Emotional Expressiveness of Successive Addition Boundary Tone in Mandarin Chinese

Aijun Li[1], *Qiang Fang*[1] , *Jianwu Dang*[2,3]

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

Abstract

The aim of the present study is to explore the relationship between the form of Chinese 'successive addition boundary tone' [1,2] and its expressive function, and whether 'SUccessive Addition Boundary Tones' (SUABT) can uniquely encode pragmatic information. To this end, a perceptual experiment was conducted by attaching a successive falling tone to the F0 curve of the neutrally expressed utterance, which varies in both length and final tone categories. The slope and duration of the successive falling tones are varied and synthesized to produce 425 stimuli. And the emotional expression of stimuli was evaluated by 20 subjects. Through GLM (General Linear Model) analysis, it is found that (1) Slope and duration of successive falling tone, length of sentence, and final tone category affect emotional expressiveness significantly, but their contribution varies across emotions. (2) When the slope (absolute value) and duration of the successive falling tone increased, the perceptual score of neutral emotion decreases, whilst the perceptual score of other 6 basic emotions increases (significantly for 'disgust, angry and fear'). It indicates that there is not a unique mapping between the form of the successive addition boundary tone and its pragmatic function, rather a 1-to-many mapping. The results also demonstrated that the SUABT alone is not enough to encode emotional information. (3) The perceptual score is affected by the length of the utterance, the longer the sentence, the smaller the effect, which further revealed that the SUABT alone is not enough to express emotions, as suggested by Xu'PANTA model[3], other components such as focus, voice quality, speech rate, energy, all could be used to encode express pragmatic information.

Index Terms: successive addition boundary tone, emotion, expression, Chinese intonation

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 [4,5]. The intonation is also considered to eventually link communicative meanings or paralinguistic or pragmatic meaning [4,6,7].

Gussenhoven [8] proposed a relatively more elaborate theory that bases intonational meaning on a set of "biological codes": frequency, effort, and production.

Liu & Xu [9] argue 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 [5] as a phonological unit for representing the internal makeup of an intonation phrase. Xu [10] 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 [7] gave a list of functions of pitch accents and boundary tones with reference to discourse structures. Vered Silber-Varod [10] 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[1] remarked early on that boundary tones showed specified pitch movements that seemed to be linked to sentence pragmatics. He 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 successively to the end of the lexical tones. 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, 11].

Lin [12, 13] investigated how 'simultaneous addition' is realized in the sentence-final syllables to convey intonational meanings. Yuan [14] proposed 3 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 [15] support that Mandarin interrogative cues distribute over an entire utterance. High pitch in both edges of the intonation phrase should be marked out.

Employing a corpus of conversational data in Mandarin, Patrick Callier [16] investigated intonational meaning and function of the boundary tone in a sociophonetic way, checking the H% or L% Boundary tone by clause type, illocutionary force, and speaker gender.

Seldom researches have been conducted for successive addition relating to pragmatic meaning of intonation. Mueller-Liu [17] listed some successive addition tones in expressive speech to signal the emotion-attitudinal message, and Lu and Lin [18] also found it in the intonational questions to signal the interrogative mood.

In our previous study [2] we found that the successive addition boundary tones are employed by the speakers to convey expressive information, such as to express 'Disgust and Angry' by a kind of 'Falling' successive addition tone, and 'Happy or Surprise' by a kind of 'Rising 'successive addition tone, as pointed

out by Chao. As shown in Fig.1, the boundary tones for 'Disgust and Angry' exhibit a falling tail compared to that for neutral, keeping the first part as its lexical tone.

In the present study, we will further tear apart the interaction between the pragmatic function of the SUABT and its acoustic form through a perceptual experiment. Moreover, we will explore whether the SUABT can independently encode emotional attitude and whether the mapping between the acoustic form and the expressive function is unique?

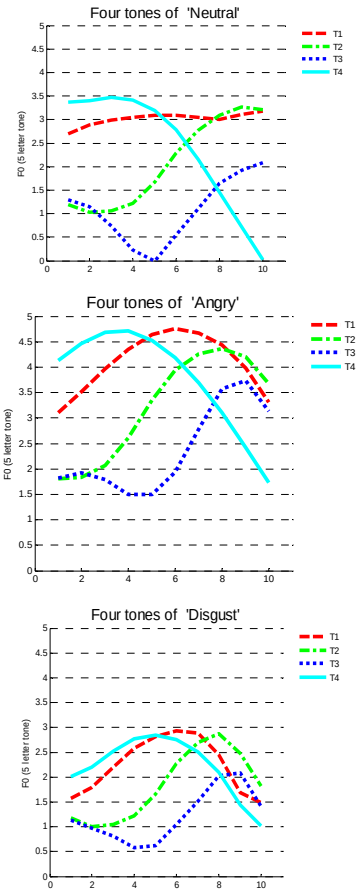


Figure 1: *A male speaker's 'Neutral, Angry and Disgust' intonation of monosyllabic utterances with four lexical tones(T1~T4), normalized in 5 tone letter scale.*

2. Stimuli preparing

Here, we wish to explore the 'gradient nature'[4] that the additive boundary tone contributes to emotional expression. In order to exclude the potential expressive factors except the SUABT, emotionally neutral utterances are used as the baseline data and an addition tone is successively added to final boundary tone of the neutral sentences, by varying the acoustic parameters of the additive tone.

