Successive Addition Boundary Tone of Chinese Emotional Intonation: Production and Perception

Aijun Li
Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China

‘Successive addition boundary tone’ (SUABT), proposed by Chao (1933) to convey pragmatic function, was found in emotional intonation of Mandarin Chinese, when speakers express ‘disgust’ or ‘anger’ by using a kind of ‘falling’ SUABT, and ‘happiness’ or ‘surprise’ by a kind of ‘rising’ SUABT. In this case, the boundary tone is composed of two components, the lexical tone to encode linguistic meaning, and the expressive tone to express emotional attitude or pragmatic meaning. In the present study, the intonation patterns and the SUABT were analyzed phonetically and phonologically. Then a perceptual experiment was conducted to separate the interaction between the form of Chinese SUABT and its expressive function, and to determine whether it can uniquely encode the emotional or pragmatic information. Through the GLM (General Linear Model) analysis on the perceptual results, we found that the acoustic features, length of sentence, and final tone category have significant effect on emotional expressiveness, but their contributions vary with different emotions. The mapping between the form of SUABT and its pragmatic function is a complex many-to-many relation. The results further revealed that SUABT is not the only element speakers use to express emotions; other components could be invoked to convey expressive information.
Index: Chinese intonation, emotion, expressiveness, boundary tone, successive addition boundary tone

1. INTRODUCTION

In intonation phonology, intonation is described as a chain of pitch accents well organized under intonation grammar known as the autosegmental and metrical approach (AM) or the Pierrehumbert model (1980, 1990). The intonation is also considered to eventually link communicative meanings or paralinguistic or pragmatic meaning (Bolinger 1989; Gussenhoven 2002; Ladd 1996;Wu 1995, 2000; Xu 2007, 2011)

Gussenhoven (2002) proposed a relatively more elaborate theory that bases intonational meaning on a set of "biological codes": frequency, effort, and production.

Liu and Xu (2005) argued that communicative meanings are conveyed through a set of separate functions that are realized by an articulatory system with various biophysical properties. This view of speech melody is summarized into a comprehensive model of tone and intonation, namely, the parallel encoding and target approximation (PENTA) model.

Boundary tone, as an important component to convey linguistic and pragmatic information, is proposed by Pierrehumbert (1980) as a phonological unit for representing the internal makeup of an intonation phrase. Pierrehumbert (1980) and Xu (2005) suggested that the most obvious F0 pattern associated with a yes-no question is the final rising, which has been attributed to a high boundary tone H% in the AM theory. Pierrehumbert & Hirschberg (1990) gave a list of functions of pitch accents and boundary tones with reference to discourse structures. Vered Silber-Varod (2011) concentrates on the continuous (C)-boundary inventory in a corpus of spontaneous Israeli Hebrew and investigates the linkage function of the communicative value of the C-boundary tone according to the syntactic relations between the word preceding and following each of the C-boundaries.

Turning to Chinese, Chao (1932, 1933, 1968) 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.

Chao (1933) 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 his tone-letter system (1980), and '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 Falling tone (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 (1933,1968).

Lin (2004, 2011) investigated how 'simultaneous addition' is realized in the sentence-final syllables to convey intonational meanings. Yuan (2006) proposed three mechanisms of question intonation in Mandarin Chinese: an overall higher phrase curve, higher strengths of sentence final tones, and a tone-dependent mechanism that flattens the falling slope of the final falling tone and steepens the rising slope of the final rising tone. Jiang and Chen (2011) supported the idea that Mandarin interrogative cues distribute over an entire utterance. High pitch at both edges of an intonation phrase should be marked out. Employing a corpus of conversational speech in Mandarin, Patrick Callier (2011) investigated the boundary tone of intonational meaning and function in a socio-phonetic way, checking the H% or L% boundary tone by clause type, illocutionary force, and speaker gender. Few studies have been conducted for successive addition boundary tone relating to pragmatic meaning of intonation. Mueller-Liu (2006) listed some successive addition tones in expressive speech to signal the emotion-attitudinal message, and Lu and Lin (2007) also found it in the intonational questions to signal the interrogative mood.

In our study, we differentiate the expressive speech into emotional speech and attitudinal speech. In previous research, based on Wu Zongji's intonation theory (1990, 1995, 2000), we found that speakers use boundary tone H% to express happiness and courtesy in speech (Li et al. 2004, Li 2008), and that sentence stress shifts a lot across the observed emotions (Wang, Li and

Fang 2008).

In another study (Li, Fang and Dang 2011), we found that the successive addition boundary tones (SUABT) are employed by speakers to convey expressive information, such as ‘disgust and anger’ by a kind of ‘falling’ successive addition tone, and ‘happiness or surprise’ by a kind of ‘rising ’successive addition tone, as pointed out by Chao. As shown in Figure1, the boundary tones of ‘disgust and anger’ exhibit a falling tail compared to that of ‘neutral’ intonation, keeping the first part as its lexical tone.

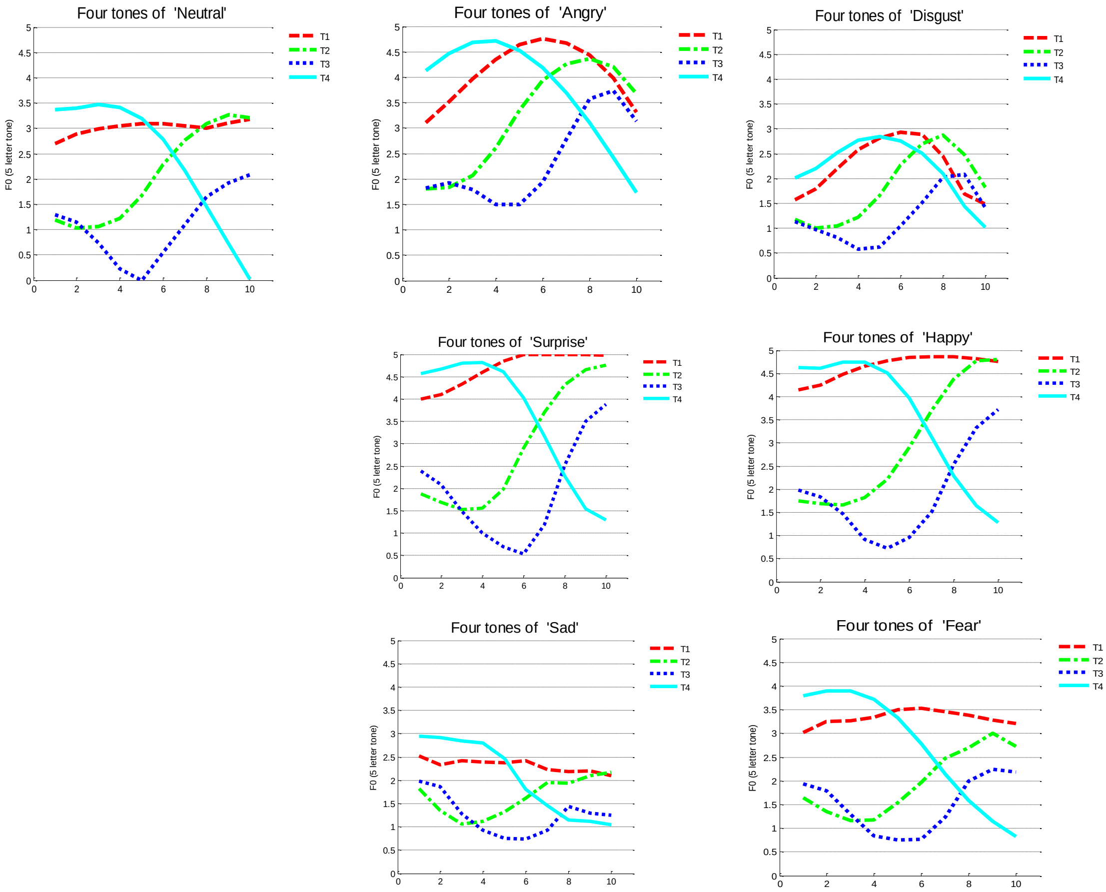


Figure 1: *A male speaker’s ‘Neutral’ and six emotional intonations of monosyllabic utterances with four lexical tones (T1~T4),* F_0 *normalized in 5 tone-letter scale, and tonal duration normalized into 10 points.*

In the present paper, the acoustic patterns of Mandarin emotional intonation were analyzed on more data: disyllabic utterances and longer utterances. Then, the tone and intonation addition patterns of the boundary tones, especially the successive addition boundary tone (SUABT), were analyzed phonetically and phonologically. Finally, a perceptual experiment was made to separate the interaction between the pragmatic function of the SUABT and its acoustic form, and to determine whether the SUABT can be independently encoded to express emotional attitude and whether the mapping between the acoustic form and the expressive function is unique.

