Emphasis, Syllable Duration, and Tonal Realization in Standard Chinese

Yiya Chen

Center of Language Studies; Radboud University Nijmegen
yiya.chen@let.ru.nl

Abstract

This study investigates how durational and F_0 cues are employed to convey degrees of emphasis in Standard Chinese (SC). Three speakers of SC produced all four lexical tones embedded in sentences in which the preceding and following tonal contexts of the target syllable varied. Subjects were primed with pragmatic contexts in which corrective focus, with two degrees of emphasis on the target syllable (i.e. Emphasis and More-Emphasis), was elicited, in addition to a No-Emphasis condition (which served as the baseline for comparison).

Results showed a gradual increase of syllable duration in that the magnitude of increase from the No-Emphasis to the Emphasis condition and that from the Emphasis to the More-Emphasis condition were comparable. F_0 range expansion, however, was non-gradual. While there was a robust increase of F_0 range from the No-Emphasis to the Emphasis condition, the expansion from the Emphasis to the More-Emphasis condition was much more reduced. Examination of the F_0 adjustment of the individual tones suggests that under corrective focus with the two degrees of emphasis, lexical tones were realized with distinctive F_0 contours, adapting to both the neighboring tonal contexts and the gradual increase of the tone-carrying syllable duration.

1. Introduction

It is by now well-known that in Standard Chinese, focus in general increases the duration of the focused syllable and expands the F_0 range of its associated lexical tone ([1], [2], [3], [4], [5]). It is not clear, however, how the durational and F_0 cues are manipulated to convey more subtle differences in the information structure of an utterance in SC. For example, one type of focus – corrective/contrastive focus, employed by speakers to make a contrast or corrections ([6]), is commonly produced with different degrees of emphasis when repeated corrections are sought in pragmatic contexts such as the dialogue in (1).

(1) A: Did Shirley buy the flower?
B: Mary bought the flower. (emphasis on Mary)
A: You said that Nara bought the flower?
B: Mary bought the flower. (more emphasis on Mary)

In Germanic languages such as English, emphasis induces a gradual expansion of F_0 range with increasingly higher F_0 peak (denoted as H tone) for higher degrees of emphasis ([7], [8]). However, there is less consensus on the effect of varied degrees of emphasis on the scaling of F_0 valleys (denoted as L tone). An F_0 valley may be lowered, which, together with peak raising, results in F_0 span expansion; or it may be raised, which then results in overall F_0 level raising (following [9]). Compared to F_0, duration seems to be a less prominent cue for degrees of emphasis in English. Arvaniti & Garding ([8]) found that although there was consistent durational increase for different levels of emphasis, the magnitude of lengthening was barely large enough to be perceptible.

Chinese differs greatly from English in prosody as far as F_0 is concerned, since Chinese is a tonal language and F_0 variation is employed to indicate lexical tonal contrasts. This makes it plausible that while speakers of English rely more on F_0 and less on duration to signal degrees of emphasis, SC is more restricted in the manipulation of F_0 and therefore relies more on duration to convey degrees of emphasis.

Salience of durational increases for degrees of emphasis does not necessarily exclude the possibility that F_0 movements are additionally employed to facilitate the cuing of more subtle differences in emphasis. F_0 span expansion predicts that as the duration of the tone-carrying syllable increases, there should be increasingly higher maxF_0 and lower minF_0. F_0 level raising predicts that with greater duration, maxF_0 and minF_0 should be both higher. Besides pitch range expansion, it is plausible that in Standard Chinese, given the lexical tonal contrasts, it is important to modify the F_0 contours in such a way that the distinctive features of the lexical tones are fully preserved even when the duration of the tone-carrying syllable is increased greatly. These possibilities make it necessary to examine not only the max and minF_0 (for F_0 range), but also the location of the F_0 turning points as well as the speed of F_0 rises/falls (for F_0 movement). All these variables will be examined in three different pragmatic contexts (i.e. No-Emphasis, Emphasis, and More-Emphasis).

2. Method

2.1. Test Materials

The test stimuli are illustrated in (1). The target syllable is indicated as *Y*, its preceding syllable as *X*, and its following syllable as *Z*. For the target syllable, all four lexical tones were included. Its syllable structure includes both simple CV (i.e. *ma*) and complex CGVG (i.e. *miao*) structures. The preceding syllable varies between *shuo* ('say') with a High tone (H) and *xie* ('to write') with a Low tone (L). The following syllable varies between *man* ('slow') with a Falling tone (HL) and *nan* ('difficult') with a Rising tone (LH). Thus the target syllable may be preceded by tones that end high or low and followed by tones that start high or low. The choices for *X*, *Y*, and *Z* were made based upon four factors: semantic meanings of the syllables, possible tonal combinations (since not all syllables carry all four lexical tones), easy segmentation, and the availability of the desired syllable structures in the lexicon. 32 stimuli sentences were included.