Table 1 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 [2]. Because it is not easy to tell apart the additive falling boundary tone from the original lexical tone, in order to clearly and easily manipulate the lexical and the additive parts respectively, tone 4 was excluded here.

Figure 2 depicts a schematized representation of final tone 2 with a successive falling addition tone; the rising part represents the lexical tone while the falling part represents the SUABT. The acoustic features of the male speaker's disgust boundary tone were analyzed according to this structure [2]. The features that we concerned 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 slop 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 is less than 65st/s and ratio less than 1.4 within the 95% Confidence Interval.

Table 1: *Nine sentences used in the perceptual experiment*

#	Content	Length (syll.s)	Final syllable	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。Mr. Mr. Zhou bought 2.5kg of sea cucumbers.	8	shen1	T1
8	去年盖的二层小楼。Two-floor flat was built last year.	8	lou2	T2
9	长篇小说梅娘曲。long novel MeiNiang Qu	7	qu3	T3

When generating the stimuli, we added a falling tail to the neutral boundary tones of the utterances in Table 1. So the new boundary tone consisted of two parts, the first part was the original neutral boundary tone, 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 9 neutral sentences. As the lowest F0 is set to be 75Hz in this process, hence, 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 have falling tone endings, the pitch registers of them are different. The pitch register of Disgust are more closed to Neutral than Angry, so we hypothesize that the synthesized stimuli would cause a greater tendency to express the Disgust emotions.

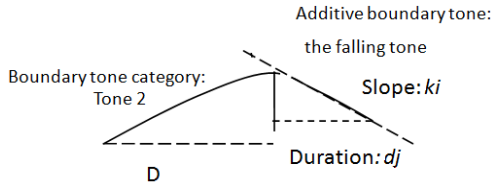


Figure 2: *A schematic representation of additive falling tone with duration dj and slope ki.*

3. 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 left group). Before start 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. The perceived score is set to 1 for selected emotion; else the score is set to 0.

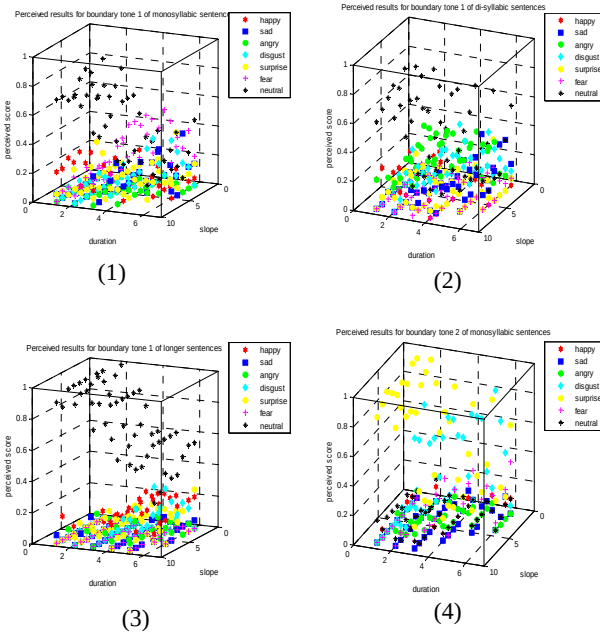
4. Analysis on the perceptual results

4.1. Perceptual results

Figure 3 shows the average perceptual results on 7 emotions of 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 tone 1 in Fig.3.1~3, 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, while increase gradually for other emotions. Unexpectedly, not only the scores for Disgust emotion but also the scores of Angry and Fear emotion increase apparently.

For final tone 2 (Fig3.4~6), the results for monosyllabic utterances are ‘misunderstood’. It’s 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 modal word ‘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 than Surprise emotion. Meanwhile, the ‘Angry’ emotion can be perceived with rather high scores.



