PHONETIC REALIZATION OF FOCUS IN SECOND LANGUAGE ENGLISH

Hu Na[1,2] Jia Yuan[1], Wang Meng[1, 3]

[1] Phonetics Lab, Institute of Linguistics, Chinese Academy of Social Sciences, China
[2] Foreign Languages Department, Graduate University of Chinese Academy of Sciences, China
[3] Shandong University of Science and Technology, China

ABSTRACT

This paper examines the realization of narrow focus in English declarative sentences by Chinese EFL learners from four dialectal regions, Zhenjiang, Jinan, Hangzhou and Beijing. The results revealed that F_0 is the closest correlate with focus in Beijing speakers' performance, mainly going through a process of going down after focus, but without significant changes in the pre-focus and on-focus regions. The statistics also shows that duration in the pre-focus region is significantly reduced in Hangzhou speakers' utterance. However, none of the two parameters is significantly varied with the presence of focus in neither Jinan speakers nor Zhenjiang speakers' performance.

Index Terms: focus realization, second language English, Chinese EFL learners, Chinese dialects

1. INTRODUCTION

Perceptually, focus in discourse refers to language items that are most prominent. To realize focus, different acoustic parameters are employed. In language such as English [1] and Mandarin [2], F_0 is identified as the closest correlate with the presence of focus: F_0 of the on-focus words is increased, F_0 in the post-focus regions decreased while that in the pre-focus regions largely unchanged. In Danish, F_0 also cues the presence of focus, but unlike in English it is expanded under focus, F_0 of the on-focus words is reduced [3]. Furthermore, in languages like Cantonese [4] and Taiwan Southern Min [5], the phenomenon of post-focus compression of F_0 is absent. In some other languages such as Spanish, focus is realized by expanding the intensity of on-focus words. Interestingly, studies concerning second language performance demonstrate that there is no post-focus compression in non-Chinese-heritage American learners' Mandarin even though they have been learning Mandarin for a considerable number of years (since high school or college) [6]. This arises the authors' interest about whether Chinese learners of English would realize focus by varying F_0 as native English speakers do. To explore this, Chinese learners of English from different dialect areas are recruited and how they realize focus in short English declarative sentences is examined. The learners' dialects are specified is because there are numerous regional

dialects of Chinese and each of them may influence second language production in different ways. Two of the major ten dialect areas will be focused in this study. One is Guan and the other is Wu. Since each dialect possesses several subtypes and it is impossible to cover them all, two subtypes in each area, which makes four dialects in total, will be focused on. For Guan dialect, Beijing dialect and Shandong dialect are chosen and for Wu dialect, Hangzhou dialect and Zhenjiang dialect are chosen. The geological locations are demonstrated in Figure 1.

Figure 1. The spot that the capital letter "A" indicates is where Zhenjiang city is; the square indicates Hangzhou; the dot, Beijing; the triangle, Jinan.

2. METHODOLOGY

2.1. Material

The stimuli are short declarative sentences with either broad focus or narrow focus, the compositions of which are in Table 1 with the focused words capitalized. Stimuli with narrow focus are prompted by *wh*-questions and the corresponding unfocused ones are said after the question "what happened?". Focus varies in two locations, sentence-initial and sentence-final. Words bearing focus are different in word lengths and stress patterns. Word length varies from mono-syllable to tri-syllable. The lexical stress is located on either word-initial or word-final. All sentences are extracted from "AESOP (Asian English Speech Corpus Project)_CASS" .

Table 1. Compositions of stimuli

Focus locations	Focus conditions	Stimuli
Sentence-initial	Narrow focus	–What about tom? Who can come with him? –JANE/ANNA/SARAN/CATHERINE comes with tom.
	Broad focus	Jane/Anna/Saran/Catherine comes with Tom.
Sentence-final	Narrow focus	–What about Jane? Who can she come with? –Jane comes with TOM/ MANNY/BURNELL/CHRISTOPHE.
	Broad focus	Jane comes with Tom/ Manny/Burnell/Christopher.

2.2. Subjects and recording

For each accent, four subjects are recruited. Two of them are female and the other two are male. They are all local and communicate in dialects on daily basis. They all have learned English for more than ten years and are all majored in English in university. Their oral English proficiency has been evaluated from the perspectives of pronunciations and intonation before recording the corpus and all of them received the rate "High". All speakers reported having no speech disorders.

Recording was conducted in the sound-treated booth at the Phonetics Laboratory at Chinese Academy of Social Sciences (CASS). The recording software was CUHK-SIAT Recording Tool, designed jointly by Chinese University of Hong Kong and CASS. The equipment of the recording was the laptop and the head-wear microphone, Sennheiser PC166. The sampling rate was 16 kHz. Subjects said the sentences once unless they pronounced the words mistakenly. They were kept unaware of the research purpose.

2.3. Data extraction

All sound tracts were first automatically processed by a segmentation program to generate both word and phone level transcriptions. The inaccurate boundaries of vowels were mended manually and irregular points was eliminated. F_0 of each voiced phone was extracted at ten equally distributed points and then transformed into semitone values with the 70 Hz as the reference frequency. Graphs displayed in Figure 2 are mean F_0 curves in semitones averaged over subjects speaking respective dialects. Values adopted in statistic tests are raw fundamental frequency in Hz.

