TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 17/22 pp533-539
Volume 13, Number 4, August 2008

Tone-3 Accent Realization in Short Chinese Sentences*

CAO Wen (曹 文)**, ZHANG Jinsong (张劲松)†

Key Research Institute in University, Center for Studies of Chinese as a Second Language, Beijing Language and Culture University, Beijing 100083, China;
† National Institute of Information and Communications Technology, ATR Spoken Language Translation Research Laboratories, 2-2-2 Hikaridai, Keihanna Science City 619-0288, Japan

Abstract: To investigate how a low tone (tone-3, T3) syllable in Chinese can be perceived to be focal accented or not, a total of 156 sentences containing tone-3 words were synthesized and used as stimuli in a perceptual study. The sentences differed in the falling value between the two high pitches, and in the duration and phonation types of the T3 syllables. Thirty-nine subjects were asked to judge where the focus or accent was for each sentence. The results show that at least three degrees of pitch drop are involved in the focus recognition: a big sized drop of about 10 semitones; a middle sized drop of about 6 semitones; a small sized drop of about 2 semitones. The results suggest that the three sizes of pitch drop have different indications in Chinese intonation, depending on both the tone and the tone combination. In perception, there are various ways to realize tone-3 focus in the T*x*-T3-T*y* sentences series, but in production or for text-to-speech synthesis, the rule simply is making a middle sized pitch drop with a long and creaky T3 syllable. Similarly, to focus on the low tone syllable in the T3-T*x*-T*y* sentences, a creaky T3 syllable is essential. However, a long T3 syllable is a strong determinant for a low tone focus in the T*x*-T*y*-T3 sentences.

Key words: accent; focus; tone-3 in Chinese; pitch drop; duration lengthening; creaky voice; downstep

Introduction

Researchers have found in many languages, including Japanese, English, Swedish, French, Dutch, Italian, and Russian, that a large fall of the top line key of the sentential pitch contour is an important characteristic of a focal accent[1-5]. This general phenomenon has also been observed in Chinese as well[6-9]. However, further studies are necessary because in standard Chinese there are four tones—from tone-1 (T1) to tone-4 (T4), which can be phonologically described as /H/, /LH/, /L/, and /HL/, respectively. In general, it is easy for a syllable which has the tonal feature [H], such as /H/ (tone-1, T1), /LH/ (tone-2, T2), and /HL/ (tone-4, T4) to realize a high pitch accent or focus (H*). The common way to do this is to make a significant pitch (f 0) drop between the intending accent [H] and the following, by raising [H] of the focused syllable, by lowering [H] of the post-focused syllable, or by using both. But what about the low tone, e.g., tone-3? Without [H], how can a focal accent be realized for this tone in Chinese? Chao reports that the tone will be lower if emphasized[6]. Shen holds that the post-focal syllable is raised when tone-3 is focused[7], whereas Xu suggests a general post-focus lowering to indicate a focal accent[8]. In some early research of Cao on Chinese intonation[9], it was found that whether a tone-3 was accented or not depended on dropped pitch values (D-values) between the [H]s around the /L/ syllable. Zheng reported that a prominent tone-3 syllable even involved a voice quality change[10]. Additionally, some

Received: 2007-09-10; revised: 2008-03-15
* Supported by the National Office for Teaching Chinese as a Foreign Language (No. HBK01-05/17)
** To whom correspondence should be addressed.
E-mail: tsao@blcu.edu.cn; Tel: 86-10-82303813

other studies have shown that a long duration played an important role in the focusing of a tone-3 syllable[11,12].

In most of the studies mentioned above, the tone-3 syllables were arranged in the middle of sentences. However, low tone syllables located in different parts of a sentence may have different acoustic correlations. It is well known that when a tone-3 syllable ends a neutral sentence, or when it is narrow-focused at the end, it will have a rising tail, described phonologically as /L+H/ or /214/ in the 5-level tone system[6,13,14]. Several years ago, Ling noted that when a sentence beginning tone-3 word was accented, the second next [H] would be lowered[14].

