The Realization of Low Tones in the Context of Downstep

Maolin Wang[1, a *], Ziyu Xiong[2]

[1] College of Chinese Language and Culture, Jinan University, Guangzhou, China

[2] Institute of Linguistics, Social Academy of Social Sciences, Beijing, China

[a]wangmaolin@hwy.jnu.edu.cn

Abstract. In this study, the pitch realization of low (L) tones in the context of downstep in Chinese is analyzed. The test sentences are composed of 'High + Low + High + Low' sequences, and the low point pitch of the L tones is investigated. It is found that there is an obvious downward trend for the L tones, that is, they always drop to a lower scale compared to the forgoing ones. Due to the final lowering effect, the degree of the last drop is always larger than the forgoing ones. When there are more downsteps, The initial L tone will be raised, but there is no significant variation for the final L tone.

Introduction

Downstep was originally observed in the tone languages in Africa, which refers to the stepwise lowering of High (H) tones in certain contexts [1]. In downstep, each successive H tone in longer downstepping sequences is lower than the preceding one, creating a cumulative 'staircase' pattern [2]. In tone languages, downstep is a common phenomenon, 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.

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] 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 values 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 realization of L tones in the context of downstep.

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. 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, 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).

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,

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

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.

As Tone 3 in Chinese is a low falling tone, the low point value is used in this study. The low point value is the minimum pitch value of the tone, which may best represents the feature of low tones. Statistic analysis is done in SPSS.

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.
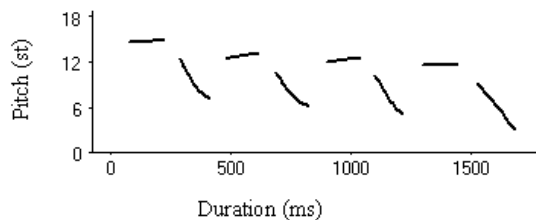
The low point value

From Fig. 1 it can be seen that there is a prominent gradual lowering of L tones throughout the utterances. Detailed analysis will be given in the following subsections.

(a) Utterance with 2 HL sequences
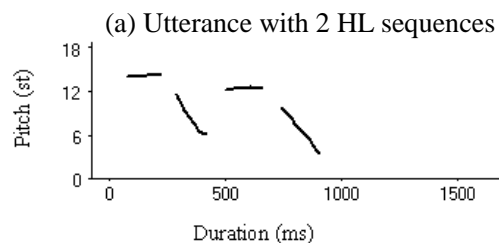
(b) Utterance with 3 HL sequences
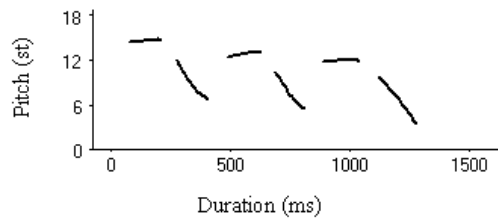
(c) Utterance with 4 HL sequences
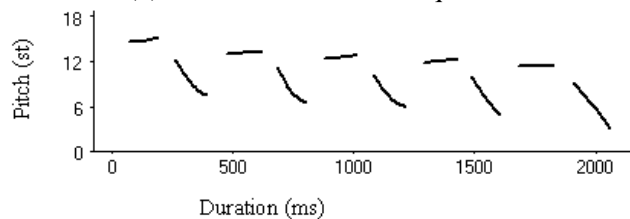

(d) Utterance with 5 HL sequences

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 L tones. It is shown from a repeated measures ANOVA result that there is significant difference between their pitch values: $F(1, 95) = 167.9$, $p < 0.001$, with the pitch of the second L tone (L_2) lower than that of the first L tone (L_1).

Utterance with 3 HL sequences. In regard to utterances with 3 HL sequences, lowering effect also exists for the L tones. Repeated measures ANOVA results show that L_2 is lower than L_1: $F(1, 95) = 83.9$, $p < 0.001$, and the third L (L_3) lower than L_2: $F(1, 95) = 138.7$, $p < 0.001$.

Utterance with 4 HL sequences. As for utterances with 4 HL sequences, there is still a gradual lowering of L tones. It is displayed from repeated measures ANOVA results that significant difference exists between pitch values of successive L tones, L_1 vs. L_2: $F(1, 95) = 129.0$, $p < 0.001$; L_2 vs. L_3: $F(1, 95) = 94.0$, $p < 0.001$; L_3 vs. L_4: $F(1, 95) = 114.9$, $p < 0.001$, with the subsequent L tone lower than the preceding 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 low point values of the successive L tones are still significant, L_1 vs. L_2: $F(1, 95) = 174.3$, $p < 0.001$; L_2 vs. L_3: $F(1, 95) = 94.8$, $p < 0.001$; L_3 vs. L_4: $F(1, 95) = 103.7$, $p < 0.001$; L_4 vs. L_5: $F(1, 95) = 122.7$, $p < 0.001$, with the subsequent L tone lower than the foregoing one.

Discussion

Results from this experiment show that, first of all, there is a downward trend for the L 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 L tones at different positions, whether those at the

earlier part of an utterance, or those at the later part. L tones always drop to a lower scale compared to the foregoing ones.

Generally speaking, in utterance with alternative H and L tones, due to the downstep effect, the subsequent H tones will be lower than the forgoing ones. In the context of downstep, the pitch of the L tones will also be affected, with the later L tones lower than the earlier ones. The pitches of tones are relative. For example, the pitch of a male speaker is much lower than that of a female speaker, but when male and female speakers articulate a Tone 1 syllable, both will be correctly perceived. In this study, there are alternative H and L tones in the utterances, which will trigger downstep. In the context of downstep, the subsequent H tone will be lower than the forgoing ones, in which case the later tones will be reset in a lower scale. The L tones are basically lower than the H tones, so the subsequent L tones will be lower than the forgoing ones.

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. The soft-landing model is put forward on the basis of the top line of the pitch contour. In this study, the bottom line is investigated, and it is found that the lowering degrees of the earlier downsteps 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].

Conclusion

In this study, the pitch realization of L tones in the context of downstep in Chinese is analyzed, and it is found that lowering effect exist for L tones, regardless of the length of the utterance, or the position of the tone in the utterance. The low points of L tones always drop to a lower scale compared to the foregoing ones, and due to the effect of final lowering, the lowering degree of the last downstep is always larger than the preceding ones. The utterance initial L tones may shift upwards as the number of downsteps increases, and this effect of great, with those of longer utterances always higher than those of shorter ones. The final L tones are not affected when there are more downsteps, that is, the utterance final pitch tends to be constant. The results from this study will be helpful for the pitch modeling in speech synthesis.

References

[1] Y. O. Laniran and G. N. Clements, "Downstep and high raising: interacting factors in Yoruba tone prodution," 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.