

Research on the Acoustic Realization of Focus in Mandarin

Qi Gong

(College of Foreign Studies, Jinan University, China)

Abstract:- The spectral emphasis value and the effect of focus on it for bi-syllabic words are investigated in Mandarin in this paper. It is found that, the spectral emphasis value of the onset is larger than that of the rhyme. The spectral emphasis value of the onset of the first syllable is larger than that of the second under unfocused condition, due to intervocalic voicing. The emphasis degree of the first syllable is larger than that of the second under focused condition, and that of the rhyme is larger than that of the onset.

Keywords:- Focus, intensity, onset, rhyme, spectral emphasis

I. INTRODUCTION

This experiment deals with the acoustic realization of focus in utterances Mandarin. Focus refers to the specific part in an utterance that expresses the centre of attention in speech interaction. It is usually the part that the speaker intends to be more informative, or that the speaker assumes as something more important to the listener. Focus can be signaled acoustically in an utterance. For instance, it is usually agreed that focus is closely correlated to acoustic parameters like durations and pitch. In an utterance, the acoustic realization of focus can be stated in this way: Firstly, there is generally a great rise of F0 on the focused phrase [1-3]; secondly, there is usually an increase in the duration of the focused syllables [4, 5]; and thirdly, there is normally a global pitch compression in the following part, either through a low plateau, a steady fall, or a constant fall till the end of the utterance [2, 3]. It is shown that spectral emphasis is also a reliable correlate of focus, besides parameters like duration and pitch. Heldner [6] argues that, spectral emphasis value is a more reliable correlate compared with intensity. On spectral emphasis, there is a significant effect of position in phrase, word accent and vowel height, and it may provide a better predictor of focus in general.

Regarding spectral emphasis, researchers have adopted several measures that might fall into this category. A measure of 'spectral balance' was defined in the influential study by Sluijter and van Heuven [7], which refers to the intensity in four frequency bands: 0–0.5 kHz, 0.5–1 kHz, 1–2 kHz and 2–4 kHz. Spectral emphasis is measured as the difference between the overall intensity value and the intensity value in a low-pass filtered signal in some other authors' work [8]. One approach is to calculate the difference (in dB) between the overall intensity value and the intensity value in a speech signal, which was low-pass filtered at 1.5 times the fundamental frequency mean value for each utterance. The reason for a filter cut-off frequency at 1.5 times the fundamental frequency is to get a normalized measure of the energy in higher frequency bands, separating the fundamental frequency from the rest of the harmonics [6].

There has been much research work on the representation of the duration and the pitch of focus in Mandarin. It has been shown that, focus patterns are realized as pitch range variations, which is imposed on different parts of an utterance. The pitch range preceding the focal part is quite similar to that of the neutral focus condition, the range of pitch contours directly under focus is significantly expanded, and the pitch range following the focus is substantially suppressed. Therefore, it seems that there are three focus-related pitch ranges: neutral in the pre-focal parts, expanded in the non-final focused parts, and suppressed in the post-focus parts. It is also displayed that the energy of the focused part increases the rising slope of the rising tones in Mandarin, and studies on focus in both English and Mandarin has demonstrated many similarities between the two languages [3, 9].

In regard to the lengthening of focus, it is revealed that, when a word occurs in utterance medial position, focus induces significant lengthening. The pattern of lengthening is not uniform if a focused domain is a phrase: the final syllable lengthened the most, revealing a tendency of edge effect. Outside the focused phrase, there is also a spill-over lengthening on the adjacent syllables. The extent of such lengthening is controlled by prosodic boundaries, where word boundaries cause lengthening more than syllable boundaries [5].

Syllables in most Mandarin words are of almost equal stress, except those words with neutral tones, as Mandarin is not a stress language. Lin et al. [10] investigated the maximum intensity values of disyllabic words in Mandarin Chinese, and found that, the maximum intensity value of the first syllable is larger than that of the second in most cases. However, they did not compare the intensity value of focused and unfocused words. The present work will analyze the effect of focus on spectral emphasis of bi-syllabic words in Mandarin. In particular, it will answer the following questions. What is the effect of focus on the realization of spectral

emphasis of disyllabic words in Mandarin? What is the pattern of spectral emphasis for bi-syllabic words under the unfocused and the focused condition?

