In Proceedings of International Conference on Speech Prosody 2004, Nara, Japan. pp. 81-84.

On the Temporal Domain of Focus

Yi Xu[*], *Ching X. Xu*[†] *& Xuejing Sun*[‡]

[*]Haskins Laboratories, New Haven, CT, USA, [†]Northwestern University, Evanston, IL, USA,
[‡]Panasonic Speech Technology Labratory, Santa Barbara, CA, USA
xu@haskins.yale.edu; xxq@northwestern.edu; xuejing@research.panasonic.com

Abstract

It is well known that focus affects the pitch of what is being focused. It is much less recognized, however, that focus also extensively affects the pitch ranges of non-focused regions in a sentence. In this paper we show evidence that the temporal domain of focus is much wider than has been generally recognized. We present acoustic, perceptual, and imitational data demonstrating that, in a declarative sentence, focus is realized not only by expanding the pitch range of the focused item, but also by compressing the pitch range of post-focus items, and possibly requiring that the pitch range of pre-focus items remain neutral. We conclude that the domain of a single, narrow focus consists of three temporal zones, with distinct pitch range adjustment for each. These pitch range specifications therefore should be treated as attributes of the focus itself rather than as anything else.

1. Introduction

The acoustic realization of focus, i.e., discourse/pragmatics motivated emphasis, has been investigated directly or indirectly in many studies e.g. [2–5, 7–9, 17]. The general consensus has been that focus is conveyed mainly through variations in F_0, although certain amount of amplitude and duration adjustment is also involved [4, 17]. However, what has not been widely recognized is the fact that focus realization involves the adjustment of pitch ranges of not only the focused components, but also regions before and after focus. Pitch range adjustments of non-focused regions have been reported before [2, 3, 5, 7, 12, 13, 17], but have not yet raised sufficient attention to modify the general view about focus. In this paper, we will briefly review lines of evidence from recent production, perception and imitation studies that all point to a pattern of tri-zone pitch range control by focus.

2. Production

The first line of evidence comes from acoustic data. Fig. 1 displays F_0 curves of Mandarin tone sequences H□H H□H□H and H□L□R□H□H with focus at three different locations or without any narrow focus, adapted from [17]. The first thing to notice in Fig. 1 is that focused words assume an expanded pitch range, i.e., with higher F_0 peaks and lower F_0 valleys, whichever is applicable. The expansion is greater upwards than downwards, and very small for the final word. In the top panel, F_0 of any syllable directly under focus is raised relative to the neutral-focus condition. In the bottom panel, F_0 of H-tone syllables (1 and 5) is raised under focus. Syllable 2, which carried the L tone, has lowered minimum F_0 under focus. Syllable 3, which carries the R tone, has expanded F_0 in both directions, although the upward expansion is much larger. Secondly, the pitch range of post-focused region is both lowered and narrowed. The only exception is the initial F_0 of the first post-focus syllable, which is raised if the preceding focused tone is H or R. But this is directly attributable to the carryover effect of the preceding tone [16, 17]. Thirdly, the pitch range of pre-focus words remains largely similar to that of the neutral focus condition. Similar findings about focus in Mandarin have been reported by other studies [5, 7, 13].

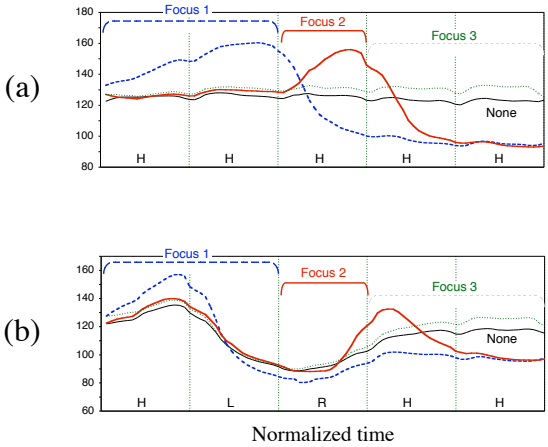

Figure 1. *Mean F_0 contours of the tone sequences H H H H H (top) and H L R H H (bottom) averaged across 24 repetitions by 4 speakers.*