(1) zhōu bīn shuō χ ʏ ʒ hěn duō.
zhōu bīn said X Y Z very more

'zhōu bīn said it is much more *Z* (difficult/slow) to *X* write/say) *Y* (target syllable with four different tones).'

2.2. Pragmatic Contexts

The stimulus sentences were elicited in three different pragmatic contexts. Taking the sentence (2) as an example, subjects were first given the sentence in Chinese characters (shown in pinyin here) on the computer screen; they were told that it provided the correct information. They were then also given the wrong information (3) as well as the relevant pragmatic context (4). A typical answer, with emphasis on *miao* (bold and underlined), is shown in (5).

(2) *Correct information:*
zhōu bīn shuō shuō miāo nán hěn duō.
'Zhoubin said that it is more difficult to say *miao*.'

(3) *Wrong information:*
zhōu bīn shuō shuō dǎ nán hěn duō.
'Zhoubin said that it is more difficult to say *da*.'

(4) *Context for emphasis*:
Suppose you gave the correct information in sentence (2), and the experimenter thought you said sentence (3), how would you correct the experimenter?

(5) *Response with emphasis:*
zhōu bīn shuō shuō **miāo** nán hěn duō.

To elicit more emphasis, the experimenter pretended that she did not hear the subject clearly and so the subject had to make the correction once more. This led the subject to repeat the answer (5) with greater emphasis on the syllable *miao* (indicated with double underline in (6)).

(6) *Response with more emphasis:*
zhōu bīn shuō shuō **miāo** nán hěn duō.

A base-line condition was also elicited for the target syllable with question on the last part of the sentence (7). And a typical answer would have emphasis *NOT* on the target syllable, but on the last syllable.

(7) *Baseline condition:*
zhōu bīn shuō shuō miāo zěnme yàng?
What did zhoubin say about saying the word miao?

(8) *Response with No-Emphasis*:
zhōu bīn shuō shuō miāo nán hěn duō.

2.3. Subjects and Recording

2 male and 1 female speakers of SC participated in the experiment. Two were born and grew up in Beijing. One was not born in Beijing but grew up there and speaks SC without any detectable accent judged by the author and the other two subjects. All sentences were automatically randomized with a computer program. Three repetitions, each with a different order, were recorded with a Sony Digital Mega Bass MZ-R55 mini recorder at the sampling rate of 44100 HZ, in the sound booth of the Phonetics Lab at Stony Brook University. The recording was then downsampled to 16000 Hz. All subjects were aware that it was a study of prosody in SC, but were naïve as to the specific analyses. During the recording, subjects were asked to reproduce the sentences whenever the experimenter failed to perceive the intended pragmatic meaning.

2.4. Acoustic Analysis

Data were analyzed in Praat and then with a set of computer programs. Syllable duration, max/min F_0 values and locations, F_0 value of the start of rising/falling and their locations, and the slope of rising or falling (derived by F_0 range divided by the distance of F_0 max-min) were taken ([3] for details). Repeated measures by Subjects were conducted in SPSS with four factors: Pragmatic Context (i.e. No-Emphasis, Emphasis, and More-Emphasis), Tone of the target syllable (all four lexical tones) (but only in §3.1), Syllable Structure (complex vs. simple), Preceding Tone (High vs. Low), and Following Tone (Falling vs. Rising) were performed. Due to space limit, only crucial results were reported in the following.

3. Results and Discussions

3.1. Magnitude of Lengthening as Correlate of Degree of Emphasis

Pragmatic Context significantly affected the duration of the target syllable [$F(2, 4) = 16.62$, $p < .025$]. Bonferroni Post-hoc tests showed that all three contexts differed (Figure 1). The mean duration of the target syllable (*ma* and *miao* pooled together) in the Emphasis condition was on average 74 ms longer than that in No-Emphasis condition. In the More-Emphasis condition, it was 94 ms longer than the Emphasis condition. The magnitude of lengthening in percentage remained consistent (34% increase from No-Emphasis to Emphasis and 32% from Emphasis to More-Emphasis). As a contrast, neither the preceding nor the following syllable exhibited comparable change in the three contexts, excluding the possibility that the observed durational increase on the target syllable was due to the adjustment of the speaking rate. Syllable Structure was also found to be a significant factor [$F(1, 2) = 26.40$, $p < .05$]. There was, however, no interaction between Pragmatic Context and Syllable Structure, suggesting that the durational difference between the two syllable structures was maintained in all three contexts. The durational pattern thus confirmed that corrective focus induced significant lengthening. Under corrective focus, two levels of emphasis were indeed elicited and durational adjustment was a robust manifestation of degrees of emphasis: the greater emphasis with which the target word was produced, the longer its duration.

