Downstep and its interaction with focus and boundary in Mandarin Chinese

Bei Wang[1], *Frank Kügler*[2*], *Susanne Genzel*[3]

[1]Minzu University of China, China
[2]Goethe University of Frankfurt, Germany
[3]University of Potsdam, Germany

bjwangbei@qq.com, f.kuegler@gmail.com, sgenzel@uni-potsdam.de

Abstract

It has been observed that in a HLH (High-Low-High) tone sequence, the second H tone is lowered in pitch, as compared to a HHH tone sequence, which was termed as downstep. To calculate the downstep effect and test its scope, we compared sentences in the tonal structure of (HHHHH**H**)(<u>H</u>HHHHHH) to (HHHHH**L**)(<u>H</u>HHHHHH). The syntactic boundary between the L (X) and the H (Y) was either a syllable boundary or a phrase boundary. Moreover, corrective focus was elicited either on the syllable X, Y, sentence final (Z) or none (wide focus). The results showed that: (1) in the wide focus and sentence-final focus (ZF) condition, downstep can last for several syllables; (2) a phrase boundary did not block downstep; (3) when syllable Y was focused, on-focus F0 raising did not override downstep, however, it stopped the trend of downstep in the following H tones; (4) when syllable X was focused, it caused a post-low-bouncing effect on the following H tones instead. Taken together the data showed that downstep is relatively stable and independent.

Index Terms: downstep, focus, boundary strength, intonation

1. Introduction

In many tone languages, the F0 height and contour of a tone are affected by adjacent tones. In a HLH tone sequence, the F0 height of the second H tone is lower than that of the first H, presumably because it is lowered by the preceding L tone [1]. This phenomenon is known as downstep and it has been found in many African tone languages [2-5] and Asian tone languages such as Mandarin [1, 6-8]. It should however be noted that the processes of downstep may not have the same underlying mechanism in different languages. Downstep might be a phonological feature governed by OCP (Obligatory Contour Principle) or a phonetic feature due to articulatory co-articulation.

In Mandarin, a L tone has the lowest minimum F0 and causes much larger downstep effect than a R or F tone (rising and falling tone) [6, 7]. Xu [1] suggested that intonation declination is probably equivalent to a series of downsteps, which is also observed in Akan (Niger-Congo)[4, 5]. However, it is not quite clear from the previous studies how stable downstep effect is in Mandarin. A deep understanding of downstep seems important for understanding intonation declination in general.

When studying pitch movement, we need to take consideration of the maximum speech of F0 shift. As shown in [9], the speed of pitch change varies linearly with the size of pitch change, and it varies also with the direction of pitch change, i.e., pitch rise takes longer time than pitch fall. In Xu's PENTA model[10], it shows that it takes almost the time of a syllable to reach its tonal target. Based on this evidence, downstep might be understood as an articulatory co-articulatory process (carry-over lowering).

In [1], when comparing a sentence with the HHHHH tone sequence with the HLHHH sentence, we can see that the H tones after the L are with lower F0 than the all H tone sentence (see Fig. 4, pp. 66). It indicates that the downstep effect is not just on the adjacent tone, but can last for several syllables. This is, however, not necessarily expected from a purely articulatory co-articulation account.

Our first research question is, whether downstep is constrained by a certain prosodic constituent. Are there any structural or phonological conditions that determine the domain of downstep. The commonly agreed prosodic components above words are prosodic word, prosodic phrase (or phonological phrase), and intonation phrase[11, 12]. We will test whether a boundary of prosodic phrase blocks downstep.

Our second research question concerns focus, because focus causes intonation variation. In Mandarin, focus raises F0 and expands the pitch range in the focused words, and lowers F0 and compresses the pitch range in the post-focus words (PFC) [1, 13-15]. It has been found that the downstep effect applies in the post-focus region [1, 14, 16]. Does on-focus F0 raising override downstep?

Another interesting phenomenon caused by a L tone needs to be mentioned is the post-low bouncing effect as reported in [17-19], i.e., after reaching a very low pitch in a low lexical tone, F0 bounces up and then gradually drops back to the low pitch level in the following syllables. It is like a bouncing ball hitting the ground. This phenomenon was long known for Mandarin that a neutral tone was raised in F0 after a L tone. [19]has

* Corresponding author: Frank Kügler.

reported that post-low bouncing lasts for several neutral tones after a L tone, and occurs in the full tones as well. It has been found that a L tone is lowered when it is under focus[1]. In this case, we expect to see a post-low-bouncing effect instead of downstep in the H tones after the L tone.