Similar focus patterns have also been found in non-tone languages such as English. Cooper et al. [3] found that the effect of a narrow focus is to raise the F_0 of the focused word and lower the F_0 of the post-focus words in a sentence. In contrast, F_0 of the pre-focus words remained much the same as in a focus-neutral sentence. A recent investigation of focus and accent in English [18] further confirmed the main findings of [3]. Some examples are shown in Fig. 2. In Fig. 2a, the sentence "Lee may know my niece" is said with a narrow focus on "Lee" or "niece", or without a narrow focus. When focus is on "Lee", not only is its F_0 raised, but also F_0 of all following words is lowered. The shallow peak in the word "niece" in the no-focus condition now has only a faint trace. When focus is on "niece", its F_0 is raised. But the F_0 of the preceding words does not differ much from that of the no-focus condition. In Fig. 2b, the sentence "Lee may lure my niece" is said with or without focus on "lure." With focus, F_0 of "lure" is much higher than in the no-focus condition. Furthermore, F_0 of the post-focus words, especially that of "niece", is lowered from the no-focus version, and there is hardly any visible peak in "niece." Slightly different from Fig. 2a, pre-focus words in Fig. 2b seems to have lower F_0 than in the neutral-focus condition.

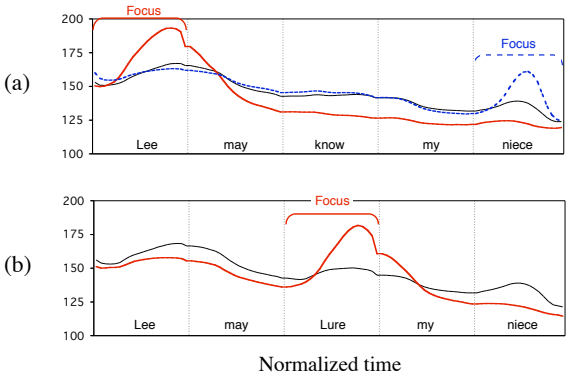

Figure 2. *Mean F_0 contours of the sentences "Lee may know my niece" averaged across 49 repetitions by 7 speakers.*

3. Perception

The second line of evidence comes from perception data. Findings from several studies on focus perception indicate that low F_0 after focus is critical for the correct identification of focus. Rump and Collier [11] examined the effect of relative height of early and late F_0 peaks on Dutch listeners' perception of focus. They found that to perceive a single early focus, not only did the early F_0 peak have to be very high, but also the later F_0 peak had to be extremely low. In contrast, to perceive a late focus, the early F_0 peak did not have to be very low. It just needed to be lower than the later peak. Hasegawa and Hata [6] investigated the effect of the falling slope of an F_0 peak on the perception of accents in Japanese and English. They found that by merely increasing the post-focus downward slope, the location of perceived focus could be shifted to the previous syllable. While it could be argued that accents in Japanese and English are not the same thing, the finding about English is in agreement with the fact that, in production, a sharp F_0 fall always occurs immediately after the focused component, as can be seen in Fig. 2.

To test the general idea that certain intonations consist of multiple temporal regions, we developed two closely related experimental paradigms inspired by the phenomenon of "phonemic restoration" [15]. We refer to them as "*Prosodic Restoration*" (PR) and "*Imitation via Prosodic Restoration*" (IPR). In both paradigms, an intonation under scrutiny is first recorded by a native speaker. Then words carrying a potential constituent of the intonation are replaced by a noise that is loud enough to have actually masked them. During the experiment, the sentence containing the replacement noise is presented to the subject together with the text. In PR, the subjects' task is to identify the prosodic component or determine the meaning of the intonation. In IPR, the subjects' task is to repeat the sentence in exactly the same way as they hear it. Our first experiments using PR and IPR were conducted to determine whether there are multiple temporal regions intrinsic to focus. The main results of the PR experiment will be presented in the following. The results of the IPR experiment will be discussed in the next section.

A total of 12 sentences were recorded by a native speaker of Mandarin with focus on word 0, 1, 2, 3 or 4 where 0 indicates no narrow focus. Some examples are shown below.

Group 1 — Tone sequence: H H H H H H H H H
E.g. Zhōngguāncūn jīntiān gōngkāi zhāobiāo.
[There is an open bidding today in Zhongguancun]

Group 2 — Tone sequences: F H F H F H F H F
H F H F H F H F H
E.g. Zhùtīngqì xūyào jīnyè xiūfù
[The hearing aid needs to be fixed tonight]

Group 3 — Tone sequence: R H R H R H R H R
H R H R H R H R H
E.g. Huáng Zhōngxiáng chūmén ānquán guīlái
[Huang Zhongxiang returned home safely]

0–3 of the four words in the sentences were replaced by a pink noise of the same duration that was 8 dB above the peak amplitude of the sentence. Fig. 3 shows the mean F_0 curves of the all-H sentences used as stimuli. The five curves are the average F_0 tracings of the sentences produced with focus on the 1st, 2nd, 3rd, or 4th word, or with no narrow focus. The dotted square shows the location of the noise when the second word of the sentence is replaced by the noise.