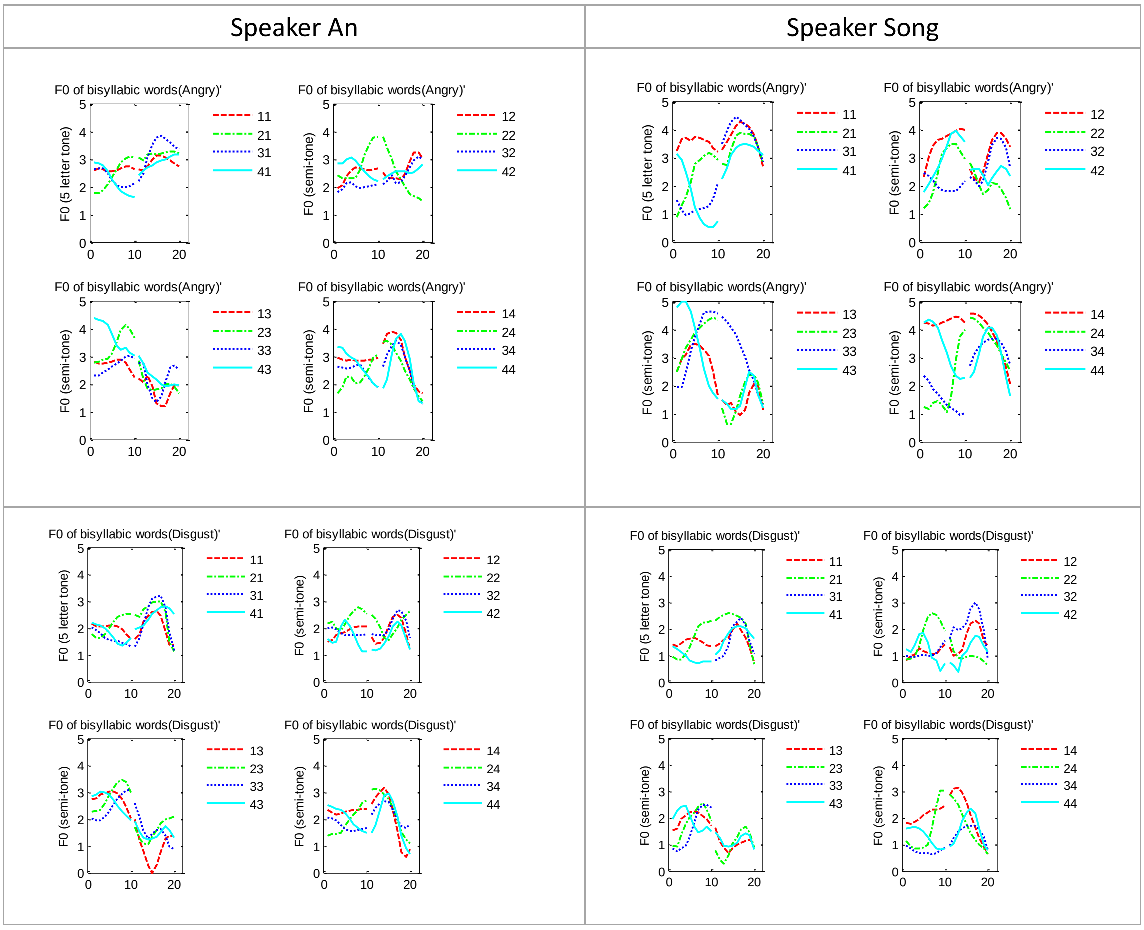
2. MATERIALS AND RECORDING

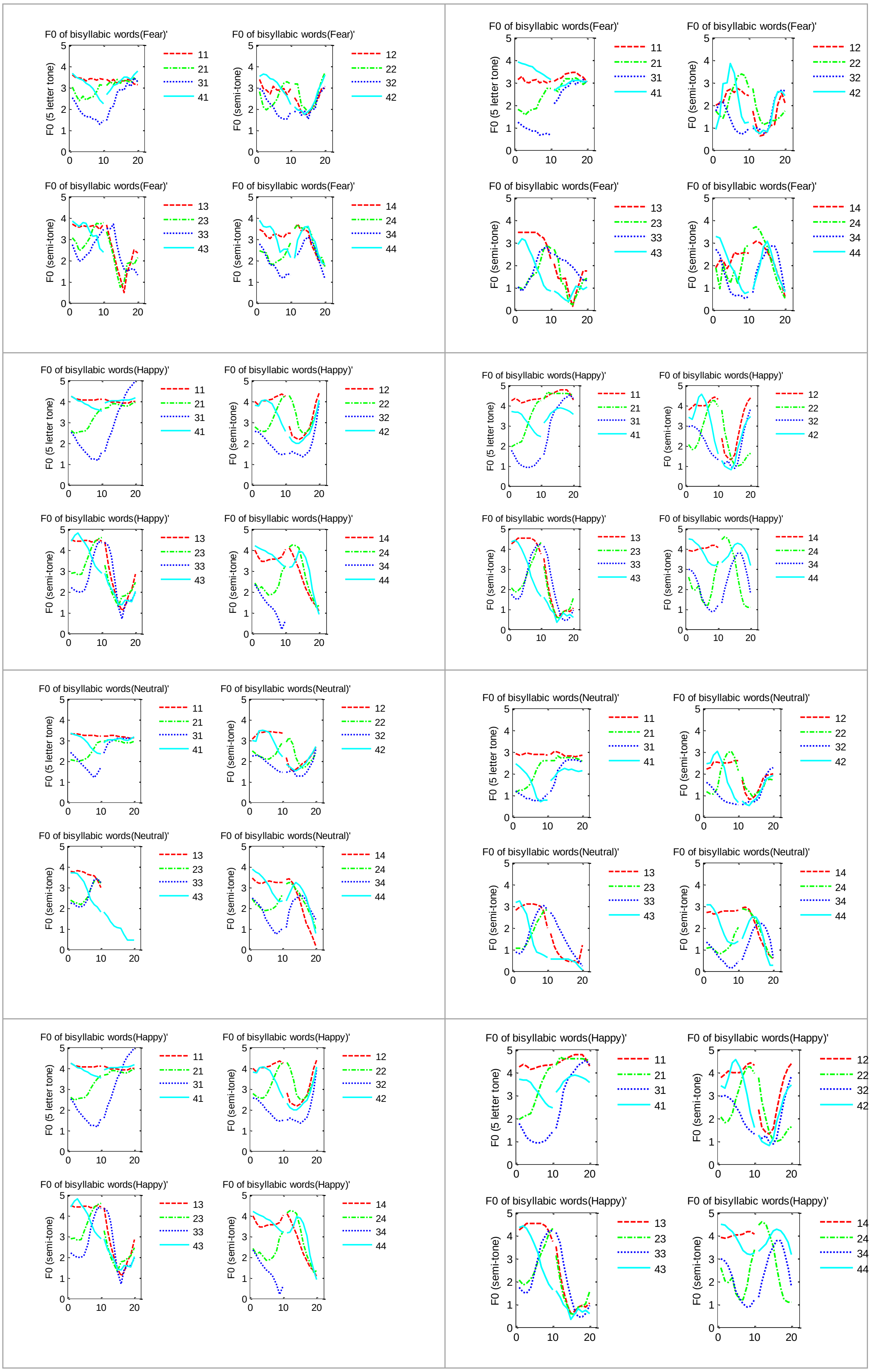
Data used here were from our emotional speech corpus Emotio-CASS. A set of 111 sentences with varying lengths (from 1 to 14 syllables), sentence types (narrative or interrogative) and structures were recorded. The contents of all these sentences were emotionally 'neutral'. The monosyllabic sentences covered the combinations of four lexical tones (T1~4) and all the vowels. The disyllabic sentences covered 16 tonal combinations (T1~4 ×T1~4) , among which T3T3 have the same surface pattern as T2T3 for the phonologyical tone sandhi. A male and a female professional voice dubbing actors were recruited to produce the utterances in seven kinds of emotions: Disgust(D), Sad(Sa), Angry(A), Happy(H), Surprise(SU), Fear(F) and Neutral(N). The sampling rate and resolution were 16KHz and 16bits respectively.

All the utterances were annotated with initial and final boundaries in Praat. F0 data of these utterances and segments (syllable, initial and final) were extracted by using Praat and manually checked. F0 values of the tonal bearing part (final) 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 (Chao 1980, Wu 2000).