2. Experiment

The experiment aimed to study the mechanism of downstep in Mandarin Chinese. Focus and boundary strength were systematically manipulated to investigate the scope of downstep.

2.1 Methods

2.1.1 Reading materials

The general idea was to compare a HL<u>H</u> sequence with a HH<u>H</u> sequence under varied conditions, in which the H tone after the L tone was the target for testing downstep, indicated as underscored. We embedded these tone sequences in a sentence with all H tones except for the sentence final particle, which was a neutral tone. In this way, we can study how long the downstep effect lasts. We used a sentence with the syntactic structure as $SV_1O_1V_2O_2$ (see (1) for an example).

(1) An example set of the reading material (in pinyin). The target words are bold. The boundary between the target syllable X and Y was manipulated as either a syllable (1a) or a phrase (1b) boundary. ω and φ indicate the prosodic word and prosodic phrase boundary. The syllables are all with H tones except "le".

(1a) *Syllable* boundary between syllable X (H 欧 or L 藕) and Y(dou 兜):

((Wangying)ω(qingchu)ω(**<u>yingoudou)</u>**ω) φ
wangying find **X** **Y** (<u>yingoubag</u>)
<u>((</u>lingchu)ω((gongsi)ω (chejian.le)ω))φ.
take-out company **Z**(factory) ASP
'Wangying has found a **<u>bag of yingou</u>** and took it out of the factory of the company'.

(1b) *Phrase* boundary between syllable X (H 欧 or L 藕) and Y(dou 都)

((Wangying)ω(qingchu)ω(**<u>yingou)</u>**ω**<u>)</u>**φ
wangying find **X** (yingou)
(**<u>(dou</u>**lingchu)ω(gongsi)ω (chejian.le)ω)φ.
Y(all) take-out company **Z**(factory) ASP
'Wangying has found **<u>Yingou</u>** and took **<u>all</u>** of them out of the factory of the company'.

Three independent variables were manipulated, that is, tone of syllable X (either H or L tone before the target H tone); boundary (syllable or phrase boundary between syllable X and Y); and focus (wide focus, focus before the boundary (XF), after the boundary (YF), and sentence final (ZF)).

In the condition of the syllable boundary (1a), the HLH tone sequence spanned one word, whereas in the condition of the phrase boundary (1b), the HL tone sequence spanned a word, and the second H tone belonged to a following word that is contained in a verb phrase (VP). The strategy was to use a monosyllabic homophone of the target syllable in these two boundary conditions, as in [10]. In the condition of syllable boundary, the HHH sequence (yin1ou1<u>dou1</u> 樱欧<u>兜</u>) means a bag printed with ying1ou1(a make-up word for an exotic plant), whereas in the phrase boundary condition, 'dou1' in the HHH sequence (ying1ou1.dou1 樱欧都) is an adverb, means "all". In this way, the two boundary conditions were clearly distinguished by using two different characters.

We assumed a stronger prosodic boundary in condition (1b) than in (1a) due to the syntactic structure. According to syntax-prosody interface constraints[12], the VP matches to a prosodic phrase. The left edge of this prosodic phrase constitutes the boundary in question.

Focus was elicited by correction of the corresponding word. Taken the HHH sentence in the syllable boundary (see 1a), the four focus conditions are presented in (2). Here, "dou" is the target syllable to be manipulated as either pre-, on- or post-focal, with a focus either on the X, Y or Z syllables correspondingly. And a wide focus condition served as the baseline. Similar contexts were constructed for the other sentences.

(2) The context sentence of the four focus conditions for sentence: "Wangying.qingchu.**ying ou(X)** **<u>dou(Y).</u>** lingchu. gongsi. Chejian(Z).le."

Wide focus: "Wo yaogaoshu ni yi jian shi." (I want to tell you something.)

X-focus: "bushiying**an**" (It is not "Yingan")(focus on "ou")

Y-focus: "bu2shi4**bao**" (It is not the tote) (focus on "dou ")

Z-focus: "bu2shi4**lou2dao**" (It is not the corridor.) (focus on "chejian")

Two sets of the sentences were constructed. The total number of sentences for analysis were:

2(tone)×4(focus)×2(boundary)×2(sentence set)×3(repetitions) ×9(speakers)=864

2.1.2 Speakers

Nine native Mandarin speakers participated in the experiment at Minzu University of China (6 female and 3 male speakers), from the age of 20 to 28. They were born and brought up in Beijing, spoke no other Chinese dialects and reported no hearing or speaking impairments. They were paid with small amount of money for joining the experiment.

