International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 12, Issue 3 (March 2016), PP.39-45

Research on the Realization of Tones in Mandarin Sentence

Ruixian Huang
School of Foreign Language and Trade, Guangzhou City Construction College, Guangzhou, China

Abstract:- In this study, the realization of tones in Mandarin sentences is analyzed, and it is found that there is downstep for high (H) tones. They always decline to a lower level compared to the previous ones. The lowering degree of the initial downstep is always larger than the subsequent ones, but there are no significant differences in the lowering degrees for the later HL sequences. The speakers do not only raise the first high tone when there are more tones, but also lower the final high tone. Regarding pitch range, compared with utterance with two HL sequences, the total pitch ranges of longer sentences are great. The pitch range of the final HL sequence is the largest, and the range of the first HL sequence is the second largest, with the range near the middle of the sentence to be the least.

Keywords:- Tone, pitch, sentence

I. INTRODUCTION

This study focus on the realization of tones in the context of downstep, which refers to the stepwise declining of High (H) tones in the contexts in tone languages. In automatic downstep, high tones are lowered in utterances of alternating high and Low (L) tones. In non-automatic downstep, often represented as H!H, there is no obvious conditioning low tone between the two high tones [1]. For downstep, each successive high tone in longer downstepping sentences is lower than the forgoing one, resulting a cumulative staircase pattern. The idea of downstep has been further extended to studies of many non-tonal languages, and has been realized in quantitative models of intonation pattern in languages such as English, Swedish, and Japanese [2].

In tonal languages, the speakers may apply the mechanism ‘foresight’ in producing long downstepping sequences. Stewart [3] argued that the pitch of high tones in downstepping utterances in Akan is sensitive to the number of the following tones. He claimed that the pitch of any high tone may be raised by as many levels as there are HL sequences in the following part of the phrase, while the last high tone in the sentence tends to be realized at a constant level, i.e., its basic pitch. On the contrary, Schachter [4] argued that the pitch of the first high tone in Akan is usually phonetically the same, regardless of the number of the subsequent downsteps, while later high tones decline to lower and lower values as the number of tones increases.

A number of studies related to tonal downstep in Mandarin Chinese has been done, Xu [5] stated that anticipatory and carry-over tonal effects co-exist in Mandarin Chinese, and they differ both in scale and in nature. Carry-over effects are usually assimilatory: the starting pitch of a tone is assimilated to the end value of a forgoing tone. Anticipatory effects, on the other hand, are mostly dissimilatory: a lower onset value of a tone raises the maximum pitch value of a forgoing tone. Shih [6] argued that the pitch contour of a Mandarin Chinese utterance is affected by a lot of factors, such as declination, downstep and final lowering, etc. Huang et al. [7] reported that in downstep in Mandarin Chinese, the low tone will reduce the pitch range of the subsequent syllables, and the main effect of downstep is mostly applied on the topline.

The study reported here will examine the pitch values of sentences with 2 to 5 HL sequences, i.e. utterances with 4 to 10 syllables, and it is aimed to find out whether Mandarin Chinese utterances display downstep effects across high and low tones. At the same time, the realization of F0 range will also be probed. The following questions are addressed:

(a) Do high and low tone sequences show lowering effects? How about the lowering degree at different position of the sentence?

(b) Is the first high and low tone scaled higher as the number of tones increases? Does the final high and low tone decline to lower values in sentences with more downsteps?

(c) Does the overall F0 range of the utterances expand as the number of tones increases? Are there variation of F0 ranges of the HL sequences at different position of the sentence?

(d) Does the first HL sequence expand as the number of downsteps increases? Does the final HL sequece expand with more downsteps?

II. METHODOLOGY

A. Stimuli

There are four tones in Mandarin Chinese. Tone 1 is high, Tone 2 is rising, Tone 3 is low falling, and Tone 4 is a falling one. For the purpose of addressing the questions of the present experiment, only Tone 1 (H) and Tone 3 (L) sequences are used. In the sentences designed, high and low tones alternate on successive syllables, that is, in the pattern of HLHL, HLHLHL, etc. In the sentences, each set contained 4 utterances, with 2 to 5 HL sequences, i.e. with 4 to 10 syllables in length. The following is one set of the utterances,

(1) Bianxie chugao. (2 HL sequences)
To compile the draft.
(2) Bianxie gepu chugao. (3 HL sequences)
To compile the music draft.
(3) Qinshou bianxie gepu chugao. (4 HL sequences)
To compile the music draft himself.
(4) Kaishi qinshou bianxie gepu chugao. (5 HL sequences)
To begin to compile the music draft himself.