3. INONATION PATTERNS AND SUABT OF THE EMOTIONAL UTTERANCES

3.2 Bisyllabic intonation





F0 of bisyllabic words(Fear)'
F0 (5 letter tone)
F0 (semi-tone)
11 21 31 41
12 22 32 42
13 23 33 43
14 24 34 44
F0 of bisyllabic words(Happy)'
F0 of bisyllabic words(Neutral)'
0 10 20
0 1 2 3 4 5

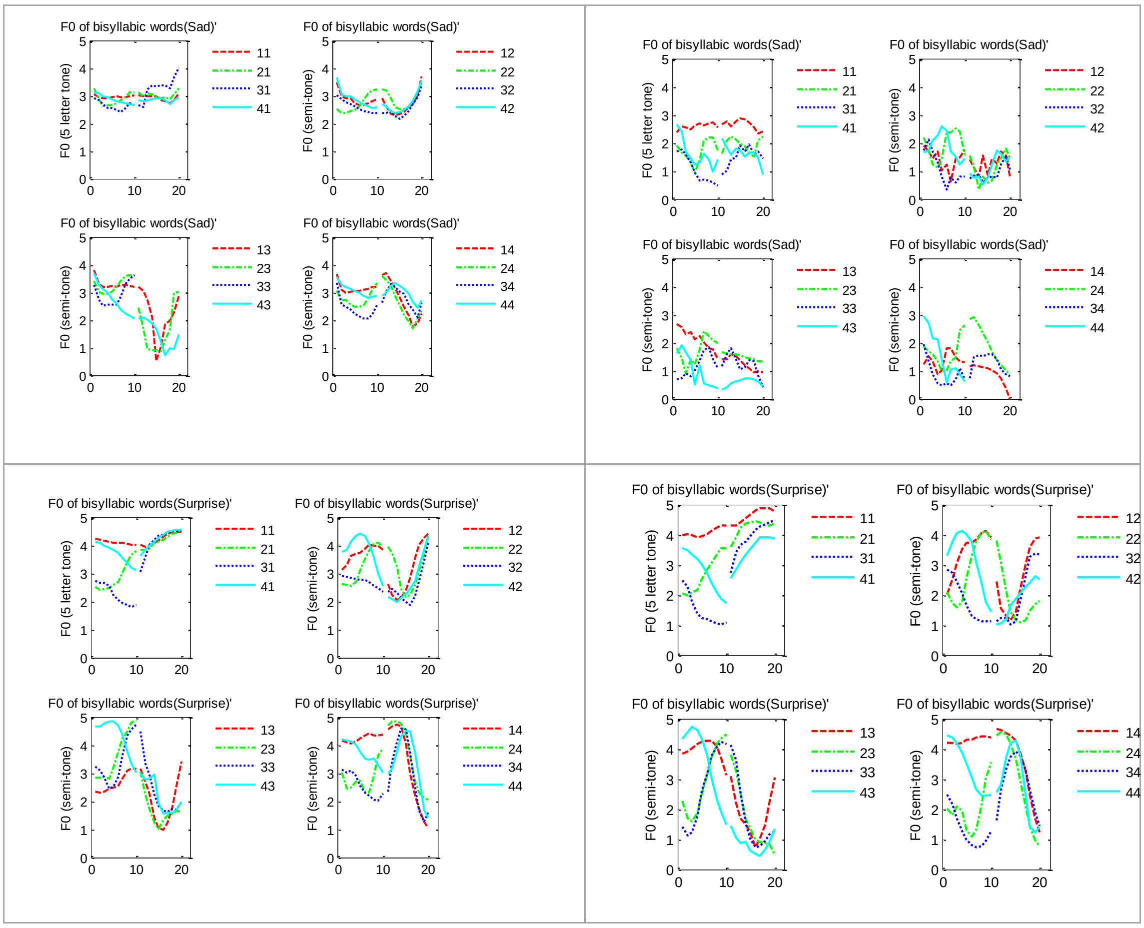


Figure 2: *The F0 Patterns of disyllabic utterances in seven emotions, disyllabic tones were normalized in 20 points and the F0 data are normalized into 5 tone-letter scale. The tonal combinations are organized according to the first tonal category. For instance, '31' refers to the combination of first syllable in T3 and the boundary tone as T1.*

In previous study (Li, Fang and Dang 2011), we analyzed the F0 features of these two speaker's emotional intonation for monosyllabic utterances. Compared to the 'Neutral' emotion, the tone pattern varies significantly across seven emotions in tonal range, tonal register and tonal contours. Except for "Angry" emotion, the tone patterns are similar between the two speakers. 'Fear' has an obvious tremor voice, higher F0 and narrower pitch range than the 'Neutral' emotion. 'Sad' emotion is characterized by reduced pitch rang and lower F0. 'Disgust' emotion is similar to 'Sad' emotion, but the boundary tone is a kind of successive addition falling tone. 'Happy' and 'Surprise' emotions are comparable, with higher pitch range and register, and rising boundary tone. 'Surprise' emotion has higher bottom pitch, but a slightly narrower range than 'Happy' emotion. It is also expressed a bit faster in speech rate. Both 'Happy' and 'Angry' emotions 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 boundary tones.

Emotional intonations of monosyllabic utterances exhibit the tone and intonation addition patterns different from what we have found in 'Neutral' speech (Lin 2004, Lin & Li 2011; Wu 1995, 2000). The speakers express some emotions like 'Disgust' or 'Anger' by using a kind of falling successive addition tone and 'Happiness' and 'Surprise' by a rising one. This part does not belong to the lexical tone of the syllable when it is produced to express the speaker's 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 (Li et al, 2011, 2012), so the encoding mechanism may depend on emotional prosody.

In the present study, F0 features of disyllabic utterances were investigated and compared with those of the monosyllabic utterances across seven emotions. The F0 patterns, grouped according to the tone of the first syllable in each emotion (7 subplots), were plotted in Figure 2, and they reveal the following:

(1) The F0 patterns of 'Neutral' intonation are the same as described by Wu (2000) and boundary tone patterns the same as described by Lin (2004).

(2) Comparing to 'Neutral' intonation, all the boundary tones (T1~T4) of 'Disgust and Anger', have an additive falling boundary, the typical SUABT with the first part as lexical component and the second part the expressive component to express 'Disgust or Anger', similar as one syllabic utterance shown in Figure 1.

(3) Comparing to the 'Neutral' intonation, the F0 patterns of 'Happiness and Surprise' have a rising SUABT with broader F0 range. The 'rising' feature makes the H tone (T1:HH; T2:LH,; T3:LLH) raise higher with a steeper slope, while for T4(HL), the slope become smaller (flatter). 'T2+T2' and 'T2+T3' have almost overlapped contours caused by the emotional expression.

(4) Comparing to the 'Neutral' intonation, the F0 patterns of 'Sad and Fear' intonation have some variations on the F0 range, in addition to distinctive quivers observed on F0.

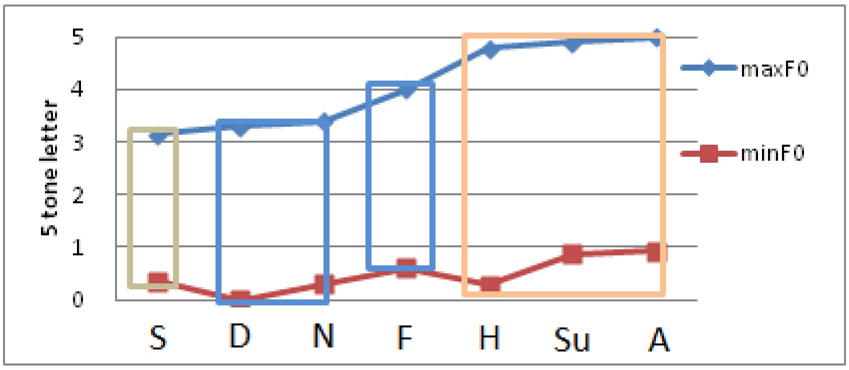


Figure 3: *F0 range variations of disyllabic utterances for 7 emotions (male speaker). 'maxF0' and 'minF0' are maximum and minimum values of F0*

