# EMOTIONAL INTONATION IN A TONE LANGUAGE: EXPERIMENTAL EVIDENCE FROM CHINESE

*Aijun Li[a], Qiang Fang[a] & Jianwu Dang[b,c]*

[a]Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China;
[b]Tianjin University, Tianjian, China; [c]Japan Advanced Institute of Science and Technology, Japan
liaj@cass.org.cn; fangqiang@cass.org.cn; jdang@jaist.ac.jp

## ABSTRACT

Chinese is a tonal language. How lexical tones and intonation interplay with each other is an interesting question. In this study, we investigated emotional intonations by analyzing monosyllabic utterances from two speakers. We found that the tonal space, the edge tone and the duration differ greatly across 7 emotions, and that different speakers showed consistent production patterns. The speakers expressed 'Disgust' or 'Angry' by using a kind of 'Falling' successive addition tone, and 'Happy' or 'Surprise' by a kind of 'Rising' successive addition tone, as pointed out by Chao [1]. The function of the successive addition boundary tone is to express the speaker's emotion rather than linguistic information. We show that it is more reasonable to use both traditional boundary tone features (H% or L%) and the successive addition tone features (r, f or le) to describe boundary tones of emotional intonation. For instance, 'H-f%' stands for a high boundary tone but with a falling successive addition tone.

**Keywords:** emotion, Chinese, intonation, successive addition tone, boundary tone

## 1. INTRODUCTION

The intonation of tonal languages like Chinese is an autosegmental element independent of the lexical tone element, although these two elements are expressed by the same F0 curve. However, the F0 curve conveys both the linguistic and paralinguistic functions [4, 5, 13, 14, 16]. Hence, to understand the encoding mechanism of expressive speech in speech communication, we examined the interplay between tone and intonation in Chinese emotional speech.

Xu, et al. [16, 17] suggested that emotional meanings are encoded along a set of benefit-oriented bio-informational dimensions which involve both segmental and prosodic aspects of the vocal signal. Xu et al. further argued that the bio-informational dimensions allow emotional meanings to be encoded in parallel with non-emotional meanings; thus there is unlikely to be an autonomous affective prosody.

Chao is one of the pioneers who studied Chinese emotional/expressive intonation. He proposed that the actual melody or pitch movement of a tonal language differs from the mere succession of the few fixed tones of that language. It is in fact a resultant of three elements: tone or etymological tone, the neutral intonation, and the expressive intonation, the latter two together forming sentence intonation. He emphasized that the expressive intonation depends on the quality of the voice, unusual degrees of stress (or weakness), general pitch of the whole phrase and tempo of speech [1].

Chao [1] distinguished at least two types of tone and intonation addition patterns: *simultaneous addition* and *successive addition*. The simultaneous addition refers to the tones that are the algebraic sums or the resultants of two factors: the original lexical tone and the sentence intonation proper. The successive addition refers to the clause that has a rising or falling intonation, which is not added simultaneously to the last syllables but added on successively after the lexical tones are completed. He described the successive addition rising ending (↗) and the falling ending (↘) with the following formulae:

T1: ↗55:= 56: ⌐        ↘55:=551: ⌐\
T2: ↗35:= 36: ∧       ↘35:= 351: ∧\
T3: ↗214:= 216: √    ↘214:= 2141: ∿\
T4: ↗51:= 513: √      ↘51:= 5121: ∨

Where T1~T4 are four lexical tones, 1~5 are tonal values in 5 tone-letter system [3], 6 represents extra high pitch. Numbers on the left of ':=' are the lexical tone values and the numbers on the right are the addition tone values. The most interesting effect is shown with the Qusheng (T4), resulting in a circumflex tone.

Chao even enumerated 40 intonation patterns to demonstrate the forms and the functions of the intonation by grouping them according to pitch/duration elements, voice quality and intensity elements [1, 2].

The successive addition tone has been found in expressive speech to signal the emotion-attitudinal messages by Mueller-Liu [11], and also has been found in the intonational questions to signal the interrogative mood [10].