2.1.3 Recording procedure

The subjects were recorded individually in the speech lab at Minzu University of China. They were asked to read aloud both the context and the target sentences at a normal speed and in a natural way. They sat before a computer monitor, on which the test sentences were displayed, using AudiRec, a custom-written recording tool. To make the reading task a little easier for the speakers, the focused words were highlighted with color. A Shure 58 Microphone was placed about 0-5 cm in front of the speaker. All sentences were digitized directly into a Thinkpad computer and saved as WAV files. The sampling rate was 48 KHz and the sampling format was one channel 6-bit linear. Each subject read the sentences three times, once in each session, with about 5-minute breaks between sessions. In each session, all the 32 sentences were randomized, and each subject had a different random order. The total recording time was about an hour.

2.1.4 Acoustic measurement

The target sentences were extracted and saved as separate WAV files. ProsodyPro, a Praat script [20] running under Praat [21], was used to take F0 and duration of each syllable measurements from the target sentences, which were all segmented into syllables manually, and at the same time hand-checked vocal cycles markings generated for errors, such as double-marking and period skipping. ProsodyPro then generated syllable-by-syllable F0 contours that are either time-normalized or in the original time scale. At the same time, the script extracted various measurements, including maximum F0, minimum F0 and duration of each syllable.

2.2 Results

2.2.1 Graphic analysis

First, we look at how focus is encoded in intonation, see Fig. 1 for the HHH (left) and the HLH (right) sentence. In Fig. 1 and the following figures, 10 time-normalized F0 points were taken for each syllable. The contours were averaged of 3 repetitions of the 9 speakers across the two sets of the sentences.

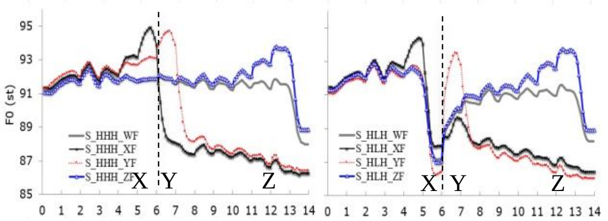

Fig.1 : *The intonation contours of the four focus conditions in the HHH(left) and the HLH (right) sentences under the syllable boundary condition.*

We can see in Fig.1 that focus is reliably realized as the tri-zone pattern, similar to [1, 13, 15]. The on-focus syllable showed raised F0 and expanded pitch range; the post-focus parts exhibited lowered and compressed pitch; and the pre-focus parts are largely intact. Such a pattern holds for both the HHH and HLH sentences. A very similar pattern was found in the sentences with a phrase boundary between X and Y as well.

Second, we will look at how boundary strength is encoded in intonation, see Figure 2. We can see that there is no clear difference in F0 between the two boundary conditions in both the HHH (left) and HLH (right) sentences. In any focus condition, the contours of the phrase (P) or syllable (S) conditions are almost identical, as shown in Fig. 2. To save space, we do not present those figures here. It is in consistence with [15] that the two boundaries are not systematically marked in F0.

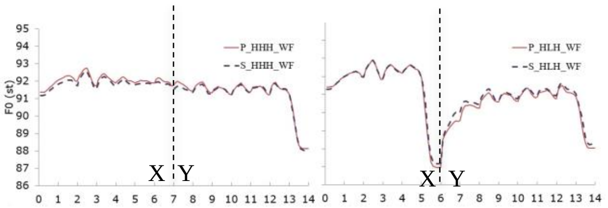

Fig. 2: *The intonation contours of the two boundary conditions in the HHH (left) and the HLH (right) sentences under the wide focus condition. Here P and S stands for phrase and syllable boundary between the X and Y syllable respectively.*

Third, we will study the downstep effect, which is of the most interest to the current study. Fig. 3 (syllable boundary) and Fig. 4 (phrase boundary) presents the direct comparison between the HHH and HLH sentences under different focus conditions.