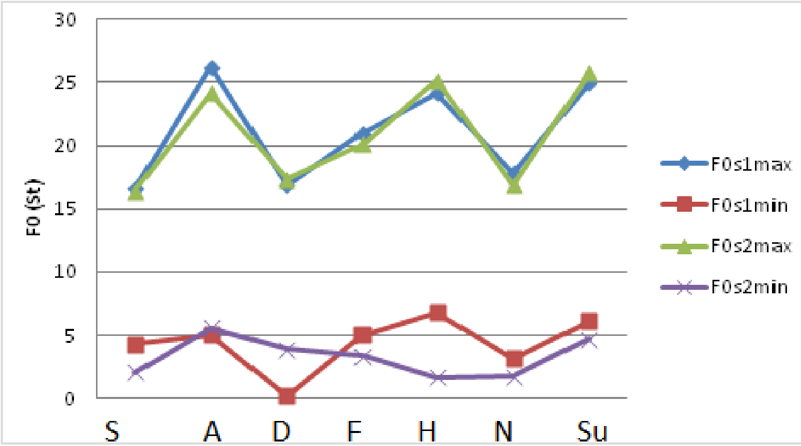


Figure 4: *F0 range variations of the first and the second syllable for 7 emotions (male speaker). 'S1' and 'S2' denotes the first and second syllable respectively.*

Figures 3 and 4 depict the F0 range variations of the whole utterances and the individual syllables for male speaker (female speakers' data are similar and omitted here). Figure 3 shows that the F0 range can be categorized into three levels: Narrow for 'Sad '(range < 3), Middle for 'Neutral, Disgust and Fear' (range between 3 to 4) and Broad for 'Happy, Surprise and Angry' (range between 4 to5) intonations. The register of the intonation (maxF0) can also be categorized into three levels: low register for 'Sad, Disgust' (3~3.5), middle register for 'Fear and Neutral' (3.5~4) and high register for 'Angry, Happy and Surprise' (4~5). Here the 'Angry' intonation has a broader range, while others keep the same as the monosyllabic intonations (Li et al. 2011).

For the F0 variations of the two syllables (Figure 4), the maxF0 of the two syllables are almost the same, but the 'minF0' is lower for the second syllable except 'Disgust and Angry' intonation. The 'Happy and Surprise' intonation come with broader F0 range on the final syllable, but the 'Angry and Disgust' have narrower F0 range on the final syllable.

3.3 F0 patterns and SUABT of longer utterances

We checked the F0 patterns of longer utterances for 7 emotions, see Figures 5~7. The F0 ranges and registers keep the same patterns in different contexts as those of the disyllabic utterances. 'Disgust and Sad' intonation have low register, 'Neutral and Fear' have middle register, and 'Happy, Angry and Surprise' have high register. 'Sad' has narrow range, 'Neutral, Disgust and Fear' have middle range, while 'Happy, Angry and Surprise'' have broad range.

For SUABT, the patterns should be categorized by the lexical tone of the boundary syllable. For boundary syllable whose lexical tone has a 'H' offset tone (T1~T3), as in syllables 'fu1' and 'jing1' (T1:HH) in Figures 6 and 7, a falling tone is added after the H tone for 'Disgust and Angry', and a rising tone is added after the H tone for 'Surprise and Happy'. 'Fear and Sad' intonation shows a tremble F0. For boundary syllable with lexical tone 4 (HL), the offset tone is L tone. In this case, the boundary tone keeps falling for all emotions with different slopes. See 'hui4' and 'sai4' in Figures 5 and 6, but a closer inspection shows that 'Surprise' intonation has a small rising tail.

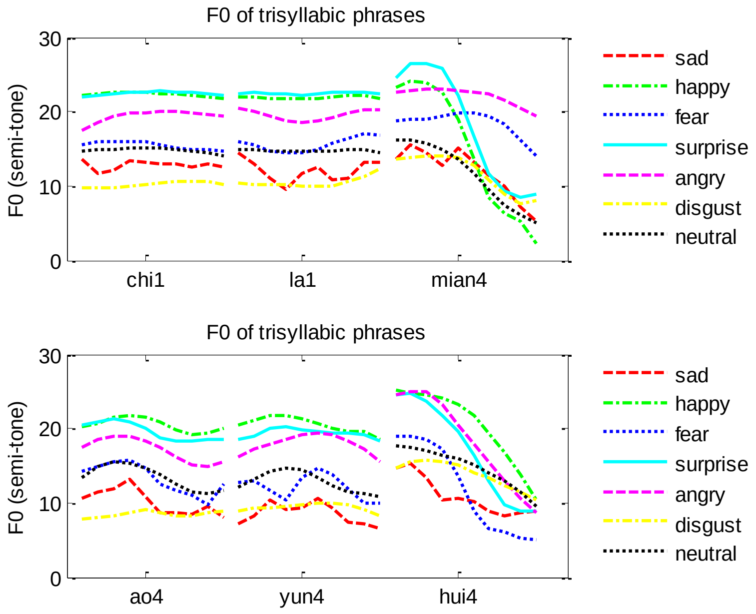


Figure 5: *F0 of trisyllabic utterances in 7 emotions（male speaker）*

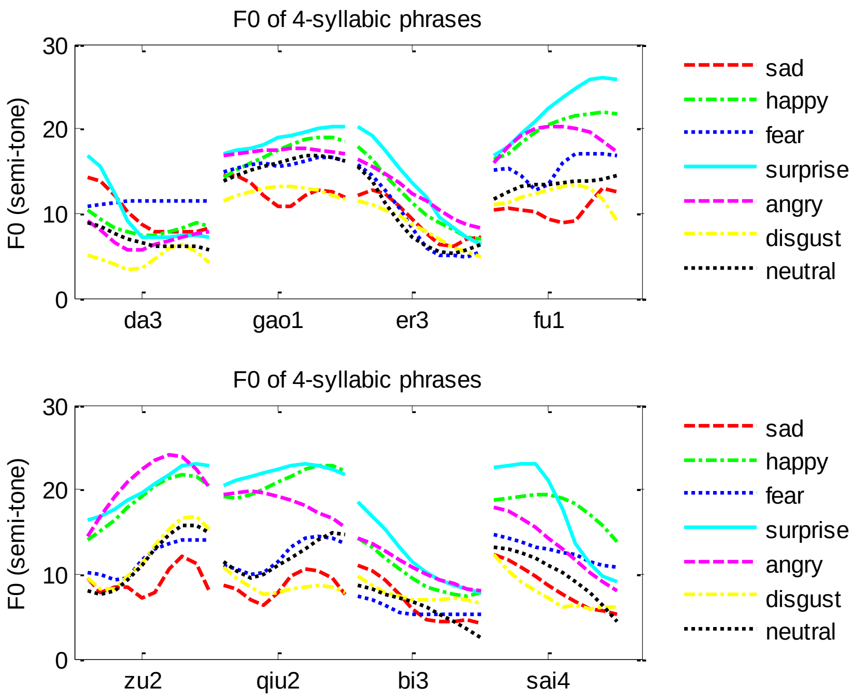


Figure 6: *F0 of quadrisyllabic utterances in 7 emotions (male speaker)*

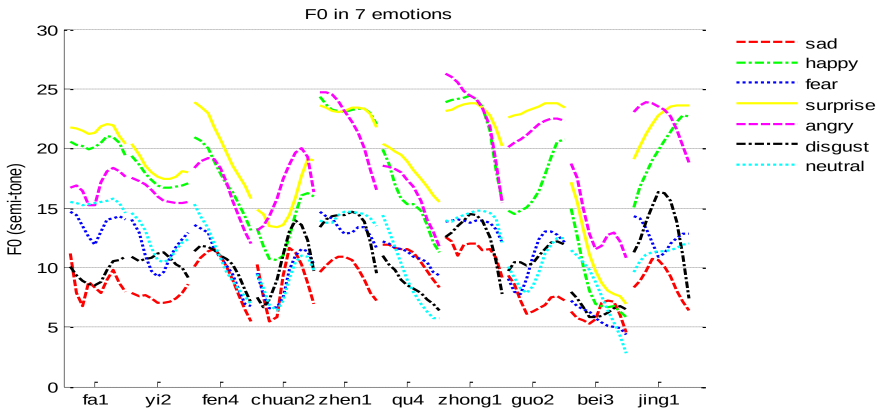


Figure 7: *F0 of utterances in 10-syllable length in 7 emotions (male speaker)*

3.4 Phonological representation of emotional intonation and the *SUABT*

Based on the previous analysis, Table 1 summarizes the intonation patterns and SUABT of 7 emotions in phonological features. 'H, M and L' are features to describe pitch range and register. We suggest combining traditional boundary features 'H% or L%' and successive addition tone features 'r, f or le' to describe the successive addition boundary tones of Chinese emotional intonations. For example, boundary tones of 'Disgust' and 'Angry' intonations can be described as 'L-f%' and 'H-f%', and 'Happy and Surprise' intonations as 'H-r%' respectively, where 'x' stands for no successive addition tone.