II. METHODOLOGY

2.1 Speakers And Stimuli

Twelve native speakers of Standard Mandarin, six male and six female, participated in the experiment. The materials are twenty bi-syllabic verbs, in the pattern of ‘Onset1 Rhyme1 Onset2 Rhyme2’ forms, such as ‘Shanghai’ (hurt) and ‘Xinshang’ (appreciate). Most of the syllables are consisted of two parts in Mandarin, i.e., the onset and the rhyme, except for the ‘zero-onset’ syllables. For instance, in the syllable ‘shang’, the onset is the consonant ‘sh’, and the rhyme is the vowel-nasal combination ‘ang’. There is no onset in a zero-onset syllable like ‘ai’, with only the rhyme. Only syllables with both onset and rhyme were studied in the present research work, and the spectral emphasis value of the onset and rhyme were investigated separately. For the twenty stimuli, the onsets include fricatives, like ‘x’, ‘sh’, etc, and nasals, like ‘n’, ‘m’. The rhymes include monophthongs, like ‘i’, ‘u’, etc, diphthongs, like ‘ai’, ‘ao’, etc, triphthongs, like ‘iou’, and VN combinations, like ‘in’, ‘ang’, etc.

All the twenty verbs are normally stressed in the stimuli, without neutral tone words. They occur in utterance medial position, in the carrier structure of ‘Nana VERB Lili’, where the words ‘Nana’ and ‘Lili’ are supposed to be the names of two girls. The utterances were read under two focal conditions, with one focusing on the initial name ‘Nana’, and the other focusing on the VERB. Therefore, there yielded two focal conditions for the VERB, i.e., unfocused and focused. The position of focus was elicited by questions. For the utterance initial focus condition, the question is ‘Shui VERB Lili? (Who VERB Lili?)’, and for the medial focus condition, it is ‘Nana zenme Lili? (What did Nana do to Lili? or How does Nana like Lili?)’.

2.2 Procedure And Measurements

The orders of the utterances were randomized when recording. The questions for eliciting focus were recorded beforehand, and played by a loudspeaker. The speakers were instructed to read the answers after the question was played. Each speaker read the utterances on each focus condition once, yielding a total of 480 recorded utterances (12 speakers × 20 sentences × 2 focus conditions).

Acoustic data of the utterances were segmented and labeled after the recording process, with the onsets and the rhymes of both the first and the second syllables of the verbs marked, and intensity values extracted using Praat [11]. The segmentation of the onset and rhyme was first done by a segmenting program, and then manually corrected. For spectral emphasis value, the signal was low-pass filtered at 1.5 times the mean fundamental frequency for each utterance, and the difference (in dB) between the overall intensity value and the intensity value was computed. Analysis was conducted by a self-written visual basic program, with which the average of the spectral emphasis within the onset and the rhyme of each syllable of the key word were computed. Statistic analysis was conducted in SPSS.

III. RESULTS

Fig. 1 shows the spectral emphasis values of the onset and the rhyme, of both the first and the second syllable, under both unfocused and focused conditions. In the subsequent sub-sections, detailed analysis will be given about them.

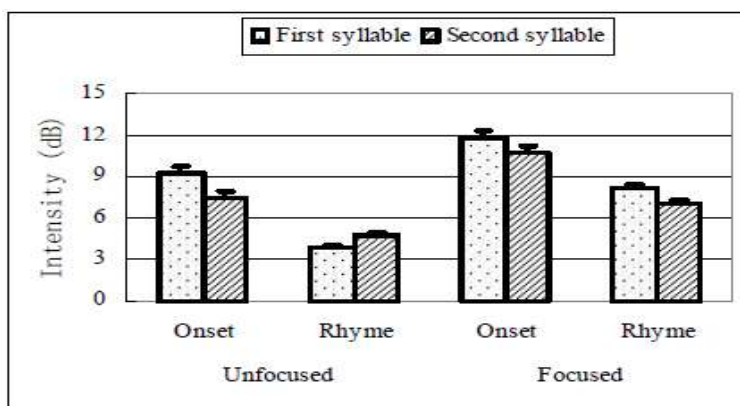


Fig. 1 Spectral emphasis of onset and rhyme at two syllable positions, under two focused conditions

3.1 Onset versus rhyme

It is presented from repeated measures ANOVA results that, the main effect of the onset versus the rhyme is significant, i.e., there are significant difference between the spectral emphasis values of the onset and

the rhyme: $F(1, 239) = 125.7, p < 0.001$, with the spectral emphasis value of the onset much larger than that of the rhyme.