In Fig. 3 (a), we can see downstep lasts for several H tones after the L tone in the wide focus sentence, which becomes weaker when the H tones are in a longer distance from the L tone. In Fig. 3(d), when the focused word is sentence final, the downstep effect is similar to the wide focus sentence. In Fig. 3(c), when the H tone after the L tone is under focus (Y-focus), the downstep effect shows in the target H tone, i.e. the F0-rise due to focus is clearly pronounced but its H target is undershot compared to the HHH tone sentence. The downstep effect holds in the remaining H tones as well. In Fig. 3(b), when the L tone is under focus (X-focus), instead of downstep, we see the post-L-bouncing effect in the following H tone, as reported in [17-19]. It shows that the H tones after the L tone goes up first, then drops graduate, as compared to the all-H sequence.

In Fig. 4 (a-d), when the boundary between syllable X and Y is a phrase boundary, a very similar pattern can be seen as discovered in Fig. 3. It indicates that the downstep effect is not blocked by a phrase boundary (see Fig. 4(a, d)), neither is post-L-bouncing effect (see Fig. 4(b)).

We need to notice that the boundary effect on downstep can be seen when syllable Y is focused, i.e., downstep after Y holds when XY is a word (Fig. 3c), but disappears when there is a phrase boundary between X and Y (Fig 4c).

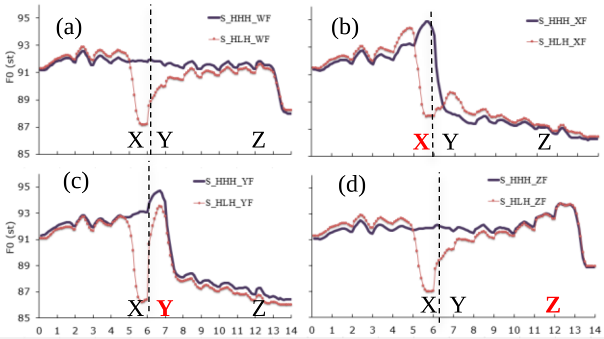

Fig.3: *The comparison between the HHH and HLH sentences in four focus conditions (a-d) under the condition that the boundary between X and Y is a syllable boundary, as indicated by the vertical line. The red bold letter of X, Y and Z is the focus.*

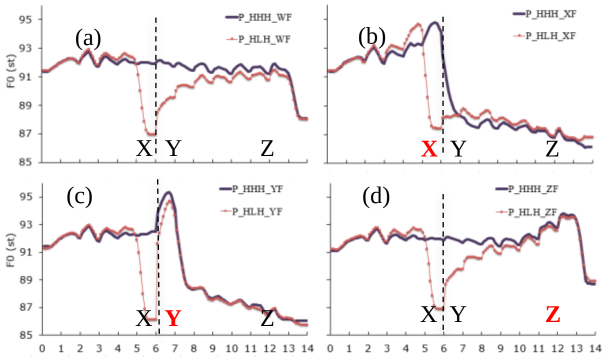

Fig. 4: *The comparison between the HHH and HLH sentences in four focus conditions (a-d) the boundary between X and Y is a syllable boundary, as indicated by the vertical line.*

2.2.2 F0 analysis

Since the encoding of focus and boundary is well analyzed in [15], and the current results are quite similar (see Fig. 1 and 2), we will not analyze that further. The main questions to be tested are: (1) How far does the downstep effect last? (2) Is the downstep effect interfered by focus? (3) Is the downstep effect blocked by a phrase boundary?

We calculated the downstep effect by comparing the HLH sentence with the HHH sentence under the same condition.

Table 1 and 2 present maximal F0 of the H tones after the L tone (from the 1st to the 5th H tone), the corresponding H tones in the HHH sequence and their difference (the size of downstep effect) in the syllable-boundary sentence and phrase-boundary sentence respectively. The F0 values are in semitones ($st = 12 \log_2[F_0]$). There is no downstep effect in the 6h and 7th H tones, except for the wide focus condition in the phrase boundary sentences. Paired-sample T tests were applied in each syllable to test whether the difference of the H tones between the HHH and HLH sentences reached statistic difference at the level of $p<0.05$, which is marked as * in Table 1 and 2.