Table 1: F0 *features of emotional intonation and SUABT*

Emotions	Range	Register	Additive tone	Boundary tone transcription
Sad	L	L	x	L-%
Happy	H	H	r	H-r%
Angry	H	H	f	H-f%
Fear	M	M	x	L-%
Surprise	H	H	r	H-r%
Disgust	M	L	f	L-f%
Neutral	M	M	x	L-%

3.5 Schematic representation of SUABT

As shown in Figure 8, a schematic presentation is proposed to describe the SUABT which is composed of two components, the lexical tone to express lexical meaning, and the expressive tone to express emotional attitude. The expressive tone has two patterns: a rising tone to express 'Surprise and Happy' emotions, a falling tone to express 'Anger and Disgust'.

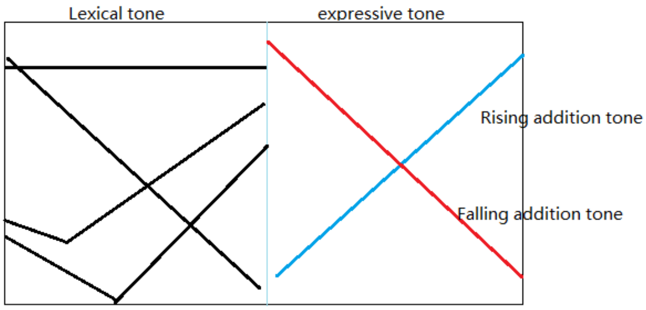


Figure 8: *A schematic representation of additive boundary tone with two tonal components: lexical tone and expressive tone*

4 Perceptual experiment on the expressiveness of SUABT

We would like to explore the 'gradient nature' (Ladd, 1996) that the additive boundary tone contributes to emotional expression. In order to exclude expressive factors other than SUABT, emotionally neutral utterances were used as the baseline data and an addition tone was successively added to final boundary tone of the neutral utterances. The successive falling addition tone was selected in the experiment. The stimuli were obtained by varying the acoustic features of the falling expressive tone.

4.1Stimuli preparation

Table 2 lists the nine emotionally neutral sentences used in the experiment, which have diverse length and final tone categories. Additive falling tones are simulated based on the variation of the acoustic parameters for 'Disgust' intonation of a male speaker (Li, Fang, et al, 2012). In order to separate the additive falling boundary tone from the original lexical tone, and to clearly and easily manipulate the lexical and the additive parts respectively, T1~T3 were selected as the boundary lexical tone and T4 was excluded.

Table 2: *Nine sentences used in the perceptual experiment*

#	Content	Length (sylls)	Final syll.	Final tone
1	一。One	1	yi1	T1
2	姨。Aunt	1	yi2	T2
3	椅。Chair	1	yi3	T3
4	老翁。Old man	2	weng1	T1
5	母羊。Ewe	2	yang2	T2
6	熊蕊。Stamen	2	rui3	T3
7	老周买了五斤海参。Mr. Zhou bought 2.5kg of sea cucumbers.	8	shen1	T1
8	去年盖的二层小楼。Two-story building built last year.	8	lou2	T2
9	长篇小说梅娘曲。long novel Mei Niang Qu	7	qu3	T3

The schematized representation of a falling SUABT is given in Figure 9. The rising part represents the lexical tone while the falling part represents the expressive tone. The acoustic features of the male speaker's 'Disgust' boundary tone were analyzed according to this structure (Li, et al, 2012). The features that we are concerned about here are the duration vs. F0 slope of the additive falling tone, and the duration ratio of the additive falling tone to the preceding lexical tone.

The maximum absolute slope value of the additive falling tone is 110st/s (F0ref=75Hz); the maximum duration ratio of final falling to the preceding tone is 3.5, while the absolute slope value is less than 65st/s and ratio less than 1.4 within the 95% confidence interval.

When generating the stimuli, we added a falling tail to the neutral boundary tones of the utterances in Table 2. So the new boundary tone consisted of two parts: the first part was the original neutral boundary tone, and the second part was the additive falling tone varying in lengths and slopes based on the acoustic data for the 'Disgust' emotion. The range of additive falling slope *ki* is set from 0 to -80 st/s stepping in -10st/s, *i=1~9*. The duration ratio of the final fall to the preceding tone *dj/D* is set from 0 to 1.25 stepping in 0.25, *j=1~6*.

PSOLA synthesizer was used to generate the stimuli for all the given nine neutral sentences. The lowest F0 is set to be 75Hz in this process, therefore not all of the '*dj*' and '*ki*' were realized. Finally, we got 425 synthesized stimuli.

As shown in Figure 1, although both 'Disgust and Angry' intonations have falling SUABT, the pitch registers of them are different. The pitch register of 'Disgust' intonation is closer to 'Neutral' than 'Angry' intonations, so we hypothesize that the synthesized stimuli would generate a greater tendency to be perceived as a 'Disgust' emotion.

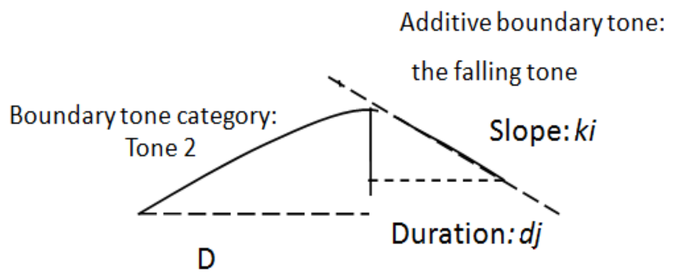


Figure 9: *A schematic representation of an additive falling tone with duration 'dj' and slope 'ki'.*

4.2 Procedure of the perceptual experiment

Twenty university students participated in the perceptual experiment, 5 male and 15 female. They speak standard Chinese and have no previous report on hearing problems.

All the stimuli were divided into 4 groups randomly (107 in 3 groups, 106 in the last group). Before starting the perceptual experiment, they were trained to use a perceptual program written in Praat script. The subject tried with 20 stimuli chosen from the emotional corpus covering 7 emotions, i.e. 'Neutral, Happy, Sad, Angry, Disgust, Surprise, and Fear', and evaluated the emotional expression for each stimulus. One or more emotions could be selected from the 7 emotions. Furthermore, they were allowed to write down other emotional or attitudinal expressions if they perceived them beyond these 7 emotions. The perceived score is set to 1 for the selected emotion; otherwise the score is set to 0.

4.3 Analysis on the perceptual experiment results

4.3.1Perceptual experiment results

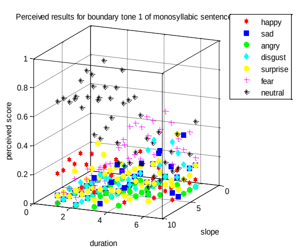
Figure 10 shows the average perceptual results on 7 emotions for 20 subjects, distributed along the two dimensions of slope and duration ratio scaled in corresponding step numbers respectively. The duration ratio steps from 1 to 6 and slope steps from 1 to 9.

For final boundary T1 in Fig.10(1)~(3), the results for monosyllbic, disyllabic and long utterances are depicted respectively. With the increasing of the absolute values of duration and slope for the additive falling tone, the perceived scores for ‘Neutral’ emotion decrease, and they increase gradually for other emotions. Surprisingly, both the scores for ‘Disgust’ emotion and the scores of ‘Angry and Fear’ emotion show clear increase.

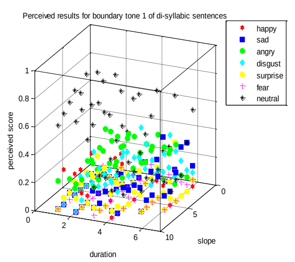
For final boundary T2 (Fig10.(4)~(6)), the results for monosyllabic utterances are ‘misunderstood’. It was perceived as ‘Surprise’ emotion rather than Neutral emotion when the absolute values of slope and duration are small. The reason for this is that the sound ‘yi2’ is a polyphonic word, which also corresponds to a neutral-tone syllable ‘咦’ when read with a rising tone to express ‘Surprise’. But with the increasing of absolute values of duration and slope of the additive falling tone, ‘Disgust’ emotion is perceived more often than ‘Surprise’ emotion. Meanwhile, the ‘Angry’ emotion can be perceived with rather high scores.

For final boundary T3 (Fig10.(7)~(9)), the perceived results were affected by the slope and duration as well. ‘Disgust, Sad and Angry’ emotions could be perceived with the increasing absolute values of slope and duration of the falling tone.