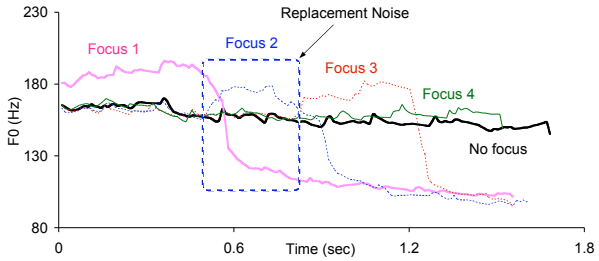

Figure 3. *Mean F_0 curves of the all-H sentences used as stimuli. The curves are the average F_0 of the sentences produced with focus on the 1st, 2nd, 3rd or 4th words, or with no narrow focus. The dotted square indicates the location of the noise in the condition where the second word of the sentence is replaced by noise.*

Ten native speakers of Mandarin (5 males and 5 females) participated as subjects. They were presented with the stimulus sentences together with the corresponding texts. The task was to determine which, if any, word in each sentence was emphasized. Due to space limit, only part of the results will be presented. Table 1 shows percentage of focus identification when all post-focus words are replaced by noise. As can be seen, focus location can be correctly identified without the presence of post-focus words.

Table 1: *Percentage correct focus identification when post-focus words were replaced by noise. The numerals in the column and row headings indicate locations in the sentence.*

Actual Focus	Noise Location		
	234	34	4
1	92.5		
2		97.2	
3			89.3

Table 2: *Percent correct identification of original focus from stimuli where only one word was replaced by noise.*

Actual Focus	Noise Location			
	1	2	3	4
0	70.0	56.9	63.5	67.7
1	*72.8* (20.5)	87.9	98.8	92.3
2	96.7	*69.4* (20.4)	95.9	97.1
3	100.0	96.8	*79.1* (12.8)	89.3
4	62.5	63.9	72.4	*53.0*

Table 2 displays percent correct identification of the original focus from stimuli where only one word is replaced by noise. The results are broken down by *actual focus* and *noise location*. Of particular interest are the numerals in italic, which are from conditions where the entire on-focus portion was replaced by noise. Although not as high as in some other conditions, they are nevertheless well above the chance level of 20%. Furthermore, the numerals in brackets are percent identification of plausible focus. For example, the 20.5% in noise location 1, focus 1, is for identifying word 2 as focused. As can be seen in Fig. 3, the noise that replaces a focused word would not replace the sharp F_0 fall after focus. When focus is on word 1, the sharp fall occurs at the beginning of word 2. When word 1 was totally absent, subjects probably sometimes heard the sharp F_0 fall, especially the initial high portion, as an indication that word 2 was originally focused. The numerals in the other two pairs of parentheses are for identifying the pre-focus word as focused. In those two cases, since the focused word is absent, the F_0 of the pre-focus word become the highest in the sentence. Thus it is not very surprising that subjects sometimes heard pre-focus words as focused. Taking these "plausible" judgments into consideration, the percentage of correct focus identification in Table 2 appears even more remarkable.

In general, the PR experiment found that focus in Mandarin could be recognized fairly well not only when the entire post-focus portion was replaced by noise, but also when the entire on-focus portion was replaced by noise. Furthermore, when both on-focus and post-focus words were present, focus could be recognized with high consistency; and when neither on-focus nor post-focus words were available, focus was quite difficult to determine. Therefore, for Mandarin at least, both on-focus pitch range expansion and post-focus pitch range compression provide critical information for the perceptual identification of focus.

4. Imitation via Prosodic Restoration (IPR)

The third line of evidence comes from our recent findings obtained in an experiment using the IPR paradigm (*Imitation via Prosodic Restoration*). As explained briefly in the previous section, this paradigm also uses noise-filled sentences as stimuli. Instead of making any prosodic judgment, however, subjects simply imitate the noise-filled sentences as accurately as possible. Since the text is provided, the only thing that needs to be "restored" during the imitation is the exact pronunciation of the missing words, including their pitch values. The consistency with which the missing parts of the target intonation are restored in the imitation would then provide indication as to how closely they are related to the parts of the intonation that are still audible.