In the corpus designed, there are four such sets, which make a total of 16 utterances.

B. Subjects and Recording

The sentences used in this experiment are recorded by eight native speakers of standard Mandarin Chinese, four males and four females. The test sentences for the experiment were recorded in a sound-proof room, with a short practice session before the recording. The sentences were presented in random order and were read three times by the subject, with the order of each repetition randomized separately. In the recording, the subjects were told to read in normal speed, in a natural style, without narrow focus. In this way, the subjects are expected to read each sentence as broad-focused. The total utterances used in this study are 384 (16 utterances × 3 repetitions × 8 speakers).

C. Measurements

The recording is segmented and labeled, and F0 is extracted using Praat [8]. The extracted pitch is manually verified with reference to the cycle in the waveform. In this experiment, for the purpose of normalizing the pitch difference among the speakers, semitone is used as the unit of F0, instead of Hertz, and the conversion is done by the following formula,

$$St = 12 \times \log_2 \left(\frac{F_0}{F_{0\min}} \right) \qquad (1)$$

In (1), F_0 is the pitch value in Hertz, $F_{0\min}$ as the lower bound of pitch range of the speaker, and St is the semitone value.

1) Average pitch value of high tone: As Tone 1 in Mandarin Chinese is a level tone, average pitch value is computed, instead of using the high point value. Average pitch value refer to the mean of the pitch values of a tone. For example, if the duration of the voiced part is 200 ms, the extraction will get 20 pitch values within it, and the average pitch value is the mean of the 20 values.

2) Lowing degree: For the purpose of inspecting the extent of the lowering effect, lowering degree is computed. It is the difference between successive high or low tones, which is calculated by the following formula,

$$D_i = St_i - St_{i+1} \qquad (2)$$

In (2), D_i stands for the value of lowering degree at position 'i', St_i for the pitch value of the tone at the same position, and St_{i+1} for the pitch value of the following tone. Statistic analysis is done in SPSS.

3) Pitch range: Pitch range is defined as the range from the top to the bottom point of the pitch, as is shown in (3),

$$R = St_t - St_b \qquad (3)$$

In (3), St_t and St_b refer to the pitch values of the top and bottom points in semitone respectively, and R is the pitch range.

III. RESULTS

Fig. 1 displays the average pitch values of utterances of various lengths for all the eight speakers, with (a) to (d) presenting values for utterances with two to five HL sequences respectively. In these graphs, the x-axis displays duration, and the y-axis displays pitch values in semitone. The line segments stand for the pitch contour, with each segment for one syllable, level one for H tone and low falling one for L tone.

