Downstep Effect on Disyllabic words of Citation Forms in Standard Chinese

XIONG Ziyu
Institute of linguistics, Chinese Academy of Social Sciences
xiongziyu@163.com

Abstract*

The purpose of this study is to specifically analyze the pitch contours of disyllabic words in Standard Chinese. Based on a huge speech corpus, the paper investigates the downstep effect on the second syllable with a zero initial or a voiced initial in disyllabic word of citation form. Statistical results show that when the tonal pitch features of a disyllabic word constitute a “.H...L..H...” tonal sequence, the H tone behind the L tone will be lowered contrasted to the preceding H tone obviously, and such phenomenon is called downstep. It is also shown in this paper that downstep happened on disyllabic words in Standard Chinese has relatively stable pitch manifestation in different kinds of tonal combinations. The results found in this paper can be applied to generate pitch contours properly to improve the naturalness of the speech synthesis.

1. Introduction

According to Autosegmental-Metrical (AM) theory, the pitch contour of any language is linearly composed by its local pitch events and pitch transitions by some rules (Ladd 1996). In English, the most important local pitch event is pitch accent, while in Standard Chinese the most important local pitch event is syllabic tone. In AM theory, local pitch event is transcribed by pitch features (e.g. “H”, “L” and its combination forms. For example, Pierrehumbert (1980) summed up pitch accent of English as seven types: H’, L’, L+H’, L+H, H+L’, H+L and H+H. The pitch realization in connected speech is not only controlled by the underlying pitch features, but also influenced by some pitch modification effects. Ladd (1983) put forward three types of pitch modification effects on intonational peaks: [± delayed], [± downstepped] and [± raised]. Thus, to generate pitch contours properly for speech synthesis, it is necessary to investigate the pitch modification effects of a language, as well as its local pitch events.

This paper will only focus on the downstep phenomenon in connected speech. Downstep formally refers to an evident pitch downdrift from a preceding H tone to the latter H tone in a phrase. There are probably many underlying factors that can lead to downstep, for example, it has been found that the pitch of the latter H tone in “.H...L..H...” tonal sequence is obviously lower than that of the previous H tone in some African tone languages (Stewart 1983). Such phenomenon not only appears in African tone language, but also in some non-tonal languages such as English (Pierrehumbert 1980) and Japanese (Pierrehumbert & Beckman 1988). Therefore, it is considered as a very common pitch changing phenomenon.

Table 1: Sample number of 16 kinds of different tonal arrays

<table>
<thead>
<tr>
<th>Tonal arrays</th>
<th>Sample number</th>
<th>Tonal arrays</th>
<th>Sample number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone1+Tone1</td>
<td>69</td>
<td>Tone1+Tone2</td>
<td>114</td>
</tr>
<tr>
<td>Tone1+Tone3</td>
<td>55</td>
<td>Tone1+Tone4</td>
<td>107</td>
</tr>
<tr>
<td>Tone2+Tone1</td>
<td>56</td>
<td>Tone2+Tone2</td>
<td>55</td>
</tr>
<tr>
<td>Tone2+Tone3</td>
<td>49</td>
<td>Tone2+Tone4</td>
<td>103</td>
</tr>
<tr>
<td>Tone3+Tone1</td>
<td>34</td>
<td>Tone3+Tone2</td>
<td>39</td>
</tr>
<tr>
<td>Tone3+Tone3</td>
<td>31</td>
<td>Tone3+Tone4</td>
<td>72</td>
</tr>
<tr>
<td>Tone4+Tone1</td>
<td>70</td>
<td>Tone4+Tone2</td>
<td>73</td>
</tr>
<tr>
<td>Tone4+Tone3</td>
<td>77</td>
<td>Tone4+Tone4</td>
<td>135</td>
</tr>
</tbody>
</table>

The statistical data in this study is taken from a speech corpus of disyllabic words, which was established by Institute of Linguistics, Chinese Academy of Social Sciences. The materials in this corpus were read by a male adult who speaks Standard Chinese. Only 1139 disyllabic words in which the second syllable consists of a zero initial or a voiced initial were selected in this study to be analyzed. The description of sample number of different tonal arrays is in table 1.