Table 1: *Maximum F0 of the 1st to the 5th H tone after the L tone in the HLH and HHH sequence and their difference (st) in the condition of syllable boundary between the L and the 1st H.*

		1st	2nd	3rd	4th	5th
Wide Focus	HHH	92.0	91.8	91.8	91.6	91.8
	HLH	90. 1	90.8	91.2	91.1	91.4
	DS	1.9*	1.0*	0.6*	0.5	0.4
Z-Focus	HHH	92.3	92.2	92.3	92.1	92.7
	HLH	90.4	91.2	91.6	91.6	92.4
	DS	1.9*	1.0*	0.7*	0.5*	0.4*
Y-Focus	HHH	94.7	92.4	88.3	87.7	87.6
	HLH	93.5	91.8	88.1	87.5	87.4
	DS	1.2*	0.6*	0.2	0.2	0.2
- Focus	HHH	92.5	88.0	88.0	87.7	87.6
	HLH	89.8	89.3	88.5	88.0	87.8
	PLB	2.7*	-1.3*	-0.5*	-0.3*	-0.2

Note: DS stands for the downstep effect, and PLB stands for post-low-bouncing effect.

Table 2: *Maximum F0 of the 1st to the 5th H tone after the L tone in the HLH and HHH sequence and their difference (st) in the condition of phrase boundary between the L and the 1st H.*

		1st	2nd	3rd	4th	5th
Wide Focus	HHH	92. 1	91.9	91.9	91.7	91.9
	HLH	89.6	90.4	91.0	90.9	91.2
	DS	2.5*	1.5*	0.9*	0.8*	0.7*
Z-Focus	HHH	92. 1	92.0	92.1	92.0	92.5
	HLH	90.0	90.9	911.5	91.5	92.2
	DS	2.1*	1.1*	0.6*	0.4*	0.2
Y-Focus	HHH	95.3	93.4	88.6	87.9	87.6
	HLH	94.7	93.1	88.6	87.7	87.5
	DS	0.7*	0.3*	0.0	0.2	0.1
X-Focus	HHH	92.7	88.2	88.1	87.8	87.7
	HLH	88.7	88.8	88.7	88.2	88.0
	PLB	4.0*	-0.6*	-0.6*	-0.4*	-0.3

First, we will analyze the sentences with a syllable boundary between the HLH and the HHH sentences (see Table 1). We can see in Table 1 that in the wide focus condition, downstep applies up to the third H tone after a L tone with declines in the effect size. When focus is sentence-final (Z-focus condition), downstep expands to the syllable before the final focus. When focus is in the H tone right after the L tone (Y-focus condition), downstep effect still applies and lasts until the second H tone after the L, although F0 of the H tone raised due to focus tone, (see Fig. 2). When focus is in the L tone (X-focus condition), as expected, we see the post-low bouncing effect, instead of downstep, i.e., the H tones after the L tones are with higher pitch in the all-H sequence than in the HLH sequence.

In Table 2, the boundary between the L and the H is a phrase boundary, we see a very similar pattern as in Table 1. It means that downstep is not blocked by a phrase boundary.

3. Discussion and Conclusions

The downstep effect in Mandarin was analyzed with the consideration of its interaction with focus and boundary strength. The results showed that: (1) in the wide focus and sentence-final focus (ZF) condition, downstep can last for several syllables; (2) a phrase boundary did not block downstep; (3) when syllable Y was focused, on-focus F0 raising did not override downstep, however, it stopped the trend of downstep in the following H tones; (4) when syllable X was focused, it caused a post-low-bouncing effect on the following H tones instead;

In [1], it is not well studied whether downstep is a phonetic or a phonological effect. In both cases a low target point of a phonological L tone causes a lowering of F0 on a following high tone. This would count as tonal coarticulation (phonetic) if the following H tone does not impose a new register reference line for following tones, if the downstep achieves its earlier high tone level at following syllables. [7] termed this phonetic effect 'downdrift'. Downstep would count as phonological according to [6,7], if the downstepped tone imposes a new register line for following tones as in Akan [3-5] or in Mambila [6]. In this case, the earlier high tone level is not achieved again. More experiments are required to find out why it takes some time to hit back to the topline.

In the present data, we observe that the H tones reach its earlier H tone level again, though only after a few syllables. The case that downstep does not impose a new pitch register line relative to which following tones are scaled would point to the fact that downstep is phonetic in Mandarin [22, 23]. However, at the same time, structural and phonological factors that otherwise would interrupt phonetic tonal coarticulation do not affect downstep in Mandarin. This points to the fact that Mandarin downstep could be interpreted as a phonological effect. This finding thus adds to [22, 23] classification suggesting that the interplay of phonological factors with downstep can shed light on its nature.