3.2 The First Versus The Second Syllable

3.2.1 uNDER UNFOCUSED CONDITION

Repeated measures ANOVA results display that, under unfocused condition, there exist significant differences between the spectral emphasis values of the first and the second syllables, for both the onset and the rhyme. However, there is interactive effect. For onset, $F(1, 239) = 6.83, p = 0.013$, with the spectral emphasis value of the first syllable larger than that of the second. For rhyme, $F(1, 239) = 28.5, p < 0.001$, with the spectral emphasis value of the second syllable greater than that of the first.

3.2.2 Under focused condition

Under focused condition, it is demonstrated from repeated measures ANOVA result that, there is no effect of position on the spectral emphasis value of the onset: $F(1, 239) = 1.56, p = 0.21$. However, the effect on the rhyme is significant: $F(1, 239) = 7.69, p = 0.006$, with the spectral emphasis value of the first syllable larger than that of the second.

3.3 Focus

3.3.1 Spectral emphasis

It is displayed that the effect of focus on the spectral emphasis value is rather great. Repeated measures ANOVA results reveal that, whether for the onset or the rhyme, and whether for the first or second syllable, the effect of focus on the spectral emphasis value is always significant, with that under the focused condition much larger than that under the unfocused one. For onset, first syllable: $F(1, 239) = 115.3, p < 0.001$; second syllable: $F(1, 239) = 136.8, p < 0.001$. For rhyme, first syllable: $F(1, 239) = 187.5, p < 0.001$; second syllable: $F(1, 239) = 98.2, p < 0.001$.

3.3.2 Emphasis degree.

It is indicated that the effect of focus on spectral emphasis value is rather great in the foregoing subsection. Emphasis degree will be investigated in this subsection. Emphasis degree indicates the difference of spectral emphasis value between focused condition and unfocused condition, as is shown in (1).

$$Dsp = SpeF - SpeU \quad (1)$$

In (1), Dsp stands for emphasis degree value, SpeF for spectral emphasis under the focused condition, and SpeU for that under the unfocused condition.

Fig 2 displays the value of emphasis degree for the onset and the rhyme at two syllable positions. Repeated measures ANOVA results demonstrate that there is no significant main effect between the values of emphasis degrees of the onset and the rhyme: $F(1, 239) = 2.86, p = 0.132$, as there is a significant syllable position \times onset/rhyme interaction: $F(1, 239) = 35.7, p < 0.001$. However, the effect of syllable position on emphasis degree is significant: $F(1, 239) = 5.68, p = 0.021$, with the value of emphasis degree of the first syllable larger than that of the second. Further analysis displays that, as for the first syllable, the emphasis degree of the rhyme is larger than that of the onset: $F(1, 239) = 25.8, p < 0.001$.

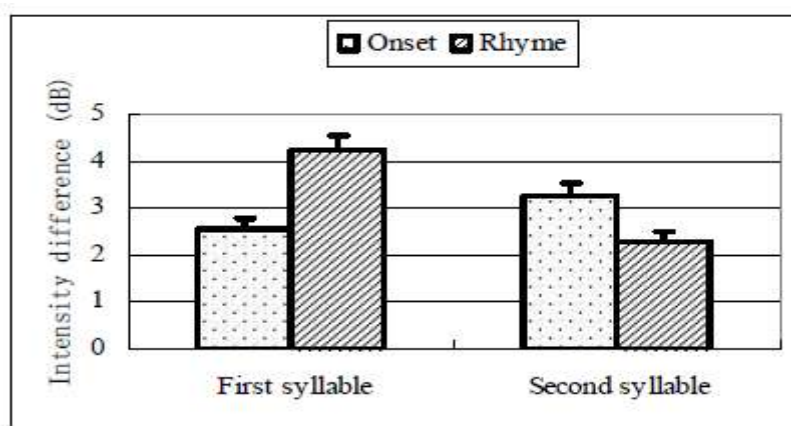
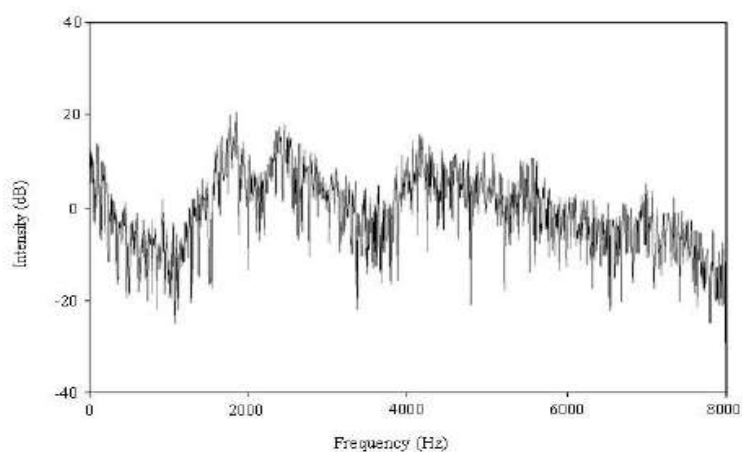