The procedures of data analysis are: 1. The syllables and tones of every disyllabic word were annotated by PRAAT program (http://www.fon.hum.uva.nl/praat/); 2. The pitch contour of each disyllabic word was analyzed by PRAAT program and then corrected by hand in order to confirm the accuracy of the pitch values; 3. Based on the annotated data of syllables and the corrected fundamental frequency (F0) data, ten points F0 values were extracted from each syllable by equal interval of duration; 4. Based on the classified data according to the different tonal combinations, the average value of each point was figured out by SPSS program (http://www.spss.com).

*This research is partially supported by National Natural Science Foundation of China (No. 60475043) and partially supported by Younger Researcher Starting Foundation of Chinese Academy of Social Sciences.
The pitch contours in Figure 1.1 to Figure 1.8 are formed based on the final average values.

**Figure 1**: The average pitch contours of disyllabic words pronounced by speaker X. In these figures, “1+1” represents “Tone1+Tone1”, “1+2” represents “Tone1+Tone2”, and so on.

It's clearly showed from those data above: (1) No matter which kind of tone follows, there will be no notable pitch changes at the beginning of the tone on the first syllable. This can be observed by the pitch values at the beginning of the first syllable in Fig.1.1, Fig.1.3, Fig.1.5 and Fig.1.7. (2) No matter which kind of tone precedes, there will be no remarkable pitch changes at the end of the tone on the second syllable. This can be seen from the pitch values at the end of the second syllable in Fig.1.2, Fig.1.4, Fig.1.6 and Fig.1.8. (3) There is a phase of pitch transition between the end of the first tone and the beginning of the second tone, which is roughly shown by the dashed lines in the above pictures. The duration of some kinds pitch transition can be accurately predicted based on some factors (Xiong, 2004b). From the statistical data, the pitch contours in pitch transition are obviously different among tonal combinations. But if ignoring the pitch transition, we can find: a) the following tone will not effectively affect the pitch values at the end of the first tone. This can be checked by observing the final pitch values of the first tone just before the left dashed line in Fig.1.1, Fig.1.3, Fig.1.5 and Fig.1.7; b) the preceding tone will not effectively affect the pitch values at the beginning of the second tone. This can be checked by observing the initial pitch values of the second tone just behind the right dashed line in Fig.1.2, Fig.1.4, Fig.1.6 and Fig.1.8. (When the tone of the second syllable is Tone1 or Tone4, the pitch realization at the beginning is obviously different, which will be explained in the later part.) All in all, if the pitch transition can be ignored, two tones have little influence on each other when the disyllabic words are in their citation forms. In addition, according to these data, it is confirmed that a speaker can accurately control the phonetic realization of target pitch value; otherwise the pitch of the data he presents wouldn’t be so steady.

Except those comparatively unanimous pitch realizations mentioned above, there are some which are not agreeable can also be found. They are summed up in the following several aspects: Phenomenon (1), according to Fig.1.2, when a syllable
with Tone1 is after another syllable with Tone4, the pitch of the Tone1 is lowered on the whole; Phenomenon (2), according to Fig.1.8, when a syllable with Tone4 is after another syllable with Tone4, the maximal pitch value of the second syllable is lowered; Phenomenon (3), according to Fig.1.4, when a syllable with Tone2 is following another syllable with Tone1, Tone2, or Tone4, the final pitch of the second syllable is lowered; Phenomenon (4), comparing the pitch range of the second syllable in Fig.1.7 with that in Fig.1.5, Fig.1.3 and Fig.1.1, it is found that the pitch range of the second syllable is obviously narrowed when the second syllable is following another syllable with Tone4.

The pitch contours in Fig.1.1 to Fig.1.8 are based on dozens of samples, so it can be affirmed that those inconsistent phenomena mentioned above are regular, and there should be some reasons behind it. Then how to explain these regular nonconformity phenomena? This paper will go on investigating these regular nonconformity phenomena, attempting to find out the reasons behind.

### 3. Downstep Effect

As mentioned above, in the "..H..L..H.." tonal sequence of Standard Chinese, the latter high tone significantly lower than the former high one. Such phenomenon is called "Downstep" by some researchers. Downstep phenomenon appears not only in African tone language, but also in some non-tonal languages, such as Japanese and English. Therefore, it is considered to be a very common pitch modification effect. There are two basic meanings in the "downstep" caused by "..H..L..H.." tonal array: (1) Only when a high tone (H) and low tone (L) combine to form a "..H..L.." array, the downstep on the following H tone is possible. Only a single "H" or a single "L" will definitely not lead to such downstep. (2) Downstep effect only acts on the H tone after "..H..L.." tonal array, but will not exert an influence on the following L tone.