In conclusion, the data showed that downstep in Mandarin is relatively stable and independent.

Acknowledgements

We thank Xiaxia Zhang and Qian Wu for running the experiment and labeling the speech data. The work was supported by a DFG fund for the SFB 632, project D5 to the second author and the "First class universities and Top Disciplines of Minzu University" project to the first author.

References

[1] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *Journal of Phonetics,* vol. 27, pp. 55-105, 1999.

[2] L. M. Hyman, "The role of consonant types in natural tonal assimilations," in *Consonant Types and Tone*, L. M. Hyman, Ed., ed Los Angeles: University of Southern California, 1973, pp. 151-179.

[3] F. Kügler, "Tone and intonation in Akan," in *Intonation in African Tone Languages*, L. Downing and A. Rialland, Eds., Mouton de Gruyter, 2017, pp. 89-129.

[4] S.Genzel and F.Kügler., "Phonetic realization of automatic (downdrift) and non-automatic downstep in Akan," presented at the Proceedings of the XVII ICPhS, Hong Kong, 2011.

[5] S. Genzel, "Lexical and post-lexical tones in Akan," PhD Thesis., Human-wissenschaftliche Fakultät, Universität Potsdam, Potsdam, 2003.

[6] C. L. Shih, "Tone and intonation in Mandarin," *Working Papers, Cornell Phonetics Laboratory,* vol. 3, pp. 83-109, 1988.

[7] C. L. Shih, "Declination in Mandarin, Intonation: Theory, models and applications," in *ESCA Workshop*, Athens, Greece, 1997, pp. 293-296.

[8] Y. Chen, "Tonal Variation," in *The Oxford handbook of laboratory phonology*, C. A. Cohn, C. Fougeron, and M. K. Huffman, Eds., ed: Oxford University Press, 2002, pp. 03-4.

[9] Y. Xu and X. Sun, "Maximum speed of pitch change and how it may relate to speech," *The Journal of the Acoustical Society of America,* vol. , p. 399, 2002.

[10] Y. Xu and Q. E. Wang, "Pitch targets and their realization: Evidence from Mandarin Chinese," *Speech Communication*, vol. 33, pp. 39-337, 2001

[11] M. Nespor and I. Vogel, *Prosodic Phonology*, Foris, Dordrecht, Holland, 1986.

[12] Selkirk, E. "The Syntax-Phonology Interface," In J. Goldsmith, et al. Eds: The Handbook of Phonological Theory, Chichester: Wiley, 435–484.

[13] Y.-Y. Chen and C. Gussenhoven, "Emphasis and tonal implementation in Standard Chinese," *Journal of Phonetics,* vol. 36, pp. 724-746, 2008.

[14] B. Wang and Y. Xu, "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese," *Journal of Phonetics,* vol. 39, pp. 595-611, 2011.

[15] B. Wang, Y. Xu, and Q. F. Ding, "Interactive prosodic marking of focus, boundary and newness in Mandarin," *Phonetica,* vol. 75, pp. 24-56, 2017.

[16] S. Genzel, S. Ishihara, and B. Surányi, "The prosodic expression of focus, contrast and givenness: A production study of Hungarian," *Lingua*, vol. 65, pp. 83-204, 2004.

[17] Y. Xu, "Contextual tonal variations in Mandarin," *Journal of Phonetics,* vol. 25, pp. 61-83, 1997.

[18] S. Prom-On, F. Liu, and Y. Xu, "Post-low bouncing in Mandarin Chinese: acoustic analysis and computational modeling," *Journal of the Acoustical Society of America,* vol. 132, p. 421, 2002.

[19] Y.-Y. Chen and Y. Xu, "Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese," *Phonetica,* vol. 63, pp. 47-75, 2006.

[20] Y. Xu, "ProsodyPro — A tool for large-scale systematic prosody analysis," presented at the Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 203), Aix-en-Provence, France, 2013.

[21] P. Boersma, & Weenink, D., "http://www.fon.hum.uva.nl/praat/," ed, 2013.

[22] Connell, B. "Tone and intonation in Mambila," in *Intonation in African Tone Languages*, L. Downing and A. Rialland, Eds., Mouton de Gruyter, 2017, pp. 131-166.

[23] Connell, B. "Downstep," In M. van Oorstendorp et al (eds.) *Companion to Phonology*. Oxford: Blackwell Publishing, 824–847.