

A Study on the Pitch Range in the Context of Downstep in Chinese

Maolin Wang, Guangling Zi

(College of Chinese Language and Culture, Jinan University, Guangzhou, China)

Abstract: - In this study, the realization of pitch range of tones in the context of downstep in Chinese is analyzed. The test sentences are composed of 'High + Low + High + Low' sequences, and it is found that, compared to utterance with 2 HL sequences, the overall pitch ranges of longer utterances are great. The pitch range of the last HL sequence is the greatest, and that of the first HL sequence is the second greatest, with that near the middle of the utterance to be the least great. The pitch range of the utterance initial HL sequence of shorter utterance tends to be greater than that of longer ones, and those of the final HL sequences of long and short utterances approach a constant value.

Keywords: - *Downstep, pitch, speech, tone, utterance*

I. INTRODUCTION

Downstep, which refers to the stepwise lowering of High (H) tones in certain contexts, was originally observed in the tone languages in Africa. It is found that there are two types of downstep, automatic downstep and nonautomatic downstep. In the former case, H tones are lowered in sequences of alternating H and Low (L) tones, and in the latter case, often noted as H^hH, 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 [2].

Besides downstep, there are other principles that contribute to pitch lowering, which must be carefully distinguished from downstep, such as declination, which is a continuous, long-term pattern of pitch decline across the utterance. Unlike declination, downstep is localized at specific junctures and is usually conditioned by the tonal, lexical, morphological, or syntactic structure of the utterance in which it applies, often serving distinctive or demarcative functions.

It is shown that downstep is a common phenomenon in tone languages, so speakers 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.

There have been a number of studies related to tonal downstep in Chinese. 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] examined downstep in Chinese by subtly designed sentences. It is found that, when there is intervening low tone, it will compress the pitch range of the following syllable. It is the top line that is mainly affected, and the bottom line keeps unaffected. When there is no intervening low tone, the top line of the intonational phrase is realized in a linear downward trend.

The experiment reported here will investigate the pitch range of utterances with 2 to 5 HL sequences, i.e. utterances with 4 to 10 syllables, and the aim is to find out the pitch range realization in the context of downstep. The following questions are addressed:

- (a) Does the overall pitch range of the utterances expand as the number of downsteps increases?
- (b) Are there variation of pitch ranges of the HL sequences at different position of the utterance?
- (c) Does the initial HL sequence expand as the number of downsteps increases?
- (d) Does the final HL sequence expand with more downsteps?

II. METHOD

2.1 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.

2.2 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, 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).

2.3 Measurements

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,

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

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

This experiment aim at investigating the pitch range, which is defined as the range from the top to the bottom point, as is shown in (2),

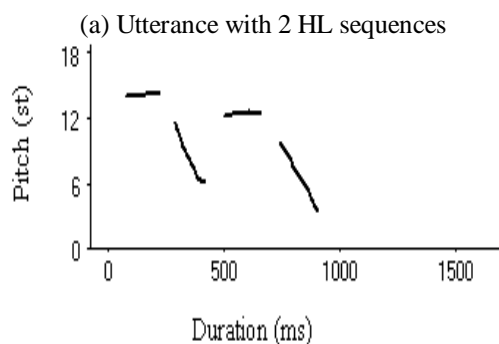
$$R = St_t - St_b \quad (2)$$

In (2), 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. Statistic analysis is done in SPSS.

III. 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.

From Fig. 1 it can be seen that there is a prominent gradual lowering of tones throughout the utterances. As for pitch range, analysis will be given in the following subsections.



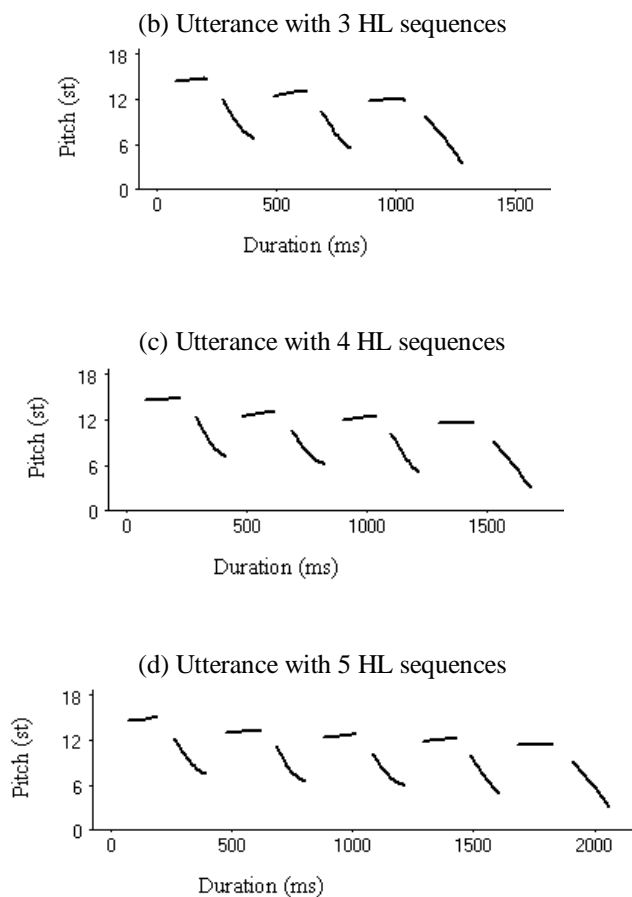


Figure 1. Pitch contour of utterances of various lengths^a

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.