The same noised-filled sentences used in the aforementioned perception experiment were used as stimuli. Eight native speakers of Mandarin participated as subjects. During the experiment, the text (in Chinese characters) of each stimulus sentence was displayed on a computer screen together with a button for playing the stimulus. The subject imitated each sentence after playing it by pressing the button.

Fig. 4 displays *normF0* of each word of the sentence produced by all subjects while imitating sentences with focus on the second word. *normf0* is calculated with the equation:

$$normF0_{wi} = (meanF0_{wi} - meanF0_s) / meanF0_s$$

where *meanF0* is the average of all F_0 points of either the whole word or the entire sentence; *wi* indicates the *i-th* word of the sentence; and *s* indicates the entire sentence. A positive bar indicates that *meanF0* of the word is above that of the sentence. The bars with four colors/shades represent *normF0* of the four words in a sentence. Therefore each four-bar sector represents the results of a particular noise-replacement condition. The plus sign "+" indicates that at least one word in a given region is not replaced by noise. The minus sign "–" means that all words in the region are replaced by noise.

Sector 2 of Fig. 4 shows *normF0* for the condition where the two post-focus words were both replaced by noise. Note that the second bar is highly positive, indicating high overall F_0 in the second word, which, as can be seen in Fig. 3, has much higher F_0 than the rest of the sentence. Note also that both the third and fourth bars have very negative values, indicating much lower F_0 than in the rest of the sentence. This is despite the fact that subjects did not actually hear the lowered F_0 in the two words, since both were replaced by noise. In sections 3–6 in Fig. 4, the post-focus words were imitated with negative *normF0*, indicating successful imitation of post-focus lowering. Note further that in sections 3 and 5, *normF0* of the second word was positive. An ANOVA found that *normF0* here is significantly higher than in the [–focus, –post-focus] condition (not shown in the figure). This indicates that subjects reproduced the raised F_0 of the focused word without hearing it. In section 4, only word 1 has positive *normF0*. In this case, the first two words were replaced by noise. Subjects seem to have determined, based on ambiguous information, that it is the first word that has been focused.

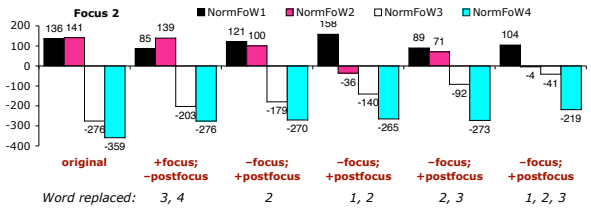

Figure 4. *Normalized F_0 of each word produced by all subjects while imitating those sentences with focus on the second word.*

Similar patterns of prosodic restoration were found in other focus conditions, with the exception of final focus, which, as can be seen in both Fig. 1 and 3, does not have an F_0 pattern very distinct from that of a neutral-focus sentence. The general findings of the IPR experiment are summarized as follows:

1. When on-focus and post-focus words were both present, subjects reproduced on-focus F_0 raising as well as post-focus F_0 lowering.

2. When only the focused word was present (sometimes together with pre-focus words), subjects also produced both on-focus F_0 raising and post-focus F_0 lowering.
3. When only post-focus words were present (sometimes together with pre-focus words), subjects, again in most cases, produced both on-focus F_0 raising and post-focus F_0 lowering, although the location of the F_0 raising may be different from that in the original sentence, especially when too many words were replaced by noise.
4. When both focused word and post-focus words were missing, imitated F_0 was atypical of any narrow focus pattern.
5. Imitation of final-focus sentences was similar to that of sentences with no narrow focus.

5. Discussion: Tri-zone pitch range control by focus

With focus, the speaker tries to indicate which particular word or words, or sometimes an even smaller unit [14], should stand out as being emphasized among all the components of an utterance. It follows, then, that what is being emphasized is given special articulatory/acoustic treatments, including larger pitch range, longer duration, greater intensity, more expanded vowel space, more clearly enunciated consonants and vowels, and more forcefully implemented pitch targets, etc. What may also follow from the function of focus is that portions of an utterance that are not being focused should be *deemphasized*. As was found in recent studies, however, deemphasis does not happen evenly in all non-focused regions. Pitch range of post-focus words is compressed extensively. Pitch range of pre-focus words, in contrast, remains largely the same as in utterances without narrow focus, although cases of pre-focus pitch lowering has also been observed (in English) [18].

