here does demonstrate an effect on the sentential intonation, which is seen in both the different number of grades and the values in each continuum. For example, Tone 1(high level tone) utterances have the most grades and the vastest range of values in both duration and pitch continuums, Tone 3 (low tone) has the lowest pitch in the pitch continuum, and Tone 2 and 3 which bear the low feature have a shorter duration compared with Tone 1 and 4.

3.3. Setting up the global features

Table 1. *Acoustic boundaries of prosodic features in three continuums.*

Utterance types	Fea-tures	Duration (ms)	Pitch (St)	Intensity (dB)
Tone 1	L	82.8-96.6	5.7-10.3	61.4-68.6
	M	96.6-152.2	10.3-20.7	68.6-74.1
	H	152.2-890.4	20.7-29.7	74.1-80.9
Tone 2	L	70.6-123.4	4.5-9.5	59.8-72.1
	M	123.4-293.4	9.5-14.1	72.1-73.9
	H	293.4-701.2	14.1-25.9	73.9-78.1
Tone 3	L	91.8-145.4	6.2-8.8	59.7-68.3
	M	145.4-225.8	8.8-12.6	68.3-74.6
	H	225.8-640.6	12.6-20.9	74.6-78.8
Tone 4	L	93.2-126.2	6.2-10.4	64.5-70.1
	M	126.2-185.4	10.4-14.3	70.1-75.2
	H	185.4-857.4	14.3-21.9	75.2-81.6

Three listeners (1 male and 2 female) participate in a perception experiment. They are told to take down their impression by marking "high, middle or low" when they listen to each of these utterances, which do not follow the original sequence of production. By adopting marks agreed by at least two listeners, we have three acoustic and perceptual based features [high] (H), [mid] (M) and [low] (L) for each prosodic parameter. Notice here that rather than reflecting directly the perceptional impression, [H], [M] and [L] refers strictly to the values of the parameter, hence, a [H] of duration indicates a high value of duration, i.e. an extremely slow speed.

3.4. Feature values of intonation

Table 2. *Feature assignment for 26 types of intonation.*

No.	Types of intonation	Duration	Pitch	Intensity
	Functional			
1	Declarative	M	M L	L (M)
2	Interrogative	M	H	H (M)
3	Imperative	M (L)	M H	H
4	Exclamatory	H M	M(H L)	M H
	Functional (whispered)			
5	Declarative	H (M)		L
6	Interrogative	H (M)		L
7	Imperative	H (M)		L
8	Exclamatory	H M		L
	People of different age			
9	Children	M (H)	H	M H
10	Young people	M	H (M)	H (M)
11	Middle-aged	H (M)	L (M)	M
12	Seniors	H	L (M)	L
13	Women	H (M)	H	M
	Social roles/other situations			
14	Couples	M (H)	L M	L
15	Family	M	M L	M
16	Intimate friends	L (M H)	M (H)	H (M)
17	Waiters	M	L	M
18	Bosses	M H	M (L)	H (M)
19	Chinese teaching	H	M	M (H)
20	Calling at a distance	H	H	H
	Emotions			
21	Impatient	L (M)	M (H)	M
22	Excited	M	H	H
23	Angry	M (H)	H	H (M)
24	Hesitant	H	M	L (M)
25	Disappointed	H	M (H)	M (H)
26	Incredulous	H	M	M (L)

The speaker is told to produce the Tone 1-4 sentences with 26 types of intonation, after which the calculation in Section

2.3. is made. Based on the acoustic boundaries for the prosodic features in Table 1, we can now assign values for these types of intonation. In Table 2, the first feature [M] in the column of “Duration” means that all the Tone1-4 declarative sentences are in a middle range of tempo. Features in the brackets suggest an alternative selection.

3.5. Feature selection and combination

The three features ([L], [M] and [H]) of the three prosodic parameters can be combined into 27 (3*3*3) types of combinations, but not all these occur in actual speech. Table 3 list all the possible feature selection of 22 types of intonation based on the data in Table 2 (the four whispered functional intonations are not taken into account). In Table 3, the capital letters “D, P, I” stands for duration, pitch and intensity respectively, and the number in the columns of “Tone1-4” refers to the intonation type provided in Table 2.