Standard Chinese is a syllabic tone language. Does such downstep phenomenon exist in it? The 16 kinds of four tones’ combinations and their tonal pitch features in Standard Chinese are described as in Table 2.

**Table 2: The description of four tones’ combinations and their tonal pitch features in Standard Chinese**

<table>
<thead>
<tr>
<th>Tone combination</th>
<th>Features of combination</th>
<th>Tone combination</th>
<th>Features of combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone1+Tone1</td>
<td>H--</td>
<td>Tone1+Tone4</td>
<td>H-L</td>
</tr>
<tr>
<td>Tone1+Tone3</td>
<td>H-L-</td>
<td>Tone1+Tone4</td>
<td>H-L</td>
</tr>
<tr>
<td>Tone2+Tone1</td>
<td>LH--</td>
<td>Tone2+Tone2</td>
<td>H-L</td>
</tr>
<tr>
<td>Tone2+Tone3</td>
<td>LH-L</td>
<td>Tone2+Tone4</td>
<td>LH-L</td>
</tr>
<tr>
<td>Tone3+Tone1</td>
<td>L-H</td>
<td>Tone3+Tone2</td>
<td>L-H</td>
</tr>
<tr>
<td>Tone3+Tone3</td>
<td>L-HL</td>
<td>Tone3+Tone4</td>
<td>L-HL</td>
</tr>
<tr>
<td>Tone4+Tone1</td>
<td>H-L</td>
<td>Tone4+Tone2</td>
<td>H-L</td>
</tr>
<tr>
<td>Tone4+Tone3</td>
<td>HL--</td>
<td>Tone4+Tone4</td>
<td>H-LL</td>
</tr>
</tbody>
</table>

It is found that among the 16 kinds of four tones’ combinations in Standard Chinese, there are five kinds of combinations containing "..H..L..H.." tonal sequence. They are marked by square in the above Table 2. If there is downstep phenomenon in "..H..L..H.." tonal sequence in Standard Chinese, it must happen and only can happen on the five combinations. With the data in Fig.1.1 to Fig.1.8, it is showed that there are downstep phenomena happening in the five combinations without exception, which causes the latter H tone lowered in "..H..L..H.." tonal sequence obviously.

There is downstep phenomenon of "..H..L..H.." tonal sequence in Standard Chinese. This can be used for explaining four kinds of regular nonconformity phenomena mentioned above (in the second part of this paper): the downstep happening on "Tone4+Tone1" combination can explain the phenomenon 1; the downstep happening on "Tone4+Tone4" combination can explain the phenomenon 2; the downsteps happening on "Tone1 + Tone2", "Tone2 + Tone2" and "Tone4 + Tone2" combinations can explain the phenomenon 3; the downsteps happening on "Tone4 + Tone1", "Tone4 + Tone4" and "Tone4 + Tone2" combinations cause the pitch values of all the H tones of the second syllable to reduce. Then this leads to the pitch range of the second syllable behind Tone4 obviously narrowed. Thus the phenomenon 4 is explained. Also with the data in Fig.1, there is no downstep phenomenon happening in other combinations except that five kinds tonal combinations. To sum up, by employing the concept of "downstep effect", we can clearly explain the four kinds of regular nonconformity phenomena mentioned previously. Another related issue is that downstep effect can also influence the global pitch trend of an utterance in Standard Chinese, which was reported by Xu (1999, 2001).

### 4. The Power of Downstep Effect

On the basis of the argumentation of the previous paragraphs, under the function of downstep effect, the pitch value of a H tone will be notably reduced. Hence, the paper proposes that the actual pitch realization of one pitch position in Standard Chinese = the pitch value of tonal pitch feature at this pitch position + [the pitch value of various kinds of pitch modification effects acting on this tonal pitch feature] + random errors. It can be expressed by the formula:

\[ P = P_c + (ΣP_{m}) + k \]  

(1)

In which, \( P \) represents the actual pitch realization of a certain pitch position under some conditions; \( P_c \) represents the pitch value of this tonal pitch feature; \( ΣP_{m} \) represents the integrated power of various pitch modification effects acting on this tonal pitch feature; \( k \) represents random errors; \([ ] \) means pitch modification effect is optional, not always appears at every pitch position.

