International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 3 Issue 10 ‖ October. 2015 ‖ PP.17-23

Effect of down step on High-Low Tones in Chinese

Maolin Wang

College of Chinese Language and Culture / Overseas Chinese Education Research Institute, Jinan University, Guangzhou, China

Abstract: In this study, the downstep principle in Chinese tone production is analyzed, and it is found that downstep exists for high (H) tones. They always drop to a lower scale compared to the forgoing ones. The degree of the first downstep is always larger than the following ones, but there are no differences in the degrees of downsteps for the later HL sequences. The speaker does not only raise the initial H tone when there are more downsteps, but also depress the final H tone.

Keywords: Downstep; tone; pitch

I. Introduction

Downstep refers to the stepwise lowering of High (H) tones in certain contexts in tone languages. In automatic downstep, H tones are lowered in sequences of alternating H and Low (L) tones. In nonautomatic downstep, often noted as $H^{!}H$, there is no overt conditioning L tone between the two H tones [1]. In downstep, each successive H tone in longer downstepping sequences is lower than the preceding one, creating a cumulative ‘staircase’ pattern. The concept of downstep has been extended to studies of many nontonal languages, and has been incorporated in quantitative models of intonation in languages such as English, Swedish, and Japanese [2].

With the effect of downstep, speakers of tone languages may employ ‘foresight’ in producing long downstepping sequences. Stewart [3] claimed that the pitch of H tones in downstepping sequences in Akan is sensitive to the number of following downsteps. He stated that the pitch of any particular high tone is raised by as many levels as there are downsteps in the subsequent part of the phrase, while the last H tone in the sequence tends to be realized at a constant level, its basic pitch. On the contrary, Schachter [4] maintained that the pitch of the first H tone in Akan is normally phonetically the same regardless of the number of the following downsteps, while later H tones descend to lower and lower values as the number of downsteps increases.

In Chinese, there have been a number of studies related to tonal downstep, Xu [5] argued that anticipatory and carry-over tonal influences co-exist in Chinese, and they differ both in magnitude and in nature. Carry-over effects are mostly assimilatory: the starting F0 of a tone is assimilated to the offset value of a previous tone. Anticipatory effects, on the other hand, are mostly dissimilatory: a low onset value of a tone raises the maximum F0 value of a preceding tone. Shih [6] pointed out that the F0 contour of a Chinese utterance is affected by a number of factors, such as declination, downstep and final lowering, etc. Huang et al. [7] reported that in downstep in Chinese, the low tone will compress the pitch range of the following syllables, and the main effect of downstep is on the topline.

The experiment reported here will investigate the pitch values of utterances with 2 to 5 HL sequences, i.e. utterances with 4 to 10 syllables, and the aim is to find out whether Chinese utterances display downstep effects across H and L tones. The following questions are addressed:

(a) Do H tone sequences show downstep effects?
(b) What about the degree of downstep at different position of the utterance?
(c) Is the initial H tone scaled higher as the number of downsteps increases?
(d) Does the final H tone drop to lower values in utterances with more downsteps?

II. Method

Stimuli. In Chinese, there are four tones. Tone 1 is high, Tone 2 rising, Tone 3 low falling, and Tone 4 is a falling one. In order to address the questions of the present experiment, only Tone 1 (H) and Tone 3 (L) sequences are used. In the utterances designed, H and L tones alternate on successive syllables, that is, in the pattern of HLHL, HLHLHL, etc. In the corpus, 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.

Subjects and Recording. The utterances used in this experiment are recorded by eight native speakers of standard Chinese, four males and four females. The test utterances for the experiment were recorded in a sound-treated room at the Phonetics Laboratory in Jinan University, with a short practice session before the actual recording. The utterances were presented in random order and were read 3 times by the subject, with the order of each repetition randomized separately. In the recording, the subjects were instructed to read in normal speed, in a natural style, without narrow focus. By this means, the subjects are expected to read each utterance as broad-focused. The total utterances used in this study are 384 (16 utterances × 3 repetitions × 8 speakers).

III. Measurements.

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

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

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

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

Degree of downstep. In order to inspect the extent of the downstep effect, degree of downstep is calculated. It is the difference between successive H tones, which is computed by the following formula,

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

In (2), D_i stands for the value of downstep degree at position ‘i’, St_i for the average pitch value of H tone at the same position, and St_{i+1} for the pitch value of the following H tone.

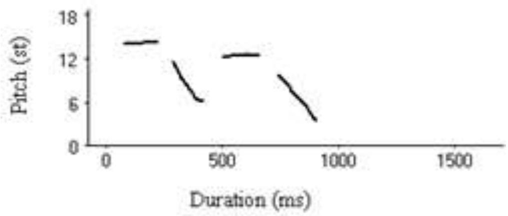
Analysis is done by a self-written visual basic program, by which the average F0 value of the H tones are calculated, and statistic analysis is done in SPSS.