Therefore, several different factors have been reported to be related with focus realization in Chinese intonation, including dropped pitch value, duration, and voice quality. How then do these factors work together assuming that they exist? Most previous research has examined this question from the perspective of speech production. In this study we consider this question instead by concentrating on the issue of L* (L* means the low tone focus in intonational phonology) recognition.

1 Method

1.1 Stimuli

Twelve three-syllabic sentences, shown in Table 1, are used as the bases of the stimuli for the study. All the sentences have the same syntax structure: subject-verb-object (SVO). In Group 1, all verbs are of /L/(T3), while the subjects and objects are varied with respect to lexical tone types to give all of the nine combinations of T*x*-T3-T*y*, except T3-T3-T3 because of the tone sandhi rule: T3-T3→T2-T3. In Group 2, a T3-T1-T2(/L/-/H/-/LH/) sentence and a T3-T4-T4(/L/-/HL/-/HL/) sentence are designed in accordance with the samples of Ling's[14]. Because there are extensive agreements on the description of T3 ending focus[6,13,14], we design only one sentence with T3 as the final syllable in Group 3.

Table 1 Sentences for synthesis

Group	Tone combination	Sentence
1	T1-T3-T1 (/H/-/L/-/H/)	Gu1 mai3 zhong1. (Aunt buys (a) clock.)
	T1-T3-T2 (/H/-/L/-/LH/)	Yi1 gei3 qian2. (She offers money.)
	T1-T3-T4 (/H/-/L/-/HL/)	Zhang1 chao3 mian4. ((Mr/Ms) Zhang cooks noodles.)
	T2-T3-T1 (/LH/-/L/-/H/)	Yuan2 mai3 dao1. ((Mr/Ms) Yuan buys (a) knife.)
	T2-T3-T2 (/LH/-/L/-/LH/)	Yi2 gei3 qian2. (Aunt offers money.)
	T2-T3-T4 (/LH/-/L/-/HL/)	Yi2 xie3 xin4. (Aunt writes (a) letter.)
	T4-T3-T1 (/HL/-/L/-/H/)	Gu4 mai3 zhong1. ((Mr/Ms) Gu buys (a) clock.)
	T4-T3-T2 (/HL/-/L/-/LH/)	Li4 jian3 qi2. ((Mr/Ms) Li picks up (a) chess.)
	T4-T3-T4 (/HL/-/L/-/HL/)	Mei4 xie3 xin4. (Younger sister writes (a) letter.)
2	T3-T1-T2 (/L/-/H/-/LH/)	Wo3 he1 cha2. (I drink tea.)
	T3-T4-T4 (/L/-/HL/-/HL/)	Jie3 dai4 cai4. (Elder sister brings dishes.)
3	T4-T2-T3 (/HL/-/LH/-/L+H/)	Mei4 qi2 ma3. (Younger sister rides (a) horse.)

With the help of Praat [http://fonsg3.let.uva.nl/praat/], each sentence in Table 1 was phonetically synthesized using a cut-and-paste approach from recorded dialogue data containing more than 8000 utterances. Two patterns were used: one has a modal tone-3 syllable; the other has the creaky kind. The length of each syllable was then normalized to a duration of 250 ms. Figure 1 shows two examples.

The dialogue data were read and recorded by a female Beijing native, aged 25. The recordings were carried out in a quiet environment, using a Maya 5.1 USB audio fast track sound card and Somic earphone, connected to a laptop computer (Toshiba Satellite 2410).

The number of the sentences for the stimuli study was firstly doubled by adjusting the duration of the tone-3 syllables. In one set, the length was unchanged, e.g., 250 ms; in the other set, a length of 375 ms (=250 ms×1.5) was used. The stimuli of Groups 1 and 2 were then further re-synthesized by changing the height of the second peak of each sentence to form sentences with different D-values: 2 semitones, 6 semitones, and 10 semitones. 100 Hz is the reference