3.2. Magnitude of F_0 Range Expansion as Correlate of Degree of Emphasis

Pragmatic Context was a significant factor on F_0 range [$F(2, 4) = 13.14$, $p < .025$]. Bonferroni Post-hoc tests showed that the F_0 range of the target syllable differed in all three contexts (Figure 2). The mean F_0 range of the target syllable in the Emphasis condition was about 42 Hz more than that in the No-Emphasis condition. In the More-Emphasis condition, it was only about 10 Hz more than that in the Emphasis condition. This is in clear contrast to the durational pattern: while emphasis did induce significant expansion of the F_0 range, just like the increase of syllable duration, there was much less expansion of the F_0 range for a greater degree of emphasis, different from what we found to be the case for durational increase.

$MaxF_0$ was also significantly affected by Pragmatic Context [$F(2, 4) = 15.70$, $p < .025$] but $MinF_0$ was not [$F(2, 4) < 5$, *n.s.*]. Because Lexical Tone either had a significant main

effect or/and interaction with Pragmatic Context on all three dependent variables, we thus examined further how the F_0 contours of the individual lexical tones were adjusted.

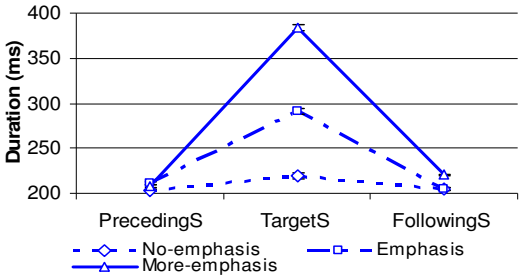

Figure 1 Mean duration of the target syllable(S), PrecedingS, and FollowingS elicited with three emphasis conditions on the targetS

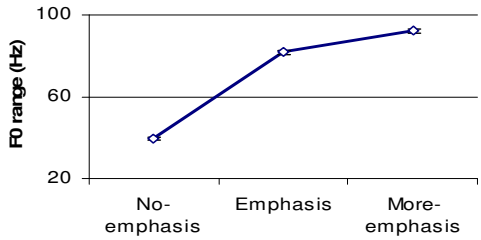

Figure 2. Mean F_0 range of the targetS elicited with three emphasis conditions

3.3. Characteristic F_0 Adjustment of the Individual Tones

3.3.1. High tone

The maxF_0 of a High tone increased significantly from No-Emphasis (185 Hz) to Emphasis (225 Hz) and then to More-Emphasis (233 Hz) [F (2, 4) = 14.29, $p < .025$]. Bonferroni Post-hoc tests showed that all conditions differed significantly. While there was an increase of 40 Hz from the No-Emphasis to the Emphasis condition, there was only 7 Hz increase from the Emphasis to the More-Emphasis condition.

No effect of Pragmatic Context was found on minF_0, which was consistently located at the onset of the syllable (i.e. the offset of the preceding syllable) and was affected by the Preceding Tone [F (1, 2) = 26.02, $p < .05$]. When the preceding tone was high, the slope of the rise from minF_0 to maxF_0 was 142 Hz/second; when the preceding tone was low, the slope of rise was 315 Hz/second. This difference in the effect of Preceding Tone on the F_0 rise slope was significant ([F (1, 2) = 31.77, $p <. 05$]). This suggests that to realize a raised High tone, speakers adjusted the speed of the F_0 rise, accommodating both the tonal context of the focused syllable as well as the durational increase of the syllable. To conclude, for a High tone, corrective focus induced the raising of maxF_0, but a higher degree of emphasis on the word with corrective focus resulted only in a slight increase of the maxF_0 from the Emphasis to the More-Emphasis condition.

3.3.2. Low tone

Pragmatic Context was significant in the minF_0 of the Low tone [F (2, 4) = 18.44, $p < .01$]. Bonferroni Post-hoc tests, however, found that only the lowering from 124 Hz in the No-Emphasis condition to 114 Hz in the Emphasis condition was significant. Pragmatic Context had no effect on the Max F_0.

