

Downstep and Pitch Range of High-Low Tones in Chinese

Maolin Wang and Wei Xiong, *Member, IACSIT*

Abstract—In this study, the lowering pattern in Mandarin 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. As for pitch range, 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.

Index Terms—Speech, downstep, tone, pitch.

I. INTRODUCTION

This study investigates the effect of downstep, which 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^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. 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].

In tone languages, 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 F_0 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 F_0 value of a preceding tone. Shih [6] pointed out that the F_0 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 Mandarin Chinese utterances display downstep effects across *H* and *L* tones. At the same time, the realization of pitch range will also be probed. The following questions are addressed:

- 1) Do *H* and *L* tone sequences show lowering effects? What about the lowering degree at different position of the utterance?
- 2) Is the initial *H* and *L* tone scaled higher as the number of downsteps increases?
- 3) Does the overall pitch range of the utterances expand as the number of downsteps increases? Are there variation of pitch ranges of the *HL* sequences at different position of the utterance?
- 4) Does the initial *HL* sequence expand as the number of downsteps increases?

II. METHOD

A. 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.

B. Subjects and Recording

The utterances used in this experiment are recorded by eight native speakers of Mandarin Chinese, four males and

Manuscript received October 9, 2013; revised December 12, 2013. 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.

M. Wang and W. Xiong are with the College of Chinese Language and Culture, Jinan University, Guangzhou, 510610, China (e-mail: wmljd@126.com).

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).

C. Measurements

Acoustic data are segmented and labeled, and F_0 is extracted using Praat [8]. The extracted F_0 is manually verified with reference to the cycle in the waveform. In this study, for the purpose of normalizing the F_0 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.

1) Average pitch value of H tone

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.

2) Lowering degree

In order to inspect the extent of the lowering effect, lowering degree is calculated. It is the difference between successive H or L 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 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, as is shown in (3),

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

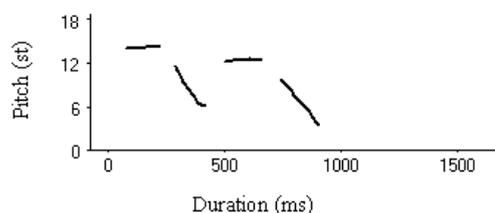
In (3), St_t and St_b refer to the pitch values of the top and bottom points in semitone respectively, and R is the pitch range. Statistic analysis is done in SPSS.

III. RESULT

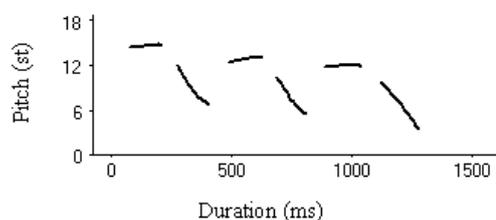
Fig. 1 graphs the mean pitch values of utterances of various lengths for all the eight speakers, with (a) to (d)

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.

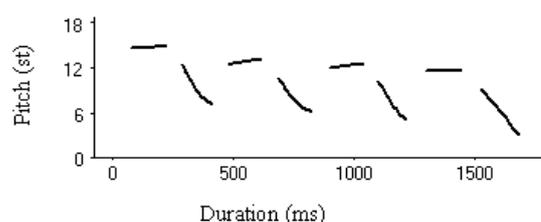
(a) Utterance with 2 HL sequences



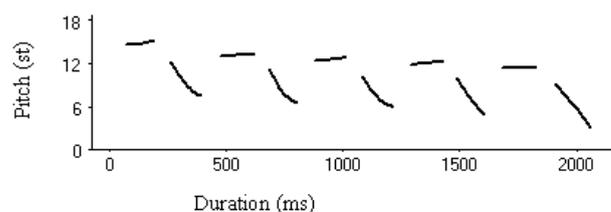
(b) Utterance with 3 HL sequences



(c) Utterance with 4 HL sequences



(d) Utterance with 5 HL sequences



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.

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

A. 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 as follow.

1) 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).

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): $F(1, 95) = 173.7, p < 0.001$.

3) *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.

4) *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.

B. *Lowering Degree*

Table I displays the average lowering degrees (in st) of *H* tones for utterances of various lengths. In Table I, 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 1.73 semitones.

TABLE I: 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	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 lowering degree 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 lowering degrees 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 lowering degrees: $F(2, 190) = 87.2, p < 0.001$. 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 lowering degrees 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 : $F(2, 190) = 1.19, p = 0.305$.

C. *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 II 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 utterances 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 II: AVERAGE PITCH VALUES (IN ST) OF THE INITIAL *H* TONES AND THE ANOVA RESULT

N. of HLs.	2	3	4	5
Pitch value	14.12	14.61	14.72	14.73
ANOVA result	$F(3, 285) = 16.7, p < 0.001$			

D. *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 III 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 III: 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	$F(3, 285) = 7.65, p < 0.001$			

E. *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.

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.

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) 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: $F(1, 95) = 3.03, p = 0.085$.

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: $F(1, 95) = 0.09, p = 0.771$.

F. The Pitch Range of the Initial HL Sequence

In this subsection, the pitch ranges of the utterance initial *HL* sequences are analyzed. Table IV 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 IV: 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	$F(3, 285) = 13.04, p < 0.001$			

IV. DISCUSSION

A. H Tones

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 lowering degree 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.

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.

As far as *H* tones are concerned, 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.

B. Pitch Range

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.

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 lowering' 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.

V. CONCLUSION

In this study, the lowering pattern in Chinese tone production is analyzed, and it is found that downstep exists 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.

In regard to pitch range, 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. 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 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 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 production," *Journal of Phonetics*, vol. 31, pp. 203-250, 2003.
- [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*, Legon, Ghana, vol. 1, 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*, Legon, Ghana, vol. 1, 1965.
- [5] Y. Xu, "Contextual tonal variations in Mandarin," *Journal of Phonetics*, vol. 25, 1997, pp. 61-83.
- [6] C. Shih, "Declination in Mandarin. Intonation: theory, models and applications," in *Proc. the 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*, vol. 32, pp. 57-61, 2007.
- [8] P. Boersma, "Praat, a system doing phonetics by computer," *Glott International*, vol. 5, no. 9-10, pp. 341-345, 2001.
- [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*, vol. 32, pp. 313-348, 2004.



Maolin Wang was born in Hebei, China, in 1965. He received his Ph.D. from the Chinese Academy of Social Sciences in 2003. He is an associate professor in Jinan University, Guangzhou, China. His current research interest is phonetics and speech signal processing.