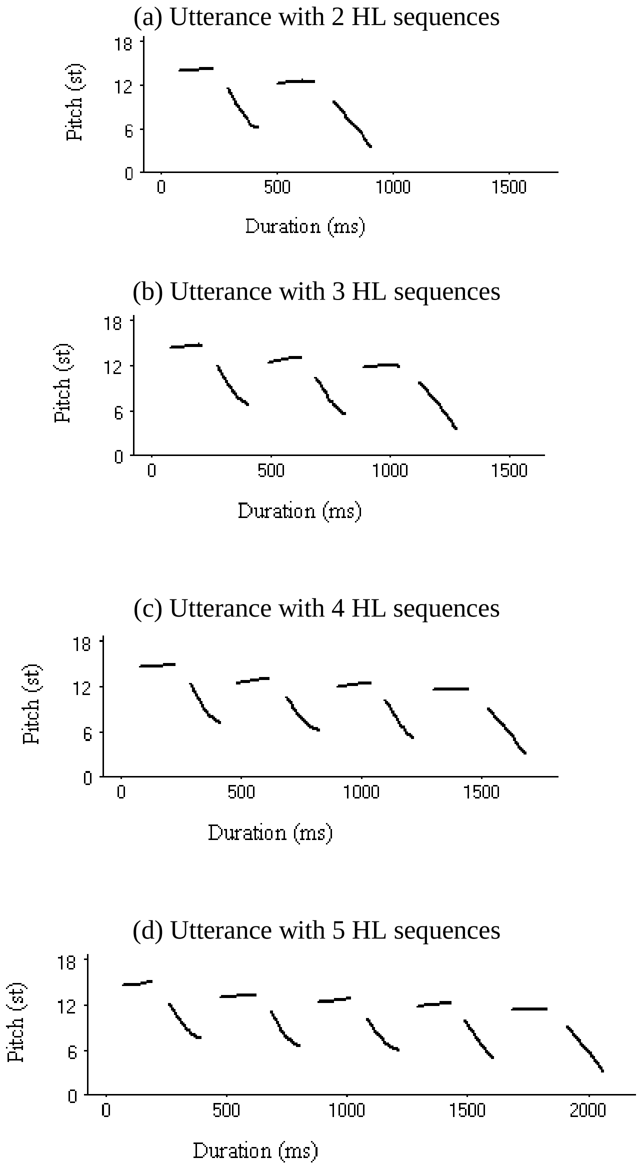

Fig. 1. Pitch contour of utterances of various lengths[a]

a. The line segments present the pitch contour, with each segment for one syllable, level one for H tone and low falling one for L tone.

A. Average pitch value

From Fig. 1, it is shown that there is a prominent gradual lowering of high tones, strongly resembling downstep throughout the utterances. Detailed analysis will be given as follow.

1) Utterance with two HL sequences: For utterances with two HL sequences, there is a downward trend for the high tones. It is shown from a repeated measures ANOVA result that there is significant difference between their pitch values: $F(1, 95) = 397.5$, $p < 0.001$, with the pitch of the first high tone (H_1) higher than that of the second high tone (H_2).

2) Utterance with three HL sequences: In regard to utterances with three HL sequences, downstep also exists for the high tones. Repeated measures ANOVA results show that H_1 is higher than H_2: $F(1, 95) = 513.2$, $p < 0.001$, and H_2 higher than the third H (H_3): $F(1, 95) = 165.8$, $p < 0.001$.

3) Utterance with four HL sequences: As for utterances with four HL sequences, there is still a gradual lowering of H tones. It is displayed from repeated measures ANOVA results that significant difference exists between pitch values of successive H tones, H_1 vs. H_2: $F(1, 95) = 376.3$, $p < 0.001$; H_2 vs. H_3: $F(1, 95) = 81.5$, $p < 0.001$; H_3 vs. H_4: $F(1, 95) = 89.6$, $p < 0.001$, with the preceding H tone higher than the following one.

4) Utterance with five HL sequences: When there are five HL sequences in an utterance, repeated measures ANOVA results show that the differences between the average pitch values of the successive H tones are still significant, H_1 vs. H_2: $F(1, 95) = 231.3$, $p < 0.001$; H_2 vs. H_3: $F(1, 95) = 82.5$, $p < 0.001$; H_3 vs. H_4: $F(1, 95) = 56.2$, $p < 0.001$; H_4 vs. H_5: $F(1, 95) = 97.3$, $p < 0.001$, with the foregoing H tone higher than the subsequent one.

B. Lowering degree

Table 1 shows the average lowering degrees (in st) of high tones for utterances of various lengths. In Table 1, D_i stands for the lowering degree between the pitch values of H_i and H_{i+1}, for example, D_1 stands for the lowering degree between those of H_1 and H_2, and D_2 for that between H_2 and H_3. For utterances with 2 HL sequences, the lowering degree between the two H tones is 1.82 semitones.

Table 1. The average lowering degree (in st) of H tones for utterances of various lengths[a]

N. of HLs.	**Lowering degree**			
	D_1	D_2	D_3	$\mathbf{D_4}$
2	1.82			
3	1.91	0.87		
4	2.05	0.68	0.76	
5	**1.61**	**0.62**	**0.48**	**0.67**

a. D_i stands for the lowering degree between the pitch values of H_i and H_{i+1}.

As for utterances with three HL sequences, it is displayed from repeated measures ANOVA results that the lowering degrees for successive high tones are significantly different, with D_1 greater than D_2: $F(1, 95) = 73.5$, $p < 0.001$.

Regarding utterances with four HL sequences, repeated measures ANOVA results show that there is significant difference among the lowering degrees: $F(2, 190) = 89.3$, $p < 0.001$. Further analysis shows that D_1 is greater than D_2 and D_3, D_1 vs. D_2: $F(1, 95) = 125.3$, $p < 0.001$; D_1 vs. D_3: $F(1, 95) = 138.1$, $p < 0.001$, but there is no significant difference between D_2 and D_3: $F(1, 95) = 0.46$, $p = 0.537$.