Fig. 2 Emphasis degree of the onset and the rhyme at two syllable positions

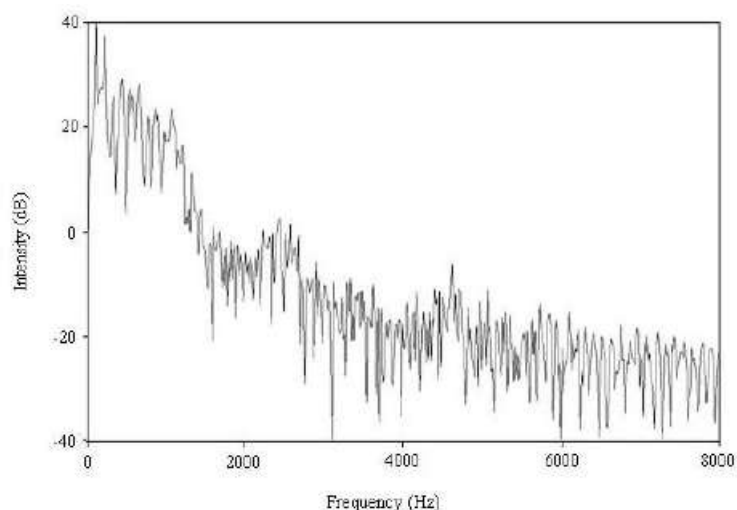
IV. DISCUSSION

Results of the previous section showed that, the spectral emphasis value of the onset is much larger than that of the rhyme. The reason for this representation is as follow. Generally speaking, the onset is a consonant and the rhyme is a vowel in most cases. Most of the consonants are unvoiced, and the vowels are always voiced in Mandarin. For the voiceless sounds, the energy in the higher frequency bands is comparatively great, but for the voiced sounds, the energy in the lower frequency bands is strong. Spectral emphasis value is a measure of the energy value in the higher frequency bands in this research work, excluding the fundamental frequency. Therefore, the spectral emphasis value of the onset is larger than that of the rhyme.

Fig. 3 graphs the spectrums of the consonant ‘sh’ (Fig. 3-a) and the vowel ‘ou’ (Fig. 3-b), from which it is obvious that, in the lower frequency bands, the energy of the vowel is much strong, but in the higher bands, the energy value of the vowel drops to a very low level, while that of the consonant remains at a medium level. As a result, the energy of the higher frequency bands of the consonant is comparatively strong.



(a) The spectrum of vowel of ‘sh’



(b) The spectrum of vowel ‘ou’

Fig. 3 The spectrums of (a) consonant ‘sh’ and (b) vowel ‘ou’

It is also displayed from the previous section that, as far as onset is concerned, the spectral emphasis value of the first syllable is larger than that of the second under the unfocused condition. We suppose that the reason for smaller spectral emphasis value of the onset in the second syllable is intervocalic voicing. A consonant tends to become voiced when it occurs at intervocalic position. As is mentioned above, the energy in the lower frequency bands is strong for voiced sounds, but that in the higher frequency bands is comparatively weak. The energy in the higher bands will be reduced when the consonant of the second syllable gets voiced, that is, the spectral emphasis value will get reduced. Therefore, the spectral emphasis value of the onset in the first syllable is larger than that of the second.

However, the pattern is just the opposite for the rhyme. The spectral emphasis value of the second syllable is larger than that of the first. Generally speaking, the energy of the consonant is weak, and that of the

vowel is strong. However, it is just the opposite for spectral emphasis. The spectral emphasis value of the consonant is large and that of the vowel is small. The energy of the rhyme of the first syllable is larger than that of the second syllable for bi-syllabic words. The spectral emphasis value of the second syllable is larger than that of the first for the rhyme, similar to the case of the consonant and the vowel.