In our study, we differentiate the expressive speech into emotional speech and attitudinal speech. In previous research, based on Wu Zongji's intonation theory [13, 14], we found that speakers use boundary tone H% to express happy and friendly

speech [8, 12], and that sentence stress shifts a lot across the observed emotions [6].

Our previous studies support the 'bio-informational dimensions' theory proposed by Xu [17] and the expressive intonation elements suggested by Chao [1]. In the present paper, we focus on the emotional intonation by analyzing the F0 and duration patterns of monosyllabic utterances in seven emotions. The boundary or edge tones were examined to demonstrate the tone and intonation addition pattern in emotional speech.

## 2. MATERIALS AND RECORDING

Data used here were from our emotional speech corpus. A set of 111 sentences with varying lengths (from 1 to 14 syllables), sentence types (narrative or interrogative) and structures were recorded. The contents of all these sentences were emotionally 'neutral'. The monosyllabic sentences covered the combinations of 4 tones and all the vowels. A male and a female professional voice actors were recruited to produce the sentences in 7 kinds of emotions: Disgust(D), Sad(S), Angry(A), Happy(H), Surprise(SU), Fear(F) and Neutral(N). Sampling rate and resolution were 16KHz and 16bits.

To clearly illustrate the tone and intonation addition patterns in the present study, we chose 36 monosyllabic sentences coving 9 mandarin vowels and 4 lexical tones for each speaker, altogether 9syllables × 7emotions × 4tones × 2 speakers = 504 monosyllabic emotional utterances. All the F0 and duration data of these utterances were extracted by using Praat and manual checking.

F0 values of the tonal bearing part of each syllable were normalized into 10 points. Then, they were transformed into Semitone scale with the reference frequency of 75Hz. Finally, all the F0 values were mapped into a 5-tone value space [3, 13, 14].

## 3. FUNDAMENTAL FREQUENCY

Figure 1 shows the F0 patterns of 7 emotional utterances in 5 letter tone space. Compared to neutral emotion, the tonal pattern varies greatly among 7 emotions in aspects of tonal range, tonal register and tonal contours. The results show that the tone patterns are consistent within the two speakers for all of emotions, except "Anger" where the male speaker has a much shrunken pattern in amplitude than the female speaker.

Figures 2 and 3 depict F0 ranges and registers of the 7 emotions for the two speakers. For the male speaker, it illustrates that 'Sad' and 'Disgust' are between 1 to 3 (here '1' stands for F0 values between 0~1, '3' for values between 2~3…etc.), with 'Neutral' and 'Fear' between 1 to 4 and 'Happy' and 'Angry' between 1 to 5. Therefore, 'Sad' and 'Disgust' have narrow band intonation in lower
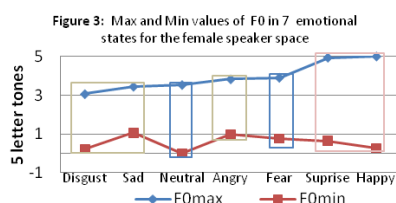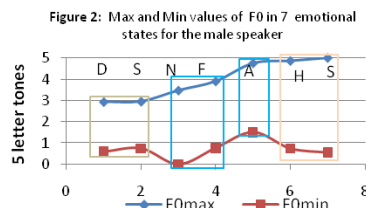
register, 'Neutral' and 'Fear' have medial band intonation in medial register, while 'Angry' and 'Happy' have broader band intonation in higher register. For the female speaker, the F0 values are almost the same as those of the male speaker except for 'Angry', which is expressed with a reduced pitch range in medial register.