3.1 The overall pitch ranges of the utterances

The overall pitch range of the utterance is the difference between the top and bottom pitch points of the utterance. Table 1 shows the average overall pitch ranges and the ANOVA result, which indicates that there is significant difference among them. Further analysis shows that, compared to utterance with 2 HL sequences, the overall pitch ranges of longer utterances are great, 2-HL vs. 3-HL: $F(1, 95) = 4.50, p = 0.037$; 2-HL vs. 4-HL: $F(1, 95) = 16.48, p < 0.001$; 2-HL vs. 5-HL: $F(1, 95) = 14.61, p < 0.001$. However, there is no significant difference among the overall pitch ranges of utterances with 3, 4 and 5 HL sequences: $F(2, 190) = 2.93, p = 0.056$.

Table 1 Average overall pitch ranges of the utterances and the ANOVA result

N. of HLs.	2	3	4	5
Pitch range	10.67	11.18	11.73	11.70
ANOVA result	$F(3, 285) = 7.65, p < 0.001$			

3.2 The pitch ranges of the HL sequences

This subsection will focus on the pitch ranges of the HL sequences within the utterance, which are the differences between the top and bottom pitch points of the HL sequences. The results of the various lengths of utterances will be given separately.

3.2.1 Utterance with 2 HL sequences

For utterances with 2 HL sequences, it is shown from a repeated measures ANOVA result that there is significant difference between the pitch ranges of the two HL sequences: $F(1, 95) = 11.05, p = 0.001$, with the second one greater than the first.

3.2.2 Utterance with 3 HL sequences

In regard to utterances with 3 HL sequences, repeated measures ANOVA results show that there are significant difference among the pitch ranges: $F(2, 190) = 14.71, p < 0.001$. Further analysis shows that the pitch range of the last HL is greater than the first one: $F(1, 95) = 5.27, p = 0.024$, and that of the first HL is in turn greater than the second one: $F(1, 95) = 10.11, p = 0.002$.

3.2.3 Utterance with 4 HL sequences

As for utterances with 4 HL sequences, it is displayed from repeated measures ANOVA results that significant difference also exists between the pitch ranges of the HL sequences: $F(3, 285) = 37.59, p < 0.001$. Further analysis shows that the pitch range of the last HL is greater than the first one: $F(1, 95) = 19.19, p < 0.001$, and the third one: $F(1, 95) = 49.44, p < 0.001$, and the pitch range of the second one is smaller than the first one: $F(1, 95) = 47.29, p < 0.001$, and the third one: $F(1, 95) = 19.91, p < 0.001$. However, there is no significant difference between the first and the third one: $F(1, 95) = 3.03, p = 0.085$.

3.2.4 Utterance with 5 HL sequences

When there are 5 HL sequences in an utterance, repeated measures ANOVA results show that the difference between the pitch ranges of the HL sequences is still significant: $F(4, 380) = 38.83, p < 0.001$. Detailed analysis shows that the pitch range of the last HL is greater than the first one: $F(1, 95) = 26.51, p < 0.001$, and the fourth one: $F(1, 95) = 50.00, p < 0.001$, and the pitch range of the fourth one is greater than the second one: $F(1, 95) = 12.07, p = 0.002$, and the third one: $F(1, 95) = 18.02, p < 0.001$. However, there is no significant difference between the first and the fourth one: $F(1, 95) = 1.58, p = 0.212$, nor is there between the second and the third one: $F(1, 95) = 0.09, p = 0.771$.

3.3 The pitch range of the initial HL sequence

In this subsection, the pitch ranges of the utterance initial HL sequences are analyzed. Table 2 displays the pitch range of the initial HL sequences and the repeated measures ANOVA result, from which it can be seen that there is significant difference among them. Further analysis shows that, the pitch range of the initial HL sequence of the two HL utterance is greater than that of the three HL utterance: $F(1, 95) = 4.35, p = 0.040$. The initial HL of the three HL utterance is in turn greater than that of four HL utterance: $F(1, 95) = 4.48, p = 0.037$, as well as that of five HL utterance $F(1, 95) = 17.83, p < 0.001$. However, there is no difference between those of four and five HL utterances.