Two observations are worth noting. The first was that many tokens exhibited varying degrees of creakiness or glottalization after the measurable minF_0 of the Low tone. Impressionistically, this variation was related to the degrees of emphasis in that more serious creakiness/glottalization was observed in the More-Emphasis condition than in the Emphasis condition, but creakiness was rarely found in the No-Emphasis condition. The second interesting observation was the rising tendency of the emphasized Low tone after its minF_0, especially after creakiness/glottalization. Such a tendency to rise was frequently observed in the More-Emphasis condition and occasionally in the Emphasis condition, but not in the No-Emphasis condition. It may be argued that creakiness/glottalization are evidence of speakers' efforts to lower the minF_0. It is not easy to relate the rising tail of the Low tone directly to speakers' effort of lowering F_0, a phenomenon which is usually observed when a Low tone is in utterance-final position ([10]), but this possibility is certainly worth exploring (see [11] for a possibly related phenomenon on F_0 rising following a Low tone).

3.3.3. Falling tone

Pragmatic Context had a significant effect on the maxF_0 of the Falling tone [F (2, 4) = 17.65, $p < .025$]. Bonferroni Post-hoc tests showed that all three conditions differed. The mean F_0 was 189 Hz for No-Emphasis, 233 Hz for Emphasis, and 246 Hz for More-Emphasis. The difference between the Emphasis and More-Emphasis conditions was again small. No effect of Pragmatic Context was found on the minF_0.

We further examined the slope of the F_0 fall. Pragmatic Context was a significant factor [F (2, 4) = 9.75, $p < .05$]. Bonferroni Post-hoc tests showed that all contexts differed significantly. The falling slope increased from 385 Hz/second in the No-Emphasis condition to 785 Hz/s in the Emphasis condition, but decreased to 703 Hz/s in the More-Emphasis condition (due to the great magnitude of the durational increase from the Emphasis to the More-Emphasis condition).

Pragmatic Context also affected the start of the F_0 falling [F (2, 4) = 15.74, $p < .025$]. The more emphasis it was, the later the falling started (No-Emphasis: 103 ms; Emphasis: 157 ms; More-Emphasis: 189 ms). Bonferroni Post-hoc showed that all three levels differed significantly. Such delayed start of falling correlated well with the duration of the tone-carrying syllable. In other words, the F_0 falling was aligned further away from the onset as the syllable duration increased. Together with the raised F_0 peak, the delayed fall ensured a distinctive falling contour despite the durational increase of the tone-carrying syllable.

3.3.4. Rising tone

Pragmatic Context affected the scaling of maxF_0 [F (2, 4) = 20.77, $p < .01$]. Bonferroni Post-hoc tests showed that all contexts differed (No-Emphasis: 160 Hz; Emphasis: 200 Hz; More-Emphasis: 207 Hz). Pragmatic Context had no effect on the minF_0 but did affect the slope of the F_0 rise [F (2, 4) = 14.39, $p < .025$]. Bonferroni Post-hoc tests showed that all three levels differed significantly (No-Emphasis: 227 Hz/s; Emphasis: 480 Hz/s; More-Emphasis: 427 Hz/s).

The start of rise was affected by Pragmatic Context [F (2, 4) = 19.8, $p < .001$]. Bonferroni Post-hoc tests showed that all three contexts differed significantly (No-Emphasis: 139 ms;

Emphasis: 199 ms; More-Emphasis: 246ms). Such delayed rise again correlated well with duration. In other words, the start of rise was aligned further away from the onset as the syllable duration increased, similar to that of the start of fall in a Falling tone. Such an alignment pattern resulted in a distinctive rising contour despite the durational increase of the tone-carrying syllable.

There was an interaction of Pragmatic Context with Preceding Tone on the start of F_0 rise [$F(2, 4) = 23.4$, $p < .01$]. In the No-Emphasis condition, Preceding tone did not affect the start of rise. In the Emphasis and More-Emphasis conditions, however, the start of rise was delayed more when preceded by a Low tone than a High tone. In other words, when the Rising tone started high (i.e. preceded by a High tone), it took less time to start rising (Emphasis: 179 ms; More-Emphasis: 216 ms) than when it started low (i.e. preceded by a Low tone) (Emphasis: 217 ms; More-Emphasis: 275 ms). This pattern is contrary to what we would have expected. Everything else being equal, it should take more time for the Rising tone to start rising when preceded by a High tone than by a Low tone, since in the former case, F_0 would have to first lower from the high F_0 offset of the preceding High tone, and then start rising. One possible explanation for such an "unexpected" later rise after a Low tone is that subjects strived for a distinctive rising F_0 contour of the rising tone under emphasis, as suggested in Figure 3, which shows the F_0 contours of a Rising tone and a High tone, both preceded by a Low tone. The later rise for the Rising tone would certainly help to maximally distinguish between the two rising contours.

