Successive Addition Boundary Tone in Chinese Disgust Intonation

Aijun LI, Qiang FANG, Yuan JIA
Institute of linguistics Chinese Academy of Social Sciences, Beijing

Jianwu DANG
Tianjin University, Tianjian; Japan Advanced Institute of Science and Technology
Background

- two elements, the intonation of tonal languages like Chinese and the lexical tone element, are expressed by the same F0 curve.
- Since the F0 curve conveys both the linguistic and paralinguistic information, to understand the encoding mechanism of expressive speech in speech communication, the interplay between tone and intonation in Chinese emotional speech has to be examined.
‘...the most obvious fact we have to recognize is that the actual melody or pitch movement of a tonal language is a different affair from the mere succession of the few fixed tone of that language. It is in fact a resultant of three elements. The characteristic tones of the individual syllable-words, their influence on each other in connected speech, and the movement of pitch, which indicates the mood or attitude of the speaker. The first is usually called tone, or etymological tone, the second I shall call neutral intonation, and the third expressive intonation, the latter two together forming sentence intonation.’
Addition of Tone and Intonation (Chao)

‘here are at least two types of tonal additions, simultaneous addition and successive addition, and in our consideration of the forms of intonation, we have not only to ascertain what they are, but also how they are applied to the tonal systems of the languages.’
Simultaneous addition

这个东西↗坏，那个东西↘好。
(That thing is bad ↗, this thing is good ↘.)

These tones are in fact the algebraic sums or resultants of two factors.
The original word-tone and the sentence intonation proper, in this case a purely logical intonation.
Successive addition

• If, however, we have an example like the following,
• (Do you say) this is↗ bad? It is (decidedly)↘good!

这个坏？(a suspense-clause) 这个好！(conclusion)

• This first clause will also have a rising intonation and the second a falling intonation in Chinese, but the fall and rise will not be added simultaneously to the last syllables, but will be joined on successively, after the word-tone are completed, thus:

这个↘坏？这个↗好！
Successive addition

The rising endings have the following effect (5-tone scale):

\[ \begin{align*}
&\uparrow \ 55: = 56: \quad \uparrow \\
&\uparrow \ 35: = 36: \quad \uparrow \swarrow \\
&\uparrow \ 214: = 216: \quad \swarrow \\
&\uparrow \ 51: = 513: \quad \searrow
\end{align*} \]

The falling endings have the following effect:

\[ \begin{align*}
&\downarrow \ 55: = 551: \quad \swarrow \\
&\downarrow \ 35: = 351: \quad \nwarrow \\
&\downarrow \ 214: = 2141: \quad \nwarrow \\
&\downarrow \ 51: = 5121: \quad \nwarrow
\end{align*} \]
Our previous findings on emotional boundary tones
Emotional intonations of one syllabic utterances:

Female speaker (Left)

Male speaker (Right)

Four tones of 'Neutral'

Four tones of 'Surprise'
Female speaker(Left) Male speaker(Right)

Four tones of ‘Angry’

Four tones of ‘Happy’
Female speaker (Left)/Male speaker (Right)

Four tones of 'Sad'

Four tones of 'Fear'

Female speaker (Left)/Male speaker (Right)
Four tones of 'Disgust'

F0 (5 letter tone)

T1  T2  T3  T4
Table 1: Features of tonal space and successive addition tones

<table>
<thead>
<tr>
<th>Emotions</th>
<th>span</th>
<th>register</th>
<th>range</th>
<th>add. Tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sad</td>
<td>1-3 (LM)</td>
<td>L</td>
<td>3 (L)</td>
<td>L-r/le%</td>
</tr>
<tr>
<td></td>
<td>1-4 (LM) for</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>1-5 (LH)</td>
<td>H</td>
<td>5 (H)</td>
<td>H-r%</td>
</tr>
<tr>
<td>Angry</td>
<td>2-5 (LH)</td>
<td>H</td>
<td>4 (M)</td>
<td>H-f%</td>
</tr>
<tr>
<td></td>
<td>2-4 (LH) for F</td>
<td>M (for F)</td>
<td>3 (L for F)</td>
<td></td>
</tr>
<tr>
<td>Fear</td>
<td>1-4 (LH)</td>
<td>M</td>
<td>4 (M)</td>
<td>L-le%</td>
</tr>
<tr>
<td>Surprise</td>
<td>1-5 (LH)</td>
<td>H</td>
<td>5 (H)</td>
<td>H-r%</td>
</tr>
<tr>
<td>Disgust</td>
<td>1-3 (LM)</td>
<td>L</td>
<td>3 (L)</td>
<td>L-f%</td>
</tr>
<tr>
<td>Neutral</td>
<td>1-4 (LH)</td>
<td>M</td>
<td>4 (M)</td>
<td>L-%</td>
</tr>
</tbody>
</table>