Table 2 Average pitch ranges (in st) of the initial HL sequences and the ANOVA result

N. of HLs.	2	3	4	5
Pitch value	8.33	7.97	7.66	7.37
ANOVA result	$F(3, 285) = 13.04, p < 0.001$			

3.4 The pitch range of the final HL sequence

Table 3 presents the average pitch range of the final HL sequence and the ANOVA result, which indicates that, from utterance with two HL sequences to that with five HL sequences, there is no significant difference among their pitch range of the final HL sequences. That is to say, the pitch ranges of the utterance final HL sequences tend to be constant regardless of the length of the utterances.

Table 3 Average pitch ranges (in st) of the final HL sequences and the ANOVA result

N. of HLs.	2	3	4	5
Pitch value	8.94	8.47	8.62	8.37
ANOVA	$F(3, 285) = 1.89, p = 0.132$			

IV. DISCUSSION

Results from this experiment show that, first of all, downstep exists in Chinese utterances, 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. High tones always drop to a lower scale compared to the foregoing ones.

In regard to the overall pitch range of the utterance, it is found that compared to utterance with 2 HL sequences, the overall pitch ranges of longer utterances are great. However, there is no significant difference among the overall pitch ranges of utterances with 3, 4 and 5 HL sequences. Generally speaking, when there are more downsteps, the overall pitch range will be great. For utterance with two HL sequences, there is only one downstep, so its overall pitch range is comparatively small. On the other side, in actual communication, the

pitch range of an utterance tends to be constant. Therefore, for utterances with 3, 4 and 5 HL sequences, their overall pitch ranges are pretty much the same.

As for the pitch ranges of the HL sequences within the utterances, it is shown that the pitch range of the last HL sequence is the greatest, and this is true for utterance of various lengths. For the 3-HL utterance, the first one is the second greatest, and the second one the the smallest. When there are four HL sequences in an utterance, the first and penultimate HL is the second greatest, with that in the middle the smallest. Coming to utterance with five HL sequences, the second greatest are also the first and the penultimate ones, with the two HL sequences in the middle the smallest.

Lieberman 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 soft-landing model is put forward on the basis of the top line of the pitch contour. In this study, the pitch range, which is the difference between the top and bottom point, is investigated, and it is found that the pitch range of the earlier HL sequences are small, with that of the last downstep the greatest, which is due to the ‘final lowering’ effect. Final lowering, the lowering of pitch at the end of an utterance, has been observed in many languages, like Spanish and Yoruba [10].

It is found that the pitch ranges of HL sequences near the middle of the utterance tend to be the smallest. This is because words in the part are less prominent in communication. Words near the end of the utterance may have ‘final lowing’ effect, which have demarcative function, i.e., marking the end of an utterance. As a result, the pitch range of the final HL sequence is the greatest. The first HL sequence is at the beginning of an utterance, with the function of drawing the listener’s attention, so it should be comparative more prominent, and its pitch range is the second greatest, only next to the final one. If the words near the middle of an utterance are not under focused condition, they will be the least prominent in communication, and their pitch ranges will be the least great.

Results from the previous section also shows that, as far as the utterance initial HL sequence is concerned, the pitch range of that of the shorter utterance tends to be greater than that of longer ones. The reason for this is that, generally speaking, the pitch range of an utterance tends to be much the same. For shorter utterances, there are less downsteps, and the pitch range of a single HL sequence will be comparatively great. On the other side, there are more downsteps in longer utterances, and the pitch range of a single HL sequence will be small.

Regarding the final HL sequence, it is shown that there is no significant difference among their pitch ranges of shorter or longer utterances, i.e., they tend to be constant. Researches show that, as far as statement is concerned, the utterance final pitch approaches a constant value, and this is true regardless of the initial pitch height of the utterance, or the variation of pitch within the utterance. Studies on both tone and non-tone languages, with both read corpus and natural speech, get similar result: the utterance final pitch tends to be constant [2]. The utterance final pitch value tends to be constant, and the pitch range also much the same.

V. CONCLUSION

In this study, the pitch range realization in the context of downstep in Chinese is analyzed, and it is found that compared to utterance with 2 HL sequences, the overall pitch ranges of longer utterances are great. As for the pitch ranges of the HL sequences within the utterances, it is shown that the pitch range of the last HL sequence is the greatest, and this is true for utterance of various lengths, which is due to the final lowering effect. The pitch range of the first HL sequence is the second greatest, and that near the middle of the utterance tend to be the least great, as it is less prominent in communicative function. As the pitch range of an utterance is much of the same, the pitch range of the utterance initial HL sequence of the shorter utterance tends to be greater than that of longer ones. The pitch ranges of the final HL sequences of long and short utterances approach a constant value. The results from this study will be helpful for the pitch modeling in speech synthesis.

VI. ACKNOWLEDGEMENTS

This work was supported in part by the Social Science Research Project of Guangdong Province, Grant No. GD11CWW04, as well as the Humanity and Social Science Research Project for Colleges in Guangdong Province, Grand No. 11WYXM012.

REFERENCES

- [1] Y. O. Laniran and G. N. Clements, "Downstep and high raising: interacting factors in Yoruba tone production," *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 system doing phonetics by computer," *Glott 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.