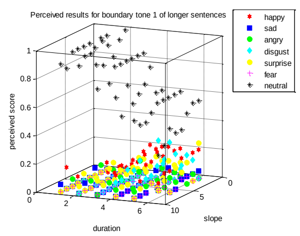
Fig10.(10)~(12) depict the results for final tone 1, 2 and 3 in all utterance conditions respectively. It clearly shows that with the increasing step number of duration and slope of the falling tail, the perceived scores decrease for ‘Neutral’ emotion and increase for other emotions, especially for ‘Disgust, Angry, Fear and Sad’ emotions. Fig10.(13)~(14) shows the overall results and the 2D plot of slope and duration.



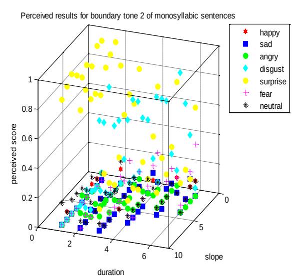
(1)



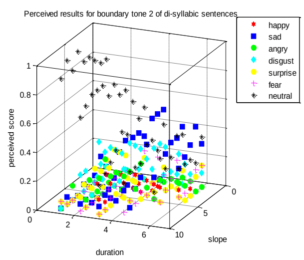
(2)



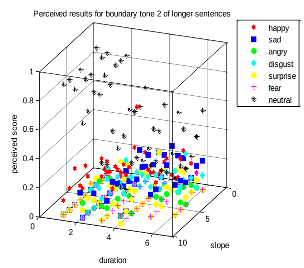
(3)



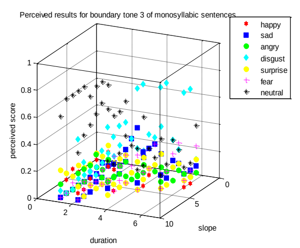
(4)

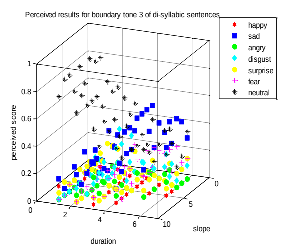


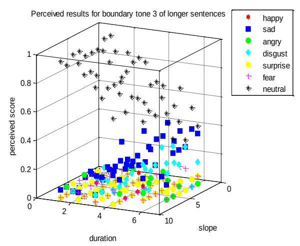
(5)



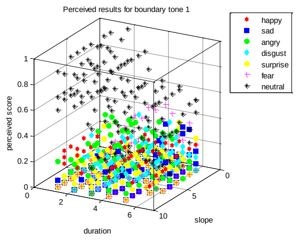
(6)

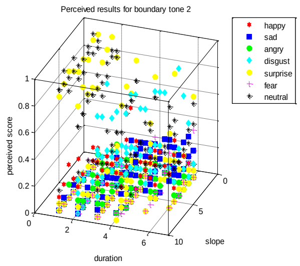


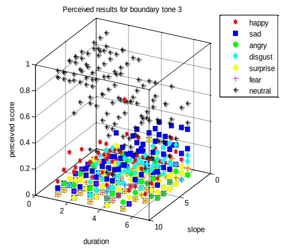




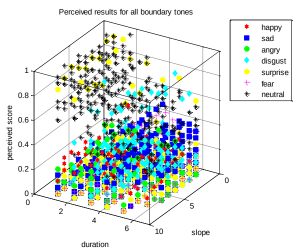
(7) (8) (9)

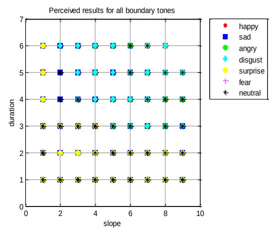






(10) (11) (12)





(13) (14)

Figure 9: *3D plots for average perceived scores of 7 emotions for boundary tone 1~3 in 3 different conditions, and general results for boundary tone1~ 3, and results for all boundary tones. Slope and duration are scaled in step numbers.*

4.3.2 Multifactor GLM analysis

Multifactor GLM analysis was applied to further reveal the factors that affect the perception of emotional attitude. The factors include duration and slope categories of the additive falling tone, final tone category, and utterance type (length category).

The results show that (1) utterance type and the final tone category have significant effect on the emotional perception (*ps=0.0, 0.0*). The slope and duration of the final falling tone also have significant effect on the perceptual scores except for 'Happy' emotion (*ps=0.0, 0.0*). (2) Interaction of 'sentence type and final tone category' is significant (all *ps=0.0*), while the interactions of 'sentence type * slope' and 'final tone category * slope' affect the perception of 'Angry, Disgust and Surprise' emotions (all *ps=0.0*), but have no significant effect on the perception of the other four emotions (*ps=0.549, 0.615, 0.641, 0.44*2). 'Sentence type * duration' and 'final tone category * duration' also have significant interaction for the perception of all emotion types (all *ps=0.0*). (3) F-values revealed that: the additive final slope contributes more to the perception of 'Disgust, Angry and Surprise' emotions than other emotions; the additive final duration contributes more to the perception of 'Neutral, Disgust and Sad' emotions than other emotions; the final tone category contributes more to the perception of 'Surprise, Neutral and Disgust and Sad' emotions than other emotions; and lastly, sentence length contributes more for the perception of 'Neutral, Surprise and Disgust' emotions than other emotions.

4.3.3 Logistic regression analysis

To model the perceptual results, we conducted a logistic regression analysis. The covariations include the duration and slope of the final additive tone, sentence length, and boundary tone category. Dependent variations are the perceptual scores of 7 emotions. Table 3 gives the estimated parameters in logistic regression formulae (1) for each emotion. Then from these P formulae, the estimation surfaces were plotted in Figure 11 for ‘Neutral’ emotion with one of the other 6 emotions by settings ‘boundarytone’= 1,2,3 and ‘stentencetype’=1,2,3 (monosyllabic, bisyllabic and long sentences) respectively.

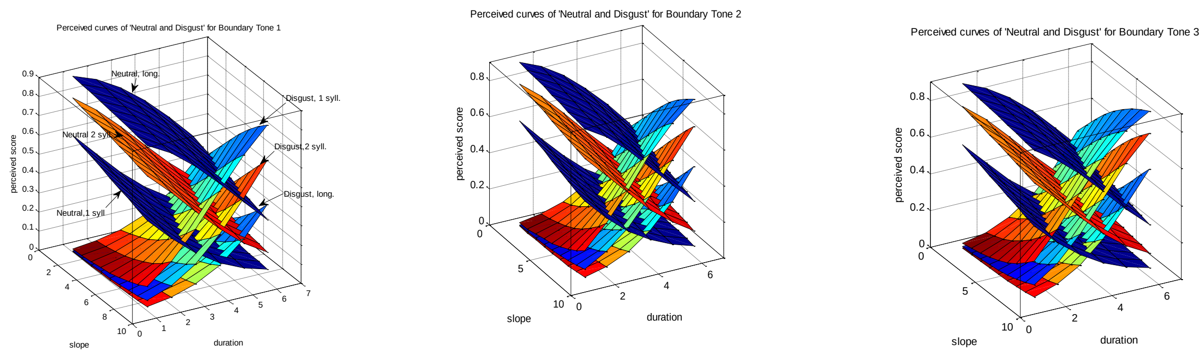
$$P = \frac{e^{b0+b1X1+b2X2+\ldots+bkXk}}{1+e^{b0+b1X1+b2X2+\ldots+bkXk}} \qquad (1)$$