**Figure 1:** F0 of 7 emotional utterances with 4 tones in 5-tone sale (female: left column; male: right column).
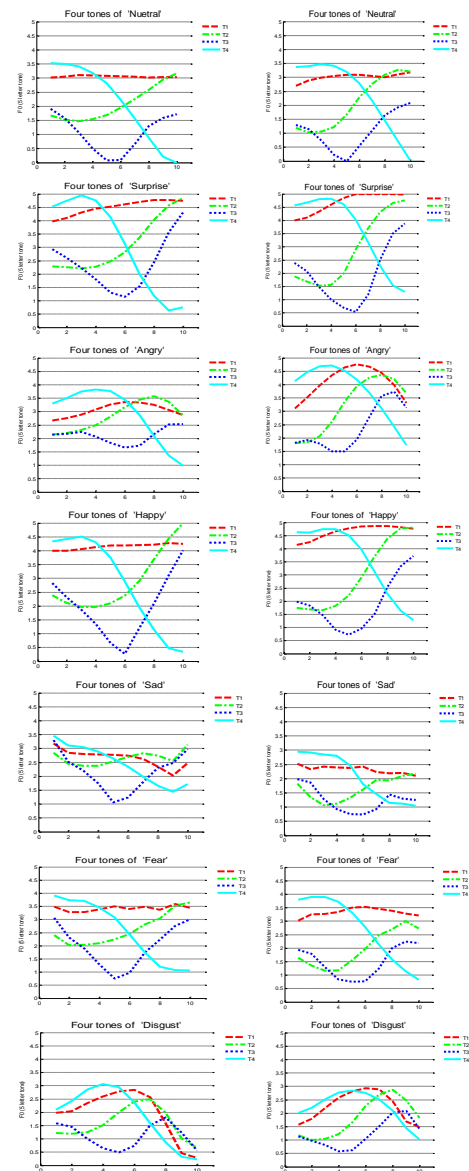


However, besides the differences in F0 ranges and registers, Fig. 1 also shows that the 'additional parts' of the emotional tones differ greatly from the 'Neutral' tone, and that the two speakers showed consistent patterns for each emotional tone. These additional parts here refer to the successive addition tones following the end of the lexical tones of the syllables. The additional parts express emotions

rather than linguistic meanings such as interrogatives as described in [9, 10].



**Figure 2:** Max and Min values of F0 in 7 emotional states for the male speaker



**Figure 3:** Max and Min values of F0 in 7 emotional states for the female speaker space

The contour features of the edge tones are different across 7 emotions: 'Surprise' and 'Happy' have a 'Rising' tone (in contrast to 'Neutral'); for falling T4 (HL), the rising feature changes the slop of the falling line and a very tiny rising tail is observed; for the three other tones with H ending, there are additive raising tones. 'Disgust' and 'Angry' have a 'Falling' addition tone following the lexical tone of the last syllable and can be obviously assembled as the successive addition tones (as mentioned by Chao [1]). For the male speaker, the 'Sad' has a level offset tone, while a slightly rising tone is present for the female speaker.

**Table 1:** Tone values for the male/female speaker in 7 emotions.

| T | S | H | A | F | Su | D | N |
|---|---|---|---|---|---|---|---|
| 1 | 33/43 | 55/55 | 454/343 | 44/44 | 55/45 | 232/232 | 44/44 |
| 2 | 23/33 | 25/35 | 254/343 | 23/23 | 25/35 | 132/132 | 24/24 |
| 3 | 212/423 | 214/314 | 2244/323 | 213/213 | 314/324 | 1132/1132 | 213/212 |
| 4 | 32/42 | 52/51 | 52/42 | 41/41 | 52/51 | 332/332 | 41/41 |

**Table 2:** Features of tonal space and successive addition tones.

| Em. | span | register | range | Add. T |
|---|---|---|---|---|
| S | 1-3 (LM) 1-4 (LM) for Female | L | 3 (L) | R/Le |
| H | 1-5 (LH) | H | 5 (H) | R |
| A | 2-5 (LH) 2-4 (LH) for Female | H M (for Female) | 4 (M) 3 (L for Female) | F |
| F | 1-4 (LH) | M | 4 (M) | - |
| Su | 1-5 (LH) | H | 5 (H) | R |
| D | 1-3 (LM) | L | 3 (L) | F |
| N | 1-4 (LH) | M | 4 (M) | - |

Table 1 describes the four tone values in 5-level scale and Table 2 summarizes the tonal features of the tonal range, register and successive addition tone in phonological features. 'R, F and Le' are used to describe the 'Rising', 'Falling' or 'Level' successive addition tones, and '-' for no successive addition tone.