The formula (1) has the following meanings: 1. the actual pitch value of certain pitch position can be the result of many factors, not only its tonal pitch feature such as H or L, but also some pitch modification effects. Xiong (2004a) reported the influence of syllabic tone on its pitch realization will be reduced when more and more pitch modification effects intervene; 2. The power of various modification effects acting on the pitch value of a tonal pitch feature will not exert any influence on its essence. Modification effects and tonal pitch feature can be theoretically separated exactly; 3. In different phonological situations, the type and number of the modification effect acting on a tonal pitch feature are variable. But the same pitch modification effect, in whatever phonetics situations, has the same power; 4. On the premise of ignoring random error, so long as the tonal pitch feature and the pitch modification effects on a pitch position are the same, then \( P \) keeps the same value. But the contrary is not true; 5. The pitch modification effects and random errors need to be distinguished,
the former has the steady and predictable modifying effect on the pitch value of a tonal pitch feature under certain conditions, and it needs to be studied systematically.

In order to investigate the modification effect acting on a tonal pitch feature, the formula (1) can be rewritten as formula (2), on the premise of ignoring random errors.

\[ \Sigma P_m = P - P_c \]  

(2)

To describe pitch modification effects, the sign "+" or "- " can be used to represent whether it is possessing a certain modification effect on a tonal pitch feature. For example, "[H] [+downstepped]" represents a downstepped H tone; "[H] [-downstepped]" represents a non-downstepped H tone. The power of downstep can be shown in the following picture:

\[ P_1 \text{[H]} [-\text{downstep}] \ldots \]

\[ P_2 \text{[H]} [+\text{downstep}] \ldots \]

*Figure 2*: a simple sketch to show the power of downstep effect.

If downstep effect can be regarded as the reason leading to the difference between \( P_1 \) and \( P_2 \) in the above figure, the power of downstep effect can be calculated out by computing the difference between \( P_1 \) and \( P_2 \).

According to the Fig.1, three groups of data correlated with downstep effect can be gotten: (1) Comparing with the Tone4 after the other three kinds of tones, the H tone of the Tone4 after a Tone4 drops from 152 Hz to about 134 Hz, the decreasing amount is about 18 Hz, and it’s about 88% (134/152) of the original pitch value. (2) Comparing with the Tone2 after a Tone3 tone, the H tone of the Tone2 after Tone1, Tone2 or Tone4 drops from 143 Hz to about 126 Hz, the decreasing amount is 17 Hz, and it’s about 88% (126/143) of the original pitch value. (3) Comparing with the Tone1 after the other three kinds of tones, the H tone of the Tone1 after a Tone4 drops from 138 Hz to about 122 Hz, the decreasing amount is about 16 Hz, and it is about 88% (122/138) of the original pitch value. This indicates that the power of downstep effect is very stable and predictable in Standard Chinese. The pitch value of downstepped H tone is about 88% of the original one, it equals to the value which is lowered 2 semitone from the original pitch value. It is widely accepted that the height of H-tone in stepping contours can be successfully predicted by using a constant ratio of decay from the previous peak's value in language like English (Liberman, M. and J. Pierrehumbert 1984 ), Dutch (R. van den Berg et al. 1992).

5. Conclusions

In Standard Chinese, there are 16 kinds of tonal combinations in disyllabic words, in which 5 kinds of tonal combinations consist of “...H..H..” sequence, namely, “Tone1+ Tone2”, “Tone2+Tone2”, “Tone4+Tone1”, “Tone4+Tone2” and “Tone4+Tone4”. The research result of this paper shows: in their citation form, all these five kinds of disyllabic words, in which the latter word consists of a zero initial or a voiced initial, are affected by downstep effect without exception, and the pitch value of downstepped H tone is about 88% of the original one. This result indicates that downstep effect can explain some regular unusual phenomena in the pitch contours of disyllabic words conveniently and accurately, and can be used to generate accurately pitch contours for speech synthesis.

6. References