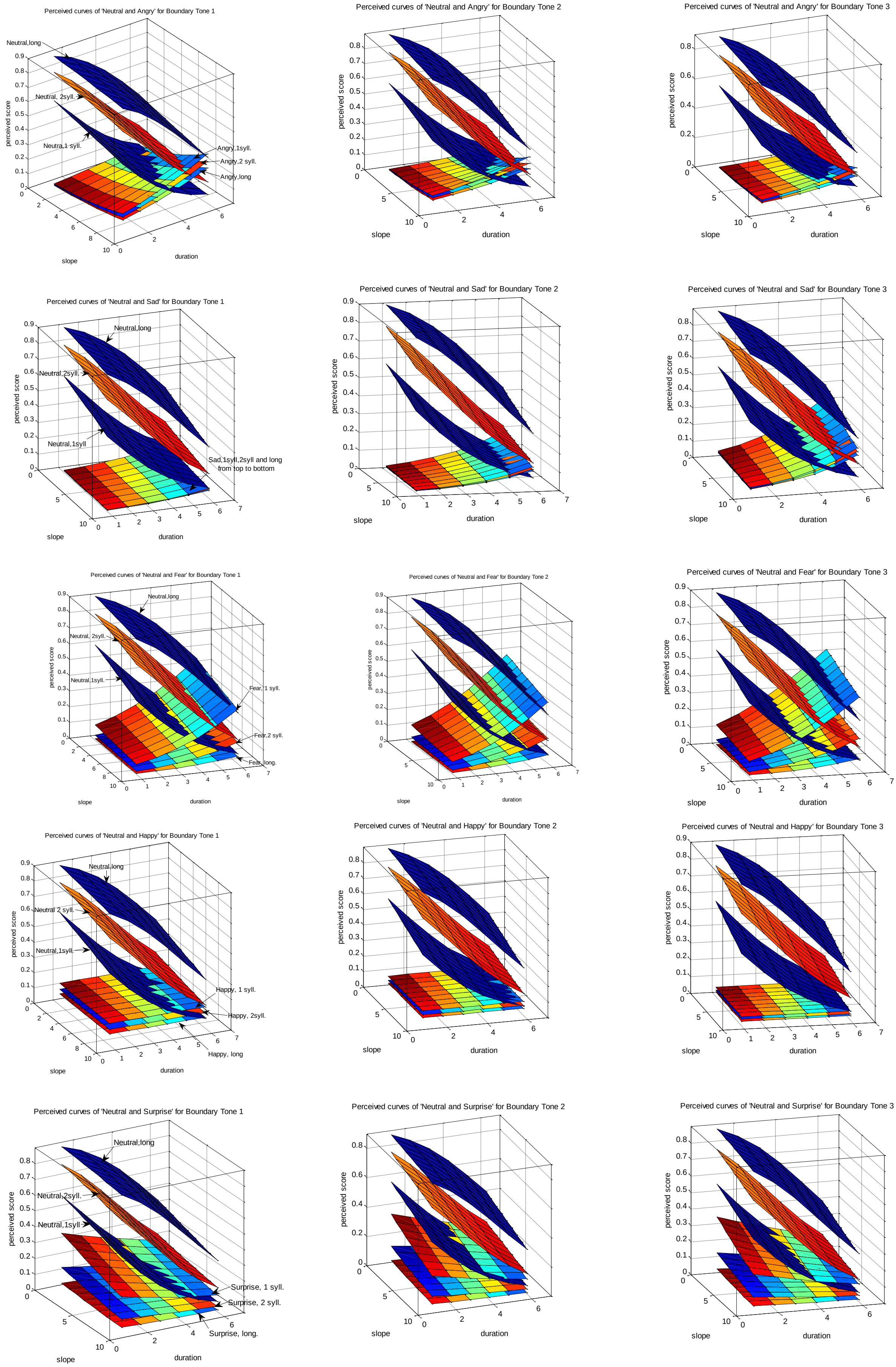
Table 3. *Parameters estimated in logistic regression formulae and the correct class rate for the model*

Emotions	b0+b1X1+b2X2+…bkX4	Correct Class %
Neutral	0.123 − 0.083 ∗ boundarytone − 0.068 ∗ slope − 0.515 ∗ duration + 0.975 ∗ sentencetype	68.8%
Angry	-3.302-0.232*sentencetype-0.424*boundarytone+0.200*slope+0.252*duration	93.6%
Disgust	-4.610-0.943*sentencetype+0.381*boundarytone+0.286*slope+0.681*duration	86.1%
Sad	-5.334-0.129*sentencetype+0.766*boundarytone-0.066*slope+0.421*duration	89.7%
Fear	-1.537-1.203*sentencetype+0.166*boundarytone-0.054*slope+0.402*duration	93.9%
Happy	-2.386+0.374* sentencetype-0.535*boundarytone	92.8%
Surprise	0.913-1.186*sentencetype-0.092*slope-0.174*duration	89.1%

It is shown in Figure 11 that with the increase of duration and slope of the falling tone, the perceived scores of ‘Neutral’ emotion decrease, while the perceived scores of other emotions increase. The longer the sentence is, the higher the perceived score for ‘Neutral’ emotion, and the lower the perceived score for other emotions. We observe similar tendencies for all 6 other emotions, but the tendency is more apparent for Disgust, Angry, Fear and Sad emotions.

The results further demonstrate that changing the acoustic form of the additive boundary tone attached after a ‘Neutral’ intonation will cause the pragmatic expression of the intonation. For the falling tail per se, ‘Disgust, Angry, Fear and Sad’ emotions are most likely to be perceived with the increasing slope and duration of the falling tone. In the same situation, the longer sentences are more likely to keep the ‘Neutral’ emotion than shorter ones. In other words, the boundary tone has less effect on emotional expression for longer sentences than for shorter ones.





Perceived curves of 'Neutral and Angry' for Boundary Tone 1
Neutral,long
Neutral, 2syll.
Neutra,1 syll.
Angry,1syll.
Angry,2 syll.
Angry,long
perceived score
slope
duration
Perceived curves of 'Neutral and Angry' for Boundary Tone 2
perceived score
slope
duration
Perceived curves of 'Neutral and Angry' for Boundary Tone 3
perceived score
slope
duration
Perceived curves of 'Neutral and Sad' for Boundary Tone 1
Neutral,long
Neutral,2syll.
Neutral,1syll
Sad,1syll,2syll and long
from top to bottom
perceived score
slope
duration
Perceived curves of 'Neutral and Sad' for Boundary Tone 2
perceived score
slope
duration
Perceived curves of 'Neutral and Sad' for Boundary Tone 3
perceived score
slope
duration
Perceived curves of 'Neutral and Fear' for Boundary Tone 1
Neutral,long
Neutral, 2syll.
Neutral,1syll.
Fear, 1 syll.
Fear,2 syll.
Fear, long.
perceived score
slope
duration
Perceived curves of 'Neutral and Fear' for Boundary Tone 2
perceived score
slope
duration
Perceived curves of 'Neutral and Fear' for Boundary Tone 3
perceived score
slope
duration
Perceived curves of 'Neutral and Happy' for Boundary Tone 1
Neutral,long
Neutral 2 syll.
Neutral,1syll.
Happy, 1 syll.
Happy, 2syll.
Happy, long
perceived score
slope
duration
Perceived curves of 'Neutral and Happy' for Boundary Tone 2
perceived score
slope
duration
Perceived curves of 'Neutral and Happy' for Boundary Tone 3
perceived score
slope
duration
Perceived curves of 'Neutral and Surprise' for Boundary Tone 1
Neutral,long
Neutral,2syll.
Neutral,1syll
Surprise, 1 syll.
Surprise, 2 syll.
Surprise, long.
perceived score
slope
duration
Perceived curves of 'Neutral and Surprise' for Boundary Tone 2
perceived score
slope
duration
Perceived curves of 'Neutral and Surprise' for Boundary Tone 3
perceived score
slope
duration

Figure.11 *Simulated perception surfaces by using Logistic Regression formulae. Each row plots the surfs of 'Neutral' vs. one of the emotions for 3 sentence length(type) and 3 tone categories (in column).*

5 SUMMARY AND DISCUSSIONS

We analyzed the F0 patterns phonetically and phonologically for the 'successive addition boundary tone' (SUABT) in monosyllabic and disyllabic utterances in order to separate the interaction between the form of Chinese SUABT and its expressive function, and to determine whether it can uniquely encode the emotional or pragmatic information. We also examined longer utterances.

We proposed that the SUABT is composed of two components, the lexical tone to express linguist meaning, and the expressive tone to express emotional attitude or pragmatic meaning.

Accordingly, emotional intonation can be represented phonologically by F0 range, F0 register and SUABT as shown in Table 1. 'Neutral and Fear' intonations have both 'M' range and 'M' register and level boundary 'L-%', but 'Fear' intonation has quiver voice. 'Sad' intonation has both 'L' range and 'L' register and 'L-%' boundary tone. 'Happy and Surprise' intonations have both 'H' range and 'H' register and 'H-r%' boundary tone. 'Disgust' intonation has 'M' range and 'L' register and L-f% boundary tone. 'Angry' intonation has both 'H' range and 'H' register and 'H-f%' boundary tone.

Based on the acoustic analysis, a perceptual experiment was conducted by elaborately manipulating the boundary tones of 9 emotionally neutral sentences with a 'successive addition falling tone' attached to the end.

After making a multi-factor GLM analysis and a logistic regression analysis on the perceptual experiment results, we found that the final falling additive tone could change the emotional expression of intonation. However, the perceived emotions vary with the change of the slope and duration of the added final fall. Except for monosyllabic sentences of 'Disgust' emotion, none of the scores can go higher than the scores for 'Neutral' intonation, besides, many stimuli have more than one perceived results (Li, Fang and Dang 2011),suggesting that successive addition boundary tone can be partly encoded to express emotional attitude, but the mapping from the form of the successive addition boundary tone to the pragmatic function is a one-to-many relation. In other words, the successive addition boundary tone can't be encoded independently to express a specific emotion. On the other hand, from the simulated surface of perceived emotions in Figure 11, we know that one could perceive some common emotion produced with different acoustic features of SUABT. So the relation between the acoustic form of the SUABT and the pragmatic function of it is 'many to many'.