## 4. DURATION

Figure 4 shows the average durations of four tones (T1~T4). Figure 5 depicts the duration ratio of emotional tones to the neutral tones. For the female speaker, 'Fear', 'Sad' and 'Happy' have longer duration than 'Neutral', while for the male speaker, only 'Sad' is longer than 'Neutral'. 'Angry' and Surprise' are the shortest for both speakers.



**Figure 4:** Duration of 4 tones in 7 emotions



**Figure 5:** Ratio of Emotional duration: Ei/EN
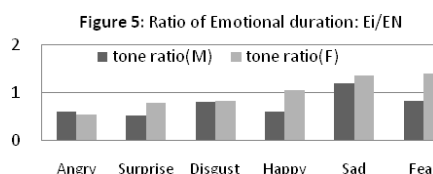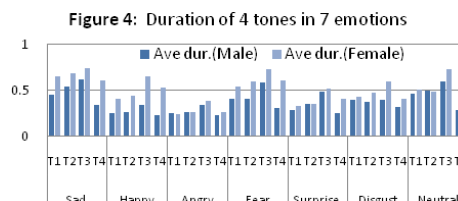
Most of the duration patterns are consistent with many previous studies on emotional speech, except that the female speaker expressed 'Happy' without using a faster speech rate.

As for those which have successive addition tones, they don't have longer durations to accommodate these additive tones. The durations for 'Fear' and 'Sad' are longer than for other emotions. As shown in Figure 1, in comparison to the 'Neutral' productions, some of the final successive addition parts account for 1/4~1/3 of the whole tone without lengthening the tone.

## 5. SUMMARY AND DISCUSSIONS

Compared to the neutral emotion, the tone pattern varies significantly among 7 emotions in tonal range, tonal register and tonal contours. Except for "Angry", the tone patterns of each of the other emotions are similar for the two speakers.

'Fear' has an obvious tremor voice, higher F0 and narrower pitch range than the neutral emotion. 'Sad' is characterized by reduced pitch range, lower F0. 'Disgust' is similar to 'Sad', but the offset tone is a kind of successive addition falling tone. 'Happy' and 'Surprise' are comparable, with higher pitch range and register, and rising offset tone. 'Surprise' has higher bottom pitch, whereas a slightly narrower range than 'Happy'. 'Surprise' is expressed a bit faster than 'Happy'. Both 'Happy' and 'Angry' have higher pitch and faster speech rate for the two speakers. Although these patterns may make them difficult to recognize, they can be easily distinguished through the offset tone.

The speakers may use different strategies to express the same emotion. For the male speaker, the 'Sad' has a kind of level offset tone, whereas it has a slightly rising tone for female speaker. The female speaker applied lower F0 register for 'Anger' and voice with more jitter for 'Fear'. Although most of the emotions have consistent duration relation with neutral speech for the two speakers, some differences still exist. For example, the female's 'Happy' speech is not as fast as the male's, and it is a bit slower than 'Neutral' speech.

Emotional intonations of monosyllabic utterances show the tone and intonation addition patterns which are different from what we have found in neutral speech [9, 13, 14]. The speakers express some emotions like 'Disgust' or 'Angry' by using a kind of falling successive addition tone and 'Happy' and 'Surprise' by a rising one. This part does not belong to the lexical tone of the attached syllable while it is produced to express the speakers' emotions. The additive tones occupy 1/4~1/2 length of the whole tone without lengthening the duration of the boundary syllable, while the successive addition tone is longer when it is used to transmit linguistic information [9], so the encoding mechanism may depend on emotional prosody.

For longer sentence like 'Play golf' shown in Figure 6, the boundary tones keep the same patterns as we discovered: 'Happy', 'Surprise' and 'Angry' have a higher boundary, while others have a lower boundary. Most significantly, there are still rising or falling final successive addition tones for some emotional intonations.