It has been forcefully argued that the occurrence, location and scope of focus are determined by discourse/pragmatic rather than syntactic factors [1, 14]. Evidence presented in this paper, including previous findings on the production and perception of focus in various languages as well as the results of the PR and IPR experiments on Mandarin, further indicates that once its location and scope are determined, focus determines the pitch ranges of not only the focused item, but also those before and after focus, as long as they are present. Thus the temporal domain of focus is much broader than that of the focused item itself. The implication of this understanding is that pitch range variations due to focus are intrinsic properties of focus, and thus should not be treated as attributes of other factors such as a hierarchical prosodic structure of the utterance, pause or prosodic break, or phrase tone, etc.

Of course, pitch range control is not the only means with which focus is manifested through F_0. There is evidence that in some languages, focus may also involve alternations of local pitch targets. For example, Xu & Xu [18] find that in English, the underlying pitch target for a focused word-final accent is a [fall] as opposed to a [high] when the accent is not focused. This change of pitch target seems to be an additional means that speakers of English employ to convey focus.

6. Conclusion

We have presented three lines of evidence all pointing to a broader temporal domain of pitch range control by focus than has been widely accepted. A narrow focus seems to control not only the pitch range of the on-focus region, but also those of the pre- and post-focus regions. In particular, the pitch range of focused region is expanded; that of the post-focus region is compressed; and that of the pre-focused region remains largely neutral. An important implication of this understanding is that, as attributes of focus, these pitch range variations should not be re-attributed to other factors when constructing a theoretical model of intonation, or developing an algorithm for intonation recognition or synthesis.

7. References

[1] Bolinger, D. L., 1972. Accent is predictable (if you're a mind reader). *Language* 48: 633-644.

[2] Bruce, G., 1977. Swedish word accents in sentence perspective. In *Travaux de L'institute de Linguistique de Lund Xii*. B. Malmberg and K. Hadding. (eds.) Lund: Gleerup.

[3] Cooper, W. E.; Eady, S. J.; Mueller, P. R., 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77: 2142-2156.

[4] Fant, G.; Kruchenberg, A.; Liljencrants, J.; Hertegard, S., 2000. Acoustic-phonetic studies of prominence in Swedish. *TMH-QPSR 2-3/2000, Royal Institute of Technology, Stockholm*. 1-51.

[5] Gårding, E., 1987. Speech act and tonal pattern in Standard Chinese. *Phonetica* 44: 13-29.

[6] Hasegawa, Y.; Hata, K., 1992. Fundamental frequency as an acoustic cue to accent perception. *Language and Speech* 35: 87-98.

[7] Jin, S. (1996). *An Acoustic Study of Sentence Stress in Mandarin Chinese*. Ph.D. dissertation. The Ohio State University.

[8] Liberman, M.; Pierrehumbert, J., 1984. Intonational invariance under changes in pitch range and length. In *Language Sound Structure*. M. Aronoff and R. Oehrle. (eds.) Cambridge, Massachusetts: M.I.T. Press: 157-233.

[9] Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation. MIT, Cambridge, MA.

[10] Pierrehumbert, J.; Beckman, M., 1988. *Japanese Tone Structure*. Cambridge, MA: The MIT Press.

[11] Rump, H. H.; Collier, R., 1996. Focus conditions and the prominence of pitch-accented syllables. *Language and Speech* 39: 1-17.

[12] Selkirk, E.; Shen, T., 1990. Prosodic domains in Shanghai Chinese. In *The Phonology-Syntax Connection*. S. Inkelas and D. Zec. (eds.) Chicago: University of Chicago Press: 313-37.

[13] Shen, J., 1994. Hanyu yudiao gouzao he yudiao leixing [Intonation structures and patterns in Mandarin]. *Zhongguo Yuwen [Journal of Chinese Linguistics]* 1994-3: 221-228.

[14] van Heuven, V. J., 1994. What is the smallest prosodic domain? In *Papers in Laboratory Phonology*. P. A. Keating. CUP. 3: 76-98.

[15] Warren, R. M., 1970. Perceptual restoration of missing speech sounds. *Science* 167: 392-393.

[16] Xu, Y., 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.

[17] Xu, Y., 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55-105.

[18] Xu, Y.; Xu, C. X., forthcoming. Intonation components in short English statements.