The factors affecting the perceptual score of 7 emotions include the slope and the duration of the final successive fall, the sentence type and the final tone category. However, each factor makes different contributions to different emotions. The final tone factor could extend Yuan's tone-dependent mechanism to pragmatic aspect.

Shorter sentences are affected more by the final successive addition tone than longer sentences. A reasonable interpretation is that longer sentences could employ richer components to express both linguistic and pragmatic formation than shorter ones, especially monosyllabic sentences. The components may include emotional focus, VQ (vice quality), speech rate, the boundary tone category and the intensity, etc. This reveals that both boundary tone and other components will be encoded in parallel to express linguistic and pragmatic information, as suggested by Xu's PENTA model (2005).

The study indicated that SUABT has the function to convey emotional information. Another question may be raised to ask whether SUABT could convey other emotions or attitudes. The answer is definitely yes. See the intonation contour in Figure 12, clipped from our children language acquisition corpus. It is a bisyllabic utterance, spoken by a 5 year-old girl. In this example, the SUABT is employed to convey the imperative attitude.

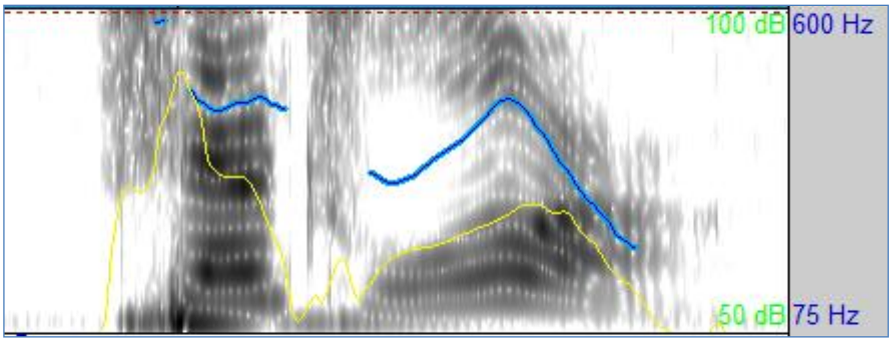


Figure 12: *The intonation (blue line) of an imperative sentence 'Delete it!'(删除!shan1 chu2!) uttered by a 5 year-old girl. The boundary tone is T2 with a falling SUABT.*

What we must emphasize here is that the emotional intonations observed in this study were all produced in isolated utterances, and only two speakers' data were analyzed. However, people communicate in spontaneous speech rather than isolated utterances. The intonation pattern has to be investigated at the discourse level in a framework such as HPG (Hierarchical Prosodic Phrase Grouping) proposed by Tseng (2008): 'all the discourse prosody is higher level specification as well as cross-phrase association in addition to discrete intonation pattern'.

6 ACKNOWLEDGEMENTS

This research was funded by JSPS Ronpaku Program and NSFC Project with No. 60975081 and CASS innovation project.

7 REFERENCES

[1] Callier Patrick. 2011. On the Edge: The Socio-phonetics of Boundary Tones and Final Lengthening in Mandarin Chinese. *eVox* 5.1:16-36.

[2] Chao, Yuan Ren. 1932. A preliminary study of English intonation (with American variations) and its Chinese equivalents. BIHP, 105-156. (The Ts'ai Yuan Pei Anniversary Volume).

[3] Chao, Yuan Ren. 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology* 4:121-134.

[4] Chao, Yuan Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

[5] Chao, Yuan Ren. 1980. A system of tone letters. *Fangyan* 2.

[6] Carlos Gussenhoven. 2002. Intonation and interpretation: Phonetics and Phonology. Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody. Aix-en-Provence, ProSig and Universit′ e de Provence Laboratoire Parole et Language. 47-57.

[7] Dwight Bolinger. 1989. *Intonation and Its Uses*. Palo Alto, CA: Stanford University Press.

[8] Jia, Yuan. 2006. Mechanisms of Question Intonation in Mandarin. In Chinese Spoken Language Processing, ed. by Qiang Huo (eds), 19-30. Berlin: Springer.

[9] Jiang, Ping, and Aishu Chen. 2011. Representation of Mandarin intonations: boundary tone revisited. Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23), ed. by Zhuo Jing-Schmidt. 97-109. University of Oregon, Eugene.

[10] Ladd, D Robert. 1996. *Intonational phonology*. Cambridge, UK: Cambridge University Press.

[11] Li, Aijun, and Haibo Wang. 2004. Friendly Speech Analysis and Perception in Standard Chinese. In Prosodic of ICSLP2004, JEJU, Korea.

[12] Li, Aijun. 2008. *Stress Pattern of Emotional Speech. Chinese Journal of Phonetics*. Beijing: Commercial Publishing House.

[13] Li, Aijun, Qiang Fang, Yuan Jia, and Jianwu Dang. 2012. Successive Addition Boundary Tone in Chinese Disgust Intonation. NACCL24, USA.

[14] Li, Aijun, Qiang Fang, and Jianwu Dang. 2011. Emotional Intonation in a Tone Language: Experimental Evidence From Chinese. ICPhS XVII, Hong Kong, 17-21.

[15] Lin, Maocan, and Zhiqiang Li. 2011. Focus and Boundary in Chinese Intonation. ICPhS XVII, Hong Kong, 17-21.

[16] Lin, Maocan. 2004. Boundary Tone of Chinese Intonation and Its Pitch Pattern, In From Traditional Phonetics to Modern Speech Processing, ed. by Gunnar Fant, Hiroya Fujisaki, Jianfen Cao, and Yi Xu, 309-327. Beijing: Foreign Language Teaching and Research Press.

[17] Liu, Fang, and Yi Xu. 2005. Parallel encoding of focus and interrogative meaning in Mandarin Chinese. *Phonetic*a 62, 70-87.

[18] Lu, Jianming, and Maocan Lin. 2009. Raising tail of the question vs. the modal particle. In Frontiers in Phonetics and Speech Science, ed. by Gunnar Fant, Hiroya Fujisaki, Jiaxuan Shen.(eds.) Beijing: Commercial Publisher Press.

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

[20] Patrick Callier. 2011. On the Edge: The Sociophonetics of Boundary Tones and Final Lengthening in Mandarin Chinese. *eVox* 5.1.

[21] Pierrehumbert, Janet Breckenridge. 1980. The Phonology and Phonetics of English Intonation. MIT dissertation.

[22] Pierrehumbert Janet, and Julia Hirschberg. 1990. The Meaning of Intonational contours in the Interpretation of Discourse. in P.

[23] Tseng, Chiu-yu. 2006. Higher Level Organization and Discourse Prosody. *The Second International Symposium on Tonal Aspects of Languages* 23-34.

[24] Tseng, Chiu-yu. 2008. Courpus Phonetic Investigations of Discourse Prososdy and Higher Level Information. *Language and Linguistics* 9.3:659-719.

[25] Varod Silber-Varod. 2011. Dependencies Over Prosodic Boundary Tones in Spontaneous Spoken Hebrew. Proceedings of Depling 2011(International Conf. on Dependency Linguistics), Barcelona.

[26] Vered Silber-Varod. 2011. Dependencies over prosodic boundary tones in Spontaneous Spoken Hebrew. Proceedings of Depling 2011(International Conf. on Dependency Linguistics), Barcelona.

[27] Wang, Haibo, Aijun Li and Qiang Fang. 2005. F0 contour of Prosodic Word in Happy

Speech of Mandarin. In Prosodic of ACII 2005, Beijing.

[28] Wu, Zong Ji. 1990. Basic contour patterns of intonation of Standard Chinese. *Essays in Honor of Professor Wang Li*. Beijing: The Commercial Press.

[29] Wu, Zong Ji. 1995. Predictability of different attitudinal intonation in Standard Chinese. *Proc. of the 13th ICPhS,* Stockholm, Sweden.

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

[31] Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251.

[32] Xu, Yi, Andrew Kelly, and Cameron Smillie. 2011. Emotional expressions as communicative signals. Under review for S. Hancil and D. Hirst (eds.) Prosody and Iconicity.

[33] Xu, Yi, and Suthathip Chuenwattanapranithi. 2007. Perceiving anger and joy through the size code, ICPhS2007, Saarbrücken, 6-10 August 2007 VI.

[This paper was published on ICPhS 2011, Hong Kong]