Table 3. *Feature selection of 22 types of intonation.*

D	P	I	Tone 1	Tone 2	Tone 3	Tone 4	Total
L	L	L					0
L	L	M					0
L	L	H					0
L	M	L					0
L	M	M	21	21		21	3
L	M	H					0
L	H	L					0
L	H	M			21		1
L	H	H			3		1
M	L	L		1	14	1	3
M	L	M	18,17	15,11	15		5
M	L	H					0
M	M	L		14	1	14	3
M	M	M	1,15, 10	3	4,16	15	7
M	M	H	3,16	18,10	10	16	6
M	H	L					0
M	H	M	9	13	2,9,23		5
M	H	H	2,22,23	2,4,22, 16,9,23	22	2,3,22, 10	14
H	L	L	12,14	12	12		4
H	L	M			11	4,11	3
H	L	H					0
H	M	L		24	24	12， 24,26	5
H	M	M	25,24, 11,26	19	19	19	7
H	M	H	4， 16,19	25	18	18,25	7
H	H	L					0
H	H	M	13,20		13		3
H	H	H		20	25,20	9,20, 23	6

Seen in Table 3, the combination with the highest frequency is [M H H] (14 times, 16.9%), which doubles the second ranked [M M M], [H M M] and [H M H]. We argue that these most selected combinations follow the principle of communicative explicitness and are easy to produce.

Meanwhile, 10 combinations out of 27 do not occur in any of the intonational type: [L L L], [L L M], [L L H], [L M L], [L M H], [L H L], [M L H], [M H L], [H L H], [H H L]. These combinations fall mainly into two categories. The first group (the first three combinations) includes the combination of [L] duration and [L] pitch, which indicates that there is barely no daily communication with both fast tempo and low pitch due to the communicative principle of being explicit and clear. The second group involves a combination of opposite values (H and L), especially between pitch and intensity, shown in the last five combinations. In other words, opposite values are forbidden for pitch and intensity in the same utterance. This group is related to the difficulty or impossibility of physical realization. Hence, it reveals a contradiction between the symmetry in phonological analysis and the unbalance in actual phonetic implementation.

4. Conclusions

When looking into Chinese intonation from a global viewpoint, we can deal with quite a few issues that the traditional analysis may fail to cope with. By analyzing the continuums of duration, pitch and intensity, the correlation triangle of prosodic parameters has been worked out. It is then suggested that the hierarchy of duration and intensity goes up in a non-linear way in a speaker’s mind, while the pitch increases linearly. A system of acoustic and perceptual based features [H], [M] and [L] are set up. The restriction of the selection and combination of these features comes from both extrinsic and intrinsic factor. A large amount of data is needed to support these results in the future.

5. Acknowledgements

This work is supported by the Shanghai social science project of “A Phonological Study on Chinese Intonation” hosted by Professor Ma Qiuwu. Special thanks also goes to Doctor Li Yunjing for providing the Praat scripts.

6. References

[1] Chao, Yuanren. *A Grammar of Spoken Chinese.* Berkeley: University of California Press. 1968.

[2] Chen,Yiya. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics*. 34 (2006) 176–201.

[3] Li, Aijun，Wang Xia. A contrastive investigation of standard Mandarin and accented Mandarin. *Report of Phonetic Research*, 2003.

[4] Lin, Maocan. Boundary tone of Chinese intonation and its pitch (F0) pattern. In Fant, G, H. et al (eds). *From Traditional Phonology to Modern Speech Processin*g: 309-328. Beijing: Foreign Language Teaching and Research Press, 2004.

[5] Shi, Feng, Wang Ping and Liang Lei. The Undulating Scale of the Intonation of Declarative Sentences in Standard Chinese. *Nankai Linguistics*. 2009 (2).

[6] Wu, Zongji. From traditional Chinese phonology to modern speech processing—realization of tone and intonation in Standard Chinese. In Fant, G, H. et al. (eds) *From Traditional Phonology to Modern Speech Processing*: 1-16. Beijing: Foreign Language Teaching and Research Press, 2004.