IV. Results

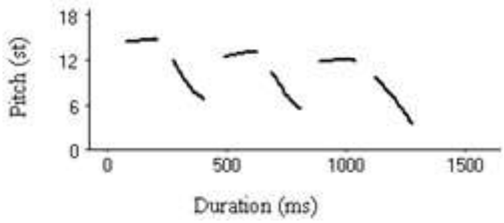
Fig. 1 graphs the mean pitch values of utterances of various lengths for all the eight speakers, with (a) to (d) presenting values for utterances with 2 to 5 HL sequences respectively. In these graphs, the x-axis displays duration, and the y-axis displays pitch values in semitone. The line segments present the pitch contour, with each segment for one syllable, level one for H tone and low falling one for L tone.

Average pitch value. From Fig. 1 it can be seen that there is a prominent gradual lowering of H tones strongly resembling downstep throughout the utterances. Detailed analysis will be given in the following subsections.

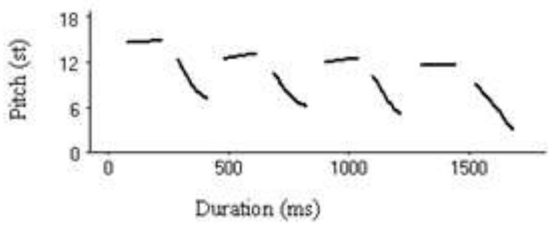
(a) Utterance with 2 HL sequences



(b) Utterance with 3 HL sequences



(c) Utterance with 4 HL sequences



(d) Utterance with 5 HL sequences

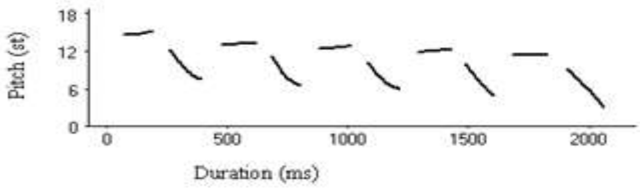


Figure 1. Pitch contour of utterances of various lengths[a]

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.

Utterance with 2 HL sequences. For utterances with 2 HL sequences, there is a downward trend for the H tones. It is shown from a repeated measures ANOVA result that there is significant difference between their pitch values: $F(1, 95) = 413.7$, $p < 0.001$, with the pitch of the first H tone (H_1) higher than that of the second H tone (H_2).

Utterance with 3 HL sequences. In regard to utterances with 3 HL sequences, downstep also exists for the H tones. Repeated measures ANOVA results show that H_1 is higher than H_2: $F(1, 95) = 492.7$, $p < 0.001$, and H_2 higher than the third H (H_3): $F(1, 95) = 173.7$, $p < 0.001$.

Utterance with 4 HL sequences. As for utterances with 4 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: $F(1, 95) = 388.5$, $p < 0.001$; H_2 vs. H_3: $F(1, 95) = 79.2$, $p < 0.001$; H_3 vs. H_4: $F(1, 95) = 88.8$, $p < 0.001$, with the preceding H tone higher than the following one.

Utterance with 5 HL sequences. When there are 5 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: $F(1, 95) = 224.1$, $p < 0.001$; H_2 vs. H_3: $F(1, 95) = 83$, $p < 0.001$; H_3 vs. H_4: $F(1, 95) = 54.1$, $p < 0.001$; H_4 vs. H_5: $F(1, 95) = 98.4$, $p < 0.001$, with the foregoing H tone higher than the subsequent one.

Degree of downstep. Table 1 displays the average degrees of downstep (in st) of H tones for utterances of various lengths. In Table 1, D_i stands for the degree of downstep between the pitch values of H_i and H_{i+1}, for example, D_1 stands for the degree of downstep 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 degree of downstep between the two H tones is 1.73 semitones.

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

N. of	Degree of downstep			
HLs.	D_1	D_2	D_3	D_4
2	1.73			
3	1.83	0.88		
4	2.02	0.67	0.74	
5	1.59	0.61	0.49	0.64

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

As for utterances with 3 HL sequences, it is shown from repeated measures ANOVA results that the degrees of downstep for successive H tones are significantly different, with D_1 greater than D_2: $F(1, 95) = 72.6$, $p < 0.001$.

In regard to utterances with 4 HL sequences, repeated measures ANOVA results show that there is significant difference among the degrees of downstep: $F(2, 190) = 87.2$, $p < 0.001$. Further analysis shows that D_1 is greater than D_2 and D_3, D_1 vs. D_2: $F(1, 95) = 101.6$, $p < 0.001$; D_1 vs. D_3: $F(1, 95) = 142.9$, $p < 0.001$, but there is no significant difference between D_2 and D_3: $F(1, 95) = 0.39$, $p = 0.529$.

When there are 5 HL sequences in an utterance, it is indicated from repeated measures ANOVA results that the difference among the degrees of downstep is also significant: $F(3, 285) = 37.2$, $p < 0.001$. Detailed analysis shows that D_1 is larger than at any of the other positions, D_1 vs. D_2: $F(1, 95) = 40.2$, $p < 0.001$; D_1 vs. D_3: $F(1, 95) = 145.6$, $p < 0.001$; D_1 vs. D_4: $F(1, 95) = 46.9$, $p < 0.001$. However, there is no significant difference among D_2, D_3 and D_4: $F(2, 190) = 1.19$, $p = 0.305$.