The effect of focus is studied in this study, and it is found that it is significant. The spectral emphasis value of the second syllable is larger than that of the first for the rhyme when the key word is under unfocused condition. However, it is just the opposite under focused condition. The spectral emphasis value of the first syllable is larger than that of the second. We speculate that the reason for this is as follows. Generally speaking, the energy of the first syllable is stronger than that of the second for bi-syllabic words. The emphasis degree of the first syllable is larger than that of the second when the word is focused. Emphasis degree refers to the difference of spectral emphasis value between the focused condition and the unfocused condition. What is more, the emphasis degree of the rhyme is also larger than that of the onset under focused condition, since the overall intensity of the rhyme is larger than that of the onset, and the rhyme contributes more on manifesting focus. The spectral emphasis value of the first syllable gets larger than that of the second under focused condition because of these dual effects.

As for the onset, the spectral emphasis value of the first syllable is larger than that of the second when the verb is under unfocused condition. However, under focused condition, there is no effect of syllable position on the spectral emphasis value. The spectral emphasis value of both of the voiced and the voiceless sounds will increase when the key word is under focused condition. Comparatively, since the voiced sounds contribute more on manifesting focus, the voiced sounds will have larger increase than the voiceless sounds. As is mentioned above, some of the onsets in the second syllable will get voiced in the intervocalic position. They will have greater increase on spectral emphasis value than the onset in the first syllable when they get voiced, and as a result, the difference between them disappears. Therefore, there is no effect of syllable position on the spectral emphasis value.

Emphasis degree for focus is also calculated in this experiment, and it is found that emphasis degree of the first syllable is larger than that of the second. The energy of the first syllable is stronger than the second in bi-syllabic word. The first syllable will have greater increase on spectral emphasis value than the second syllable when the word is focused, as it contributes more on manifesting focus. Therefore, the emphasis degree of the first syllable is larger than that of the second.

It is also found that, the emphasis degree of the rhyme is larger than that of the onset for the first syllable. The reason for this is similar to that mentioned above. The energy of the rhyme is larger than that of the onset. The emphasis degree of the rhyme is comparatively large under focused condition, as it contributes more on manifesting focus. Therefore, the emphasis degree of the rhyme is larger than that of the onset.

V. CONCLUSION

The pattern of spectral emphasis, as well as the effect of focus on bi-syllabic words in Mandarin is examined in this study. It is found that, the energy in the higher frequency bands is comparatively strong for voiceless sounds, so the spectral emphasis value of the onset is larger than that of the rhyme. The spectral emphasis value of the onset of the first syllable is larger than that of the second under unfocused condition because of intervocalic voicing. However, the spectral emphasis of the second syllable is larger than that of the first for the rhyme. There is no effect of syllable position on the spectral emphasis value of the onset under focused condition. The spectral emphasis of the first syllable is larger than that of the second for the rhyme. The emphasis degree value of the first syllable is larger than that of the second, and that of the rhyme is larger than that of the onset, as the first syllable and the rhyme contribute more on manifesting focus.

REFERENCES

- [1]. M. Beckman, *Stress and non-stress accent* (Dordrecht: Foris Publications, 1986).
- [2]. Y. Xu and C. Xu, Phonetic realization of focus in English declarative intonation, *Journal of Phonetics*, 33, 2005, 159–197.
- [3]. A. Turk and A. White, Structural influences on accentual lengthening in English, *Journal of Phonetics*, 27(2), 1999, 171–206.
- [4]. Y. Chen, Durational adjustment under corrective focus in Standard Chinese, *Journal of Phonetics*, 34, 2006, 176–201.
- [5]. D. Dahan and J. Bernard, Inter-speaker variability in emphatic accent production in French, *Language Speech*, 39(4), 1996, 341–374.
- [6]. M. Heldner, On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish, *Journal of Phonetics*, 31, 2003, 39–62.
- [7]. A. C. Sluijter and V. J. van Heuven, Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America*, 100(4), 1996, 2471–2485.

- [8]. H. Traunmuller and A. Eriksson, Acoustic effects of variation in vocal effort by men, women, and children, *Journal of the Acoustical Society of America*, 107(6), 2000, 3438–3451.
- [9]. Y. Xu, Effects of tone and focus on the formation and alignment of F0 contours, *Journal of Phonetics*, 27, 1999, 55–105.
- [10]. M. Lin, J. Yan and G. Sun, A preliminary experiment on normally stressed disyllabic words in Beijing Chinese, *Fangyan*, 1, 1984, 57–73.
- [11]. P. Boersma, Praat, a system for doing phonetics by computer, *Glott International*, 5:9/10, 2001, 341–345.