The boundary tone of Chinese is realized differently from other languages such as English and Dutch [4, 9]. Lin summarized the F0 patterns of Chinese boundary tones of statements (L%) and interrogates (H%): the contours of the boundary tones keep their lexical tonal contours unchanged, H% is realized by raising the register or the slop of the tonal contour slightly[9], which belongs to the simultaneous addition suggested by Chao[1].

If we use traditional boundary features H% or L% to describe the present emotional boundary tones, i.e. the successive addition tones, they will not be sufficiently described in the framework of the current phonological theory. Therefore, we suggest to combine traditional boundary 'H% or L%' and successive addition tone feature 'r, f or le' to account for the successive addition boundary tones of Chinese emotional intonations. For example, boundary tones of 'Disgust' and 'Angry' in Figure 6 can be described as 'L-f%' and 'H-f%', 'Happy and Surprise' as 'H-r%' respectively.

In subsequent research we plan to use the computing PENTA model to simulate these emotional intonations, especially the successive addition tones [15]. Our goal is to tease apart the affective components from the lexical components.

**Figure 6:** F0 contours for the same sentence 'da3 gao1 er3 fu1' (Play golf.) in 7 emotions (male speaker).
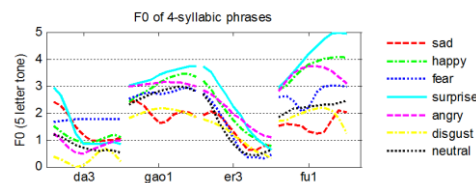


## 6. ACKNOWLEDGEMENTS

## 7. REFERENCES

[1] Chao, Y.R. 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology* 4, 121- 134

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

[3] Chao, Y.R. 1980. A system of tone letters, *Fangyan* 2.

[4] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: CUP.

[5] Ladd, D.R. 1996. *Intonational Phonology*. Cambridge: CUP.

[6] Li, A. 2008. Stress pattern of emotional speech. *Chinese Journal of Phonetics.* Commercial Publishing House.

[7] Li, A., Fang, Q., et. al. 2010. Acoustic and articulatory analysis on Mandarin Chinese vowels in emotional speech. *Proc. Of ISCSLP* Tainan.

[8] Li, A., Wang, H. 2004. Friendly speech analysis and perception in Standard Chinese, *Proc. of ICSLP* Jeju, Korea.

[9] Lin, M. 2006. Interrogative mood and boundary tone in Chinese. *Journal of Chinese Linguistics* 4.

[10] Lu, J., Lin, M. 2009. Raising tail of the question vs. the modal particle. In Fant, G., Fujisaki, H., Shen. J. (eds.), *Frontiers in Phonetics and Speech Science*. Commercial Publisher.

[11] Mueller-Liu, P. 2006. Signalling affect in Mandarin Chinese - the role of utterance-final non-lexical edge tones. *Proc. 5th of Speech Prosody,* PS6-3-0048.

[12] Wang, H., Li, A., Fang, Q. 2005. F0 contour of prosodic word in happy speech of Mandarin. *Proc. of ACII 2005,* Beijing.

[13] Wu, Z. 1997. A new method of Standard Chinese intonation processing based on the relation between tone and melody. *Collected Essays Dedicated to the 45th Anniversary of the Institute of Linguistics*. CASS. Commercial Press, Beijing.

[14] Wu, Z. 2000. From traditional Chinese phonology to modern speech processing-realization of tone and intonation in Standard Chinese. *Proc. of the ICSLP 2000* Beijing.

[15] Xu Y. 2011. PENTAtrainer. *http://www.phon.ucl.ac.uk/home/yi/PENTAtrainer/*

[16] Xu, Y., Chuenwattanapranithi, S. 2007. Perceiving anger and joy through the size code. *Proc. ICPhS* Saarbrücken.

[17] Xu, Y., Kelly, A., Smillie, C. Forthcoming. Emotional expressions as communicative signals. In Hancil, S., Hirst, D. (eds.), *Prosody and Iconicity*.