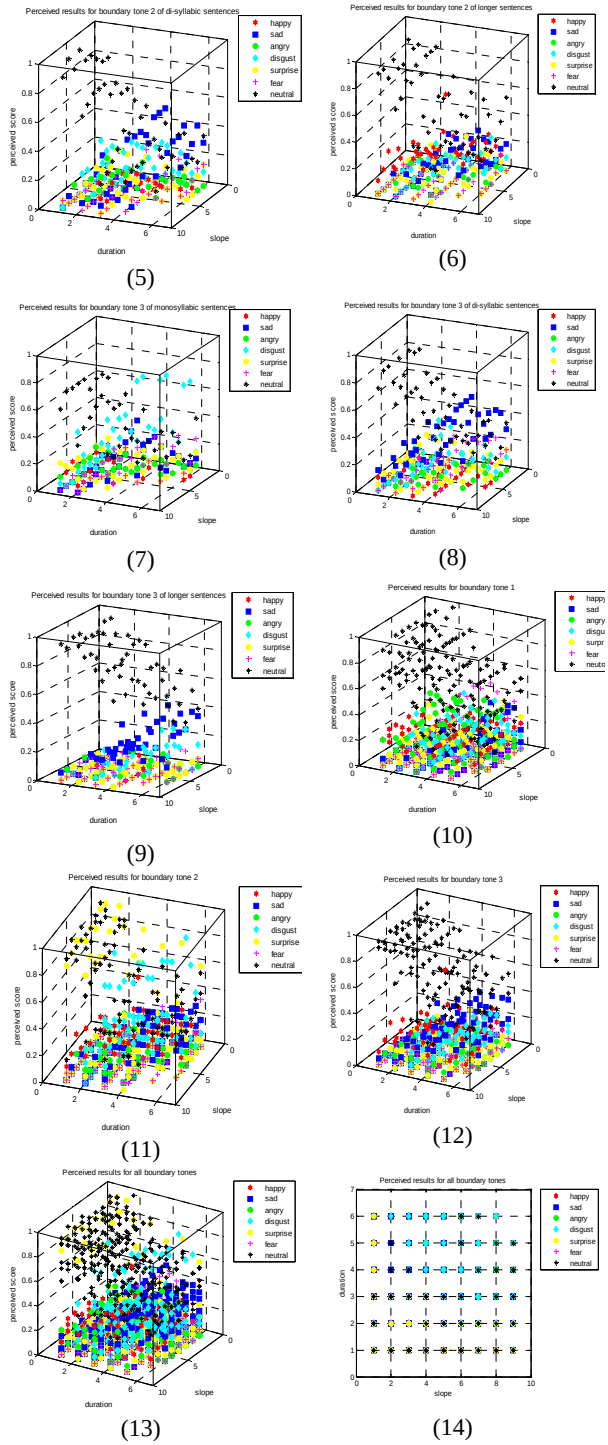


Figure 3: *3D plots for average perceived scores of 7 emotions for boundary tone 1~3 in 3 different conditions, and general results for boundary tone1~ 3, and results for all boundary tones. Slope and duration are scaled in step number.*

For final tone 3 (Fig3.7~9), 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.

Fig3.10~12 depict the results for final tone 1, 2 and 3 in all utterances conditions respectively. It clearly shows that with the increasing step number of duration and slope of the falling tail, the perceived scores for neutral emotion decrease while increase for other emotions, especially for ‘Disgust, Angry, Fear and Sad’. Fig3.13~14 shows the overall results and its 2D plot of slope and duration.

4.2. 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 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 (p=0.0, 0.0). The slope and duration of the final falling tone also have significant effect on the perceptual score except for 'Happy' emotion. (p=0.0, 0.0) (2) Interaction of 'Sentence type and final tone category' is significant (all p=0.0), while 'sentence type *slope' and 'final tone category * slope' interactively affect the perception of 'Angry, Disgust and Surprise'(all p=0.0), but have no significant effect on the perception of the other four emotions (P=0.549, 0.615, 0.641, 0.442) . 'sentence type * duration' and 'final tone category * duration' also have significant interaction for all emotion perception (all p=0.0) (3) F-values revealed that: the additive final slope contributes more to the perception of 'Disgust, Angry and Surprise' than for those of other emotions; the additive final duration contributes more to the perception of 'Neutral, Disgust and Sad' than for those of other emotions; the final tone category contributes more to the perception of 'Surprise, Neutral and Disgust and Sad' than for those of other emotions; sentence length contributes more for the perception of 'Neutral, Surprise and Disgust' than for those of other emotions.

5. Discussion and Conclusion

Based on the emotional intonations, a perceptual experiment was conducted by elaborately manipulating the boundary tones of 9 emotionally neutral utterances attached with a 'successive addition falling tone'.