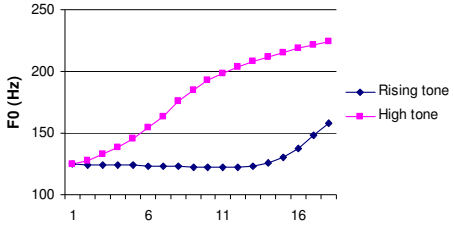

Figure 3. F_0 of a High vs. a Rising tone, in the Emphasis condition, preceded by a L tone followed by a Falling tone (time-normalized over 9 utterances of 3 subjects)

4. Conclusions

There was a robust and gradual increase in syllable duration from the No-Emphasis, to the Emphasis, and the More-Emphasis condition. The F_0 range expansion, however, was non-gradual: Although there was a robust increase from the No-emphasis to the Emphasis condition, only a limited degree of expansion was observed from the Emphasis to the More-Emphasis condition. This makes it clear that in SC, corrective focus indeed induced a significant durational increase and F_0 range expansion of the focused syllable. Under corrective focus, however, F_0 manipulation was restricted and speakers of SC had to rely more on duration to convey different degrees of emphasis, different from what we have observed in English ([7] and [8], among others). This is in line with Parallel Encoding model proposed in [12], in which focus, as an independent pragmatic function, is encoded with a specific interval of F_0 range. Therefore, despite the different degrees of emphasis induced for corrective focus, the available pitch range for focus remained.

In addition to F_0 range expansion under focus (which due mainly to the raising of the maxF_0), it is important to note that tonal targets were realized with characteristic F_0 contours, accommodating to the neighboring tonal contexts and adapting to the increased duration of the tone-carrying syllable. Briefly, the High tone was produced with a raised F_0 peak. The Low tone was produced with a slightly lowered F_0 valley, accompanied by creakiness/glottalization, and sometimes a rising tail. The Rising and Falling tones exhibited delayed start of rise and fall, which arguably contributed to the realization of sharper rising or falling contours, the characteristic F_0 patterns of the two lexical tones respectively. These F_0 adjustments suggest that tonal implementation is an important manifestation of pragmatic meanings in Standard Chinese.

5. Acknowledgement

This paper is based on data reported in Chapter 3 of my dissertation, which has benefited greatly from the questions and comments of Ellen Broselow, Marie Huffman, Chilin Shih, and Yi Xu. I also thank Carlos Gussenhoven for comments on an earlier version of this paper. Usual disclaimers apply. This work was in part supported by the VENI grant from the Netherlands Organization for Scientific Research (NWO).

6. References

[1] Jin, S. 1996. *An acoustic study of sentence stress in Mandarin Chinese*. PhD dissertation. Columbus, OSU.

[2] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F_0 contours. *Journal of Phonetics* 27(1): 55-105.

[3] Chen, Y. 2003. *The Phonetics and Phonology of Contrastive Focus in Standard Chinese*. PhD dissertation. Stony Brook Univ.

[4] Yuan, J. 2004. *Intonation in Mandarin Chinese: Acoustic, Perception, and Computational Modeling*. PhD dissertation. Cornell University.

[5] Chen,Y. & Braun, B. Submitted. The prosodic realization of information structure categories in Standard Chinese.

[6] Gussenhoven, C. 2004. Types of focus in English. In *Topic and Focus: Intonation and Meaning, Theoretical and Crosslinguistic Perspectives*, Daniel B., Gordon M., & Lee, C. (ed.). Dordrecht: Kluwer.

[7] Liberman, M. & Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length. In *Language, Sound, Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students,* Aronoff, M. & Oehrle, R. (eds.). 157-233.

[8] Arvaniti, A. & Garding, G. To appear. Dialectal variation in the rising accents of American English. In *Laboratory Phonology 9*, Cole, J. & Hualde, J. (eds.).

[9] Ladd, R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[10] Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

[11] Chen, Y. & Xu, Y. (In press). Production of weak elements: Evidence from neutral tone in Standard Chinese. *Phonetica.*

[12] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication*. 220-251.