When there are five HL sequences in an utterance, it is indicated from repeated measures ANOVA results that the difference among the lowering degrees is also significant: $F(3, 285) = 42.3$, $p < 0.001$. Detailed analysis shows that D_1 is larger than at any of the other positions, D_1 vs. D_2: $F(1, 95) = 43.5$, $p < 0.001$; D_1 vs. D_3: $F(1, 95) = 157.3$, $p < 0.001$; D_1 vs. D_4: $F(1, 95) = 51.2$, $p < 0.001$. However, there is no significant difference among D_2, D_3 and D_4: $F(2, 190) = 1.25$, $p = 0.356$.

C. The first high tone

In this subsection, the pitch values of the utterance first high tones are analyzed. If the first high tones are scaled higher as the number of downsteps increases, the pitch values should rise for longer utterances. Table 2 displays the average values of the first high tones and the ANOVA result, from which it is shown that this is the case. Result from repeated measures ANOVA shows that there is significant difference among them. Further analysis shows that, compared to that of utterance with 2 HL sequences, the initial high tones of longer utterances are higher, 2-HL vs. 3-HL: $F(1, 95) = 39.2$, $p < 0.001$; 2-HL vs. 4-HL: $F(1, 95) = 43.6$, $p < 0.001$; 2-HL vs. 5-HL: $F(1, 95) = 39.4$, $p < 0.001$. However, for utterences with three, four and five HL sequences, there is no significant difference among the pitch values of the initial H tones: $F(2, 190) = 0.79$, $p = 0.362$.

Table 2 Average pitch values (in st) of the initial H tones and the ANOVA result

N. of HLs.	**2**	**3**	**4**	**5**
Pitch	14.15	14.62	14.73	**14.76**
ANOV	**$F(3, 285) = 17.2$, $p < 0.001$**			

D. The final high tone

In the context of downstep, it is possible that as the number of downsteps increases, the final high tone may drop to a lower and lower value. Table 3 displays the average pitch values of the final high tones and the ANOVA result, which shows that there is really a continual lowering for the final high tones. It is displayed from a repeated measures ANOVA result that there is a significant difference among them. Further analysis shows that the final high tones of longer utterances are lower than those of shorter ones, regardless of the length of the sentence. To be specific, that of 3 HL utterance is lower than that of 2 HL utterance: $F(1, 95) = 72.3$, $p < 0.001$; that of 4 HL utterance lower than that of 3 HL utterance: $F(1, 95) = 35.2$, $p < 0.001$, and in turn that of 5 HL utterance is lower than that of 4 HL utterence: $F(1, 95) = 12.57$, $p = 0.002$. The final high tones do drop downwards as the number of downsteps increases.

Table 3 Average pitch values (in st) of the final H tones and the ANOVA result

N. of HLs.	**2**	**3**	**4**	**5**
Pitch	12.42	12.0	11.7	**11.5**
ANOVA	**F(3, 285) = 76.7, p < 0.001**			

E. The overall pitch ranges of the sentences

The overall pitch range of the sentence is the difference between the top and bottom F0 points of the sentence. Table 4 presents the average overall pitch ranges and the ANOVA result, which shows that there is significant difference among them. Further analysis indicates that, compared to utterance with two HL sequences, the overall pitch ranges of longer utterances are great, 2-HL vs. 3-HL: $F(1, 95) = 4.62$, $p = 0.028$; 2-HL vs. 4-HL: $F(1, 95) = 18.56$, $p < 0.001$; 2-HL vs. 5-HL: $F(1, 95) = 16.36$, $p < 0.001$. However, there is no significant difference among the overall pitch ranges of utterances with 3, 4 and 5 HL sequences: $F(2, 190) = 2.78$, $p = 0.063$.

Table 4 Average overall pitch ranges of the utterances and the ANOVA result

N. of HLs.	**2**	**3**	**4**	**5**
Pitch	10.71	11.23	11.82	**11.75**
ANOV	**F(3, 285) = 7.89, p < 0.001**			

F. The pitch ranges of the HL sequences

This subsection will examine on the pitch ranges of the HL sequences within the sentence, which refer to the differences between the top and bottom F0 points of the HL sequences. The results of the various lengths of sentences will be shown separately.

1) Utterance with two HL sequences: For utterances with two HL sequences, it is displayed from a repeated measures ANOVA result that there is significant difference between the pitch ranges of the two HL sequences: $F(1, 95) = 12.46$, $p < 0.001$, with the second one larger then the first.