The Initial H Tone. In this subsection, the pitch values of the utterance initial H tones are analyzed. If initial H 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 initial H tones and the ANOVA result, from which it can be seen that this is the case. Result from a 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 H tones of longer utterances are higher, 2-HL vs. 3-HL: $F(1, 95) = 38.9$, $p < 0.001$; 2-HL vs. 4-HL: $F(1, 95) = 41.5$, $p < 0.001$; 2-HL vs. 5-HL: $F(1, 95) = 36.7$, $p < 0.001$. However, for utterences with three, four and five HL sequences, there is no significant difference among the pitch values of the initial H tones: $F(2, 190) = 0.83$, $p = 0.436$.

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

N. of HLs.	2	3	4	5
Pitch	14.1	14.	14.7	14.7
ANOVA	$F(3, 285) = 16.7$, $p < 0.001$			

The Final H Tone. With the effect of downstep, it is expected that as the number of downsteps increases, the final H tone will drop to lower and lower values. Table 3 presents the average pitch values of the final H tones and the ANOVA result, which indicates that there is indeed a continual lowering for the final H tones. It is shown from a repeated measures ANOVA result that there is also a significant difference among them. Further analysis shows that the final H tones of longer utterances are lower than those of shorter ones, regardless of the length of the utterance. To be specific, that of three HL utterance is lower than that of two HL utterance: $F(1, 95) = 70.9$, $p < 0.001$; that of four HL utterance lower than that of three HL utterance: $F(1, 95) = 12.6$, $p = 0.001$, and in turn that of five HL utterance is lower than that of four HL utterence: $F(1, 95) = 8.04$, $p = 0.006$. The final H 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	2	3	4	5
Pitch	12.3	11.9	11.6	11.4
ANOV	$F(3, 285) = 76.7$, $p < 0.001$			

V. Discussion

Results from this experiment show that, first of all, downstep exists for H tones, and this is true for utterances of various lengths, whether those with two or three HL sequences, or those with four or five HL sequences. This is also true for H tones at various positions, whether those at the earlier part of an utterance, or those at the later part. H tones always drop to a lower scale compared to the foregoing ones.

When the degree of downstep is analyzed, it is found that the degree of the first downstep is always larger than the rest. No matter how long the utterance is, this is always the case. As for the rest of the downsteps, that is, the second, the third and the fourth one, there are no significant differences among them.

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

The downstepping pattern observed from this experiment resembles the soft-landing model to some extent, that is, the degree of the first downstep is the greatest. However, in this study, it is found that there are no differences in the degrees of downstep for the later HL sequences. The gradient for the later HL sequences keeps constant, rather than proportionally getting smaller.

In regard to the pitch values of the utterance initial H tones, it is shown that they shift upwards as the number of downsteps increases. This pattern can be attributed to 'foresight' in tone production, that is, the pitch of H tones in downstepping sequences is sensitive to the number of the subsequent downsteps. When there are more downsteps in an utterance, the speaker foresees this and will specify the initial H tone at a higher level. However, detailed analysis shows that this 'foreseeing' mechanism ceases to be in effect when there are three or more than three HL sequences in an utterance. That is to say, the effect of 'foresight' is limited, and the reason for this is that, generally speaking, there is a limit for a speaker's maximum pitch value.

Results from the previous section also show that the speaker does not only raise the initial H tone when there are more downsteps, but also depress the final H one. So there are two measures for the implementation of dowstep in Chinese, to raise the initial H tone and to depress the final H one. Further analysis shows that the measure of the depressing of the final H tone is in effect even for utterance with four or five HL sequences. As is mentioned above, 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 final H tone depressing. Compared to lowering the final H tone, raising the initial H tone is more energy consuming. Therefore, people tend to take the energy saving measure to realize the down stepping effect.

VI. Conclusion

In this study, the downstepping pattern in Chinese tone production is analyzed, and it is found that downstep exisst for H tones, regardless of the length of the utterance, or the position of the tone. High tones always drop to a lower scale compared to the foregoing ones. The degree of the first downstep is always larger than the following ones, but there are no differences in the degrees of downsteps for the later HL sequences. The utterance initial H tones will shift upwards as the number of downsteps increases, but as there is a limit for a speaker's maximum pitch value, it ceases to be in effect for longer utterances. The speaker will depress the final H tone when there are more downsteps, and this is true for utterances of any lengths. Speakers tend to take the energy saving measure to realize the downstepping effect.

This study is significant in the speech engineering practice. In speech synthesis, when modeling the pitch contour, the effect of downstep must be taken into consideration. The degree of downstep at the beginning of an utterance should be large, and those of the later downsteps can be predicted by a linear model. The raising of the utterance initial H tone is only necessary for short utterances, while the lowering of the final H tone is applicable for utterances of any lengths.

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.