‘r, f and le’ are used to describe the ‘Rising’, ‘Falling’ or ‘Level’ successive addition tones, and ‘-’ for no successive addition tone.
This research focuses on:

- ‘Disgust’ intonations: whether ‘Falling addition boundary tones’ can be observed for longer utterances.
- If so, the acoustic features of the disgust boundary tones, including the durations of the lexical tone and the additive tone, and the falling slope of the additive tone.
Materials: EmotionCASS

- A set of 111 sentences with varying length (from 1 to 14 syllables), sentence types (narrative or interrogative), and structures.
- All of these sentences are emotionally neutral.
- The monosyllabic sentences cover 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: i.e., Disgust(D), Sad(S), Angry(A), Happy(H), Surprise(SU), Fear(F) and Neutral(N).
Materials and recording

- All the F0 and duration data are extracted by Praat and manually checked.
- F0 normalized into 10 points (semitone values) and then transferred into 5 tone-letter space.

Sad  Happy  Fear  Surprise  Angry  Disgust  Neutral
Disgust intonations of disyllabic utterances with boundary tones in four tonal categories

Female speaker:

F0 of bisyllabic words (Neutral)

F0 of bisyllabic words (Disgust)

Neutral

Disgust
‘Disgust’ intonations of disyllabic utterances

Male speaker:

Neutral

Disgust
• For the ‘Neutral’ intonation, the F0 of boundary tones keep identical with their citation forms, while for the ‘Disgust’ intonations, for those boundary syllables with tone1, tone2 and tone 3, they show an additive final tailing, being added after their lexical tone;

• examining the intonations with tonal combinations of HH+HH / HH+HL /HL+HH / HL+HL (tone 1+ tone 1 / tone1+tone4 / tone 4+tone1 / tone4+tone4), we found that the ‘Neutral’ intonations have an obvious tendency of declination, while the corresponding ‘Disgust’ intonations exhibit an opposite patterns. It is caused by the lowering and compressing of the F0 on the first syllable which indicates that the final syllables could be more prominent than the first syllables in the expression of ‘Disgust’ emotion;

• ‘L-f%’ and ‘L%’ could be employed to describe the ‘Disgust and Neutral’ boundary tones as for monosyllabic constituents;

• The final addition of the falling boundary tone accounts for a considerable lengthening of the whole boundary tone.
Longer utterances: 10 syllables

Intonation pattern of the sentence /fa1 yi2 feng4 chuan2 zhen1 qu4 zhong1 guo2 bei3 jing1/ (Send a fax to Beijing, China.) with a high level boundary tone in seven emotions, the black dotted line is for the ‘Disgust’ emotion (Male speaker), light blue line for ‘Neutral’ emotion.
Longer utterances: 5 syllables

Intonation pattern of sentence /hua2 xue3 chang3 jiao4 lian4/ (a skiing coach in ski field.) with a boundary tone 4, dark line is used to express ‘Disgust’ and light blue is for ‘Neutral’, from male speaker. **Successive addition**
Longer utterances: 5 syllable interrogation

Interrogative intonation pattern of sentence /jie2 shu4 le ma?/ (Has it finished?) with an unstressed boundary tone (neutral tone), dark line for ‘Disgust’ and light blue line for ‘Neutral’, from male speaker.
• For those boundary tones with tone4(HL) and neutral/unstressed tones, the ‘Disgust’ falling boundary tone keep falling as its ‘Neutral’ counterparts, however, it also shows a slightly ‘peak delay’, even in the interrogative sentence in Figure 5.

• The ‘Neutral tone’ boundary has a rising tone, while the ‘Disgust’ intonation still has a falling boundary tone.
Acoustic analyses on the ‘Disgust’ boundary tone

- As shown in previous figures, the ‘Disgust’ boundary tones of T1~T3 (short for Tone1~Tone3) comprise two components, i.e., the lexical tone which is employed to express lexical meaning, and the expressive tone which is realized as a falling tone to express disgust emotion.

![Schematic representation of an additive falling tone with duration ‘d’ and slope ‘k’.

```latex
\textbf{Boundary tone category: T2}
\begin{align*}
\text{Additive boundary tone: a falling tone} \\
\text{Slope: } k \\
\text{Duration: } d
\end{align*}
```
Disgust utterances from two speakers

- acoustic measurements on the two components for all the utterances with final boundary tones in tone 1, tone 2 and tone 3 (leaving tone 4 and neutral tones, and integrative utterances for future discussion).