2)Utterance with three HL sequences: Regarding utterances with three HL sequences, repeated measures ANOVA results indicate that there are significant difference among the pitch ranges: $F(2, 190) = 15.67$, $p < 0.001$. Further analysis indicates that the pitch range of the last HL is largerr than the first: $F(1, 95) = 6.32$, $p = 0.021$, and that of the first HL is greater than the second one: $F(1, 95) = 11.35$, $p = 0.001$.

3) Utterance with four HL sequences: In regard to utterances with four HL sequences, it is shown from repeated measures ANOVA results that significant difference exists between the pitch ranges of the HL sequences: $F(3, 285) = 38.62$, $p < 0.001$. Further analysis indicates that the pitch range of the last HL is larger than the first one: $F(1, 95) = 21.35$, $p < 0.001$, and the third one: $F(1, 95) = 51.57$, $p < 0.001$, and the pitch range of the second one is smaller than the first one: $F(1, 95) = 49.42$, $p < 0.001$, as well as the third one: $F(1, 95) = 21.73$, $p < 0.001$. However, there is no significant difference between the first one and the third one: $F(1, 95) = 3.58$, $p = 0.093$.

4) Utterance with five HL sequences: When there are five HL sequences in an utterance, repeated measures ANOVA results indicate that the differences between the pitch ranges of the HL sequences are still significant: $F(4, 380) = 39.68$, $p < 0.001$. Detailed analysis displays that the F0 range of the last HL is larger than the first one: $F(1, 95) = 28.72$, $p < 0.001$, as well as the fourth one: $F(1, 95) = 52.36$, $p < 0.001$, and the pitch range of the fourth one is larger than the second one: $F(1, 95) = 13.39$, $p = 0.001$, as well as the third one: $F(1, 95) = 19.57$, $p < 0.001$. However, there is no significant difference between the first one and the fourth one: $F(1, 95) = 1.79$, $p = 0.243$, nor is there between the second one and the third one: $F(1, 95) = 0.148$, $p = 0.627$.

IV. DISCUSSION

A. High tones

From this experiment, it is displayed that, there is downstep for high tones, and this is the case for sentences of various lengths, whether those with 2 or 3 HL sequences, or those with 4 or 5 HL sequences. This is also the case for high tones at various positions, whether those at the earlier part of a sentence, or those at the later part. High tones always decline to a lower level compared to the previous ones.

If the lowering degree is analyzed, it is shown that the degree of the initial downstep is always larger than the rest ones. No matter how long the sentence is, this is always true. In regard to the rest of the downsteps, that is, the second, the third and the fourth one, no significant differences is found among them.

For downstep, Liberman and Pierrehumbert [9] proposed the Gradient model, which defines downstepping patterns as a gradual decline toward an abstract reference line, or asymptote. Their approach of pitch assignment describes an exponentially declining curve in which each step down is proportionally the same as the preceding one in terms of its distance from the reference line: Later downstep intervals are gradually smaller than earlier ones, and tend to become small as the reference line is approached. This model could be called a 'soft-landing' one of downstep realization as it describes a curve similar to that of an plane gliding smoothly down to a landing strip.

As for high tones, the downstepping pattern got from this experiment is similar to the soft-landing model to some extent, that is, the degree of the first drop is the largest. However, in this study, it is shown that there are no differences in the degrees of drop for the later HL sequences. The gradient for the later HL sequences keeps constant, rather than proportionally getting reduced.

Regarding the pitch values of the sentence initial high tones, it is presented that they shift upwards as the number of tones increases. This pattern can be explained by 'foresight' in tone production, that is, the pitch of high tones in sentences is sensitive to the number of the subsequent tones. When there are more tones in an utterance, the speaker foresees this, and will specify the first high tone at a higher level. However, detailed analysis shows that this 'foreseeing' mechanism stops to be in effect when there are 3 or more than 3 HL sequences in an utterance. That is to say, the effect of 'foresight' is limited to some extent, and the reason for this is that, there is a limit for a speaker's maximum pitch value.

It is shown from the previous section that the speaker does not only raise the first high tone when there are more tones, but also depress the last high one. So there are two measures for the realization of dowstep in Mandarin Chinese, to raise the first high tone and to depress the final high one. Further investigation shows that the measure of the depressing of the last high tone is in effect for utterance with 4 or 5 HL sequences. Due to the constraint of the speaker's pitch range, the 'foresight' effect is limited. However, there is no limitation for the mechanism for the last high tone depressing. Compared to lowering the last high tone, raising the first high tone is more energy consuming. Therefore, people tend to take the energy saving measure to realize the tones.