Apart from the 7 emotional expressions evaluated, 92 other attitudes were perceived by the subjects as well. 26 of them are 'uncertainty', others including 'Disgust-like 'attitudes as 'contempt, dislike, impatient, blame, query, blame', and more others as 'self-contented, jealous, suspect, emphasis, sneer, affectionate, inquiring'.

After making a multi-factor GLM analysis on the seven emotions, we found that the final falling tone could change the emotional expression of intonation. However, the perceived emotions vary with the change of the slope and duration of the addictive final fall. Except for monosyllabic utterances of Disgust emotion, none of the score can excesses the score for Neutral intonation. This suggests that successive addition boundary tone can be encoded to express emotional attitude, but the mapping from the form of the successive addition boundary tone to the pragmatic function is not unique, and the successive addition boundary tone can't be encoded independently to express a specific emotion.

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 has different contribution to different emotions. The final tone factor could extend Yuan's tone-dependent mechanism to pragmatic aspect.

The final tone has more dominate effects on the intonation of shorter utterances than longer. A reasonable interpretation is that longer sentences could employ more components to express both linguistic and pragmatic information than shorter ones. The components may include emotional focus, Voice Quality, speech rate, the final tone category and the intensity, etc. This indicates 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, even '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'[19].

6. Acknowledgements

Thank for the discussion and good suggestions from Dr. Yi Xu. This research was funded by JSPS Ronpaku Program and NSFC Project with No. 60975081 and CASS key project "Speech production: theory and research methodology".

7. References

[1] Chao, Y., "A Preliminary Study of English Intonation (with American Variations) and Its Chinese Equivalents", BIHP, The Ts'ai Yuan Pei Anniversary Volume, 105-156, 1932.

[2] Li,A., Fang,Q. and Dang J., "Emotional Intonation in a Tone Language: Experimental Evidence From Chinese", ICPhS XVII, Hong Kong, 17-21, August 2011.

[3] Xu, Y., "Speech Melody as Articulatorily Implemented Communicative Functions", Speech Communication, 46: 220-251, 2005.

[4] Ladd D. R., "Intonational Phonology", Cambridge, UK: Cambridge University Press, 1996.

[5] Pierrehumbert, J., "The Phonology and Phonetics of English Intonation", MIT dissertation, 1980.

[6] Dwight B., "Intonation and Its Uses", Palo Alto, CA: Stanford University Press, 1989.

[7] Pierrehumbert, J. and J. Hirschberg, "The Meaning of Intonational Contours in the Interpretation of Discourse", 1990.

[8] Gussenhoven, C., "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, 2002.

[9] Liu, F. and Xu Y., "Parallel Encoding of Focus and Interrogative Meaning in Mandarin Chinese", Phonetica, 62, 70-87, 2005.

[10] Varod S. V., "Dependencies Over Prosodic Boundary Tones in Spontaneous Spoken Hebrew", Proceedings of Depling 2011(International Conf. on Dependency Linguistics), Barcelona, 2011.

[11] Chao Y. R, "A Grammar of Spoken Chinese", Berkeley, CA: University Of California Press, 1968.

[12] Lin, M., "Boundary Tone of Chinese Intonation and Its Pitch Pattern", In Fant, G., Fujisaki, H., Cao, J. and Xu, Y. [Ed], From Traditional Phonetics to Modern Speech Processing, Beijing: Foreign Language Teaching and Research Press, 309-327, 2004.

[13] Lin M., Li,Z, "Focus and Boundary in Chinese Intonation", ICPhS XVII, Hong Kong, 17-21, August 2011.

[14] Yuan J. H., "Mechanisms of Question Intonation in Mandarin", Chinese Spoken Language Processing, Huo. Q., et al. [Ed], 19-30, Berlin: Springer, 2006.

[15] Jiang P. and Chen A. S., "Representation of Mandarin Intonations: Boundary Tone Revisited", Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23), 2011, Volume 1, Zhuo Jing-Schmidt [Ed], University of Oregon, Eugene, 97-109, 2011.

[16] Callier P., "On the Edge: The Socio-phonetics of Boundary Tones and Final Lengthening in Mandarin Chinese", eVox., Vol.5.1, Washington, DC: Georgetown University, January 2011.

[17] Mueller-Liu, P., "Signaling Affect in Mandarin Chinese - the Role of Utterance-final Non-lexical Edge Tones", Proc. 5th of Speech Prosody, PS6-3-0048, 2006.

[18] Lu, J. and Lin, M., "Raising Tail of the Question vs. the Modal particle", In Fant G., Fujisaki H. and Shen J. [Ed], Frontiers in Phonetics and Speech Science, Commercial Publisher, 2009.

[19] Xu, Y., Kelly, A. and Smillie, C. Emotional expressions as communicative signals. Under review for S. Hancil and D. Hirst (eds.) *Prosody and Iconicity*, 2011.