<table>
<thead>
<tr>
<th></th>
<th>Tone1</th>
<th>Tone 2</th>
<th>Tone 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>22</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Male</td>
<td>21</td>
<td>19</td>
<td>17</td>
</tr>
</tbody>
</table>
Duration of two components

Mean durations: ‘D, d and D+d’ of the final boundary tone for three tonal categories of two speakers.
• For male speaker, ANOVA analysis is conducted that ‘D’ and ‘d’ have significant difference across three tones (p=0.0 and 0.003 respectively), while the total length of Tone1~Tone3 have no significant difference (p=0.54).

• However, all these three parameters have significant difference for female speaker (p=0.034, 0.00, 0.005, respectively).
Mean duration ratios: ‘d/D’ of the final boundary tone for three tonal categories of two speakers.
• the lexical component and the expressive component for boundary T1 are almost in equal length, while the emotional components shows shorter duration in T2 and T3, for male speaker: \( \frac{d}{D} \approx 40\% \) and 36\% for T2 and T3 respectively, and for female speaker: \( \frac{d}{D} \approx 58\% \) and 44\% for T2 and T3 respectively. For both speakers, \( \frac{d}{D} \) of T2 and T3 are clustered in one group while T1 in another group.( alpha = 0.05).

• For male speaker, the minimum ratio (‘\( \frac{d}{D} \)’) is 0.15 and maximum ratio is 3.6, the 95\% Confidence Interval is from 0.25 to 1.38. For female speaker, the minimum ratio (‘\( \frac{d}{D} \)’) is 0.15 and maximum ratio is 3.59, the 95\% Confidence Interval is from 0.20 to 1.33. Therefore, both speakers have consistent change space for ‘\( \frac{d}{D} \)’.

<table>
<thead>
<tr>
<th>( \frac{d}{D} )</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>Min~Max</th>
<th>Mean ( 95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>105%</td>
<td>40%</td>
<td>36%</td>
<td>0.15~3.59</td>
<td>0.20~1.33</td>
</tr>
<tr>
<td>Female</td>
<td>102%</td>
<td>58%</td>
<td>44%</td>
<td>0.15~3.60</td>
<td>0.25~1.38</td>
</tr>
</tbody>
</table>
Average slopes of the additive falling tone

Average additive tone slopes (-st/s): ‘k’ of the final boundary tone for three tonal categories of two speakers, reference $F_0=75$Hz.
• The male speaker has slight steeper falling slopes than the female speaker in Tone 1 and Tone 2. However, Tone 1 and Tone 2 have steeper final falling than T3 for both speakers.

• While T1 and T2 are grouped in one cluster (average from -53~43st/s) and T3 (average from -35~33st/s) in another for both speakers (alpha = 0.05).

<table>
<thead>
<tr>
<th>Slope</th>
<th>Max slope</th>
<th>Min slope</th>
<th>Mean (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-0.68st/s</td>
<td>-100.30st/s</td>
<td>-50st/s to -28st/s</td>
</tr>
<tr>
<td>Male</td>
<td>-5.77st/s</td>
<td>-110.83st/s</td>
<td>-60st/s to -24st/s</td>
</tr>
</tbody>
</table>
Summary

• This study divides the boundary tone for emotion in two components: lexical tone and expressive tone. The acoustic features of successive addition boundary tone were measured for a male and a female speaker. We found some slight differences between them.

• Specifically, the male speaker keeps the boundary tone in a constant length while the female speaker has longer boundary tone especially for T3.

• The additive falling tone of the male speaker is steeping than that of the female speaker. But they all have a longer expressive tone component for T1, that is, the two components are almost in equal length in T1, but have a shorter expressive tone component for T2 and T3.

• Moreover, T1 and T2 have a steeping slope than T3 for both speakers. The mean slopes of the additive falling tones range from -53st/s to -33st/s, the maximum and minimum slope shows a broad varying range from -100st/s to -0.69st/s.
• Additionally, the ‘Disgust’ boundary tone sounds more prominent than the ‘Neutral’ boundary for both speakers. All these results indicate that speakers have similar patterns in general but have some personal characteristics as well. The statistic data have been used in our perceptual experiment (presented at sp2012) to examine the interaction between the form and function of the successive boundary tone, and to explore how the emotional information is coded in parallel to pragmatic information in Chinese.

• The results show that T1~T3 have the same falling successive addition patterns as suggested by Chao, but for our data, T4 and ‘neutral tone’ might have some variations, there is no successive addition falling after T4 and Neutral tone in ‘Disgust’ boundary tone. This suggests that the successive addition tone can have variations in expressing multifarious emotions or attitudes in additional to the patterns given by Chao.
References

• 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.
• Gussenhoven, C. 2004. The Phonology of Tone and Intonation. Cambridge: CUP.
• Wang, J. and Smith N., Studies in Chinese phonology, Walter de Gruyter.
谢谢！
[»iEó . »iE ]

Thanks!
liaj@cass.org.cn