B. Pitch range

As for overall pitch range of the sentence, it is shown that compared to utterance with 2 HL sequences, the overall pitch ranges of longer utterances are large. However, there is no significant difference among the overall pitch ranges of utterances with three, four and five HL sequences. When there are more downsteps, the overall pitch range will be large. For utterance with 2 HL sequences, there is only one downstep, so its pitch range is comparatively small. On the other side, in actual communication, the pitch range of a sentence tends to be constant. Therefore, for utterances with three, four and five HL sequences, their overall pitch ranges are much the same.

In regard to the pitch ranges of the HL sequences within the sentences, it is displayed that the pitch range of the final HL sequence is the largest, and this is the case for utterance of various lengths. For the three HL utterance, the first one is the second largest, and the second one the smallest. When there are 4 HL sequences in a sentence, the first and penultimate HL is the second largest, with that in the middle the smallest. Coming to utterance with 5 HL sequences, the second largest are also the first and the penultimate ones, with the 2 HL sequences in the middle the smallest.

Pitch range, which is the difference between the top and bottom F0 point, is investigated in this study, and it is shown that the pitch range of the earlier HL sequences are relatively small, with that of the final downstep the largest. This is due to the 'final lowering' effect. Final lowering, which is the lowering of pitch at the end of a sentence, has been observed in many languages, like Spanish and Yoruba [10].

It is indicated that the pitch ranges of HL sequences near the middle of the sentence tend to be the smallest. This is because words in that part are less prominent in communication. Words near the end of the sentence tend to have 'final lowing' effect, which have demarcative function, i.e., marking the end of a sentence. As a result, the pitch range of the last HL sequence is the largest. The initial HL sequence is at the beginning of a sentence, with the function of attracting the listener's attention, so it tends to be comparative more prominent, and its pitch range is the second largest, only next to the last one. If the words near the middle of a sentence are not under focused condition, they tend to be the least prominent in communication, and their pitch ranges tend to be the least.

V. CONCLUSION

The pattern of the production of tone in Mandarin Chinese is analyzed in this study, and it is shown that there is downstep for high tones, regardless of the length of the sentence, or the position of the tone in the sentence. H tones always decline to a lower scale compared to the previous ones. The extent of the first downstep is always larger than the subsequent ones, but there are no differences in the extent of downsteps for the later HL sequences. The sentence initial high tones will raise upwards as the number of downsteps increases, but as there is a limit for the speaker's maximum pitch value, it disappears in longer utterances. The speakers will reduce the final high tone when there are more downsteps, and this is the case for sentences of any lengths. The speakers tend to take the energy saving measure to realize the downstepping effect.

REFERENCES

[1]. Y. O. Laniran and G. N. Clements, "Downstep and high raising: interacting factors in Yoruba tone produntion," *Journal of Phonetics*, 2003, (31), pp. 203–250.

[2]. D. R. Ladd, "*Intonational phonology*," Cambridge: Cambridge University Press,1996.

[3]. J. M. Stewart, "The typology of the Twi tone system," *Bulletin of the Institute of African Studies*, Vol.1, Legon, Ghana, 1965.

[4]. P. Schachter, "Some comments on J. M. Stewart's The typology of the Twi tone system," *Bulletin of the Institute of African Studies*, Vol.1, Legon, Ghana, 1965.

[5]. Y. Xu, "Contextual tonal variations in Mandarin," *Journal of Phonetics*, 1997, (25), pp. 61–83.

[6]. C. Shih, "Declination in Mandarin. Intonation: theory, models and applications," *Proceedings of an ESCA work shop*, Athens, Greece, 1997, pp. 293–296.

[7]. X. Huang, Y. Yang and S. Lu, "Experimental studies on downstep in Chinese intonation, " *Acta Acustica*, 2007, (32), pp. 57–61.

[8]. P. Boersma, "praat, a syntem doing phonetics by computer," *Glot International*, 2001, 5:9/10, pp. 341–345.

[9]. M. Liberman and J. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language and sound structure*, M. Aronoff and R. T. Oehrle, Eds. Cambridge, MA: MIT Press, 1984, pp. 157–233.

[10]. H. Truckenbrodt, "Final lowering in non-final position," *Journal of Phonetics*, 32, pp. 313–348, 2004.