3. ANALYSIS AND RESULTS

Graphs in Figure 2 display the mean F_0 contours in semitones of individual stimulus averaged across subjects in respective dialect areas. In each graph, all four dialect areas were plotted. Speakers from different dialect areas are distinguished by lines of different styles and colors as shown in the legend and the dialects are referred by their capitalized initials: Zhenjiang, "ZJ", Hangzhou, "HZ"; Jinan, "JN"; Beijing, "BJ". Narrow focused conditions (referred to as "N") are indicated by solid lines with markers, while broad focused conditions (referred to as "B"), the dash or dotted lines. The stimuli that are plotted are displayed at the top of each graph with the narrow focused words capitalized. The upper graph is for sentence-initial focused stimulus "Jane comes with Tom" and the lower graph is for sentence-final focused stimulus "Jane comes with Tom." The y-axis is fundamental frequency in semitone and the breaks on the contours indicate word boundaries. Due to the limited length, graphs of the rest stimuli are reserved.
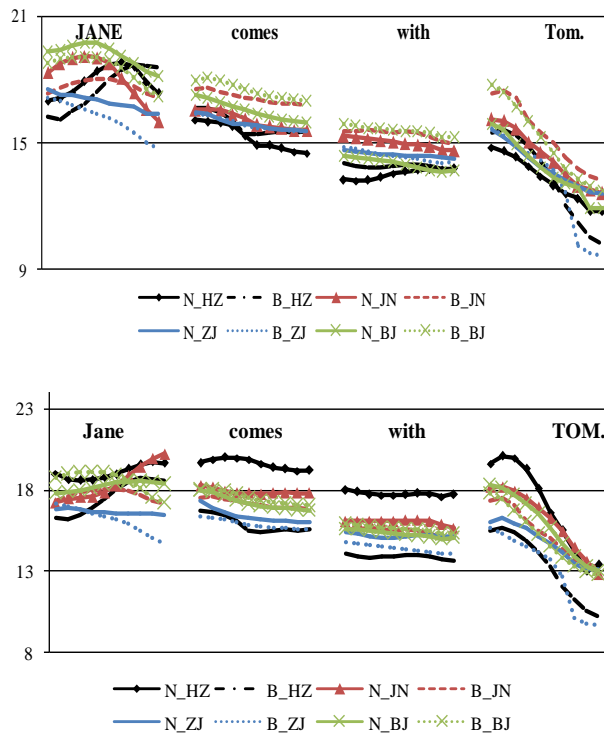

Figure 2. Mean F_0 contours in semitones under different focus conditions of the stimuli "Jane comes with Tom." with two focus locations produced by subjects from the four dialectal areas.

Graphs in Figure 2 show that when focus is at sentence-initial, F_0 of the word is higher when it is under focus than it is not for all four dialect areas. As for the post-focus region, Beijing speakers' F_0 is suppressed when the sentence bears narrow focus. Such suppression is also observed in Jinan and Hangzhou speakers' F_0 contours but

the statistic analysis is needed to evaluate its significance. For Zhenjiang, however, no F$_0$ compression in the post-focus regions is observable.

To test the significance of the observed F$_0$ variations over a focused and an unfocused condition, a series of paired sample t-test was performed with the mean F$_0$, max F$_0$, min F$_0$ and duration as the variables for all focus compositions (mono-syllable, "1S"; bi-syllable stressed at the first syllable, "2S_1"; bi-syllable stressed at the second syllable, "2S_2" and tri-syllable, "3S"), focus locations (sentence-initial, shorted as "I", and sentence-final, "F"). For the on-focus words, mean F$_0$ was computed on the syllable basis, i.e. for bi-syllable word *Manny*, mean F$_0$ of the first syllable and the second are computed separately. While for pre-focus and post-focus regions, syllables contained in each region were regarded as a whole and the mean F$_0$ was computed for all syllables in these areas.

Table 2. Results of Paired Sample t-test on mean F$_0$

Hangzhou					
		1S	2S_1	2S_2	3S
I	on	t=0.715 Sig.=.526	t=-2.211 Sig.=0.114	t=-2.019 Sig.=0.127	t=-0.152 Sig.=0.888
	post	t=-2.003 Sig.=0.139	t=-2.298 Sig.=0.105	t=-4.456* **Sig.=0.021**	T=-1.128 Sig.=0.342
F	pre	t=0.892 Sig.=0.438	t=2.999 Sig.=0.058	t=1.323 Sig.=0.423	t=-997 Sig.=0.392
	on	t=1.124 Sig.=0.343	t=1.040 Sig.=0.375	t=1.389 Sig.=0.978	t=-659 Sig.=0.557
Jinan					
		1S	2S	2S	3S
I	on	t=0.150 Sig.=0.891		t=2.460 Sig.=0.091	t=4.889 Sig.=**0.016**
	post	t=-1.658 Sig.=0.196		t=-1.725 Sig.=0.183	t=-0.708 Sig.=0.530
F	pre	t=0.907 Sig.=0.431	t=0.546 Sig.=0.623	t=-1.152 Sig.=0.333	t=1.556 Sig.=0.218
	on	t=0.155 Sig.=0.887	t=1.782 Sig.=0.173	t=0.954 Sig.=0.411	t=-0.580 Sig.=0.602
Zhenjiang					
		1S	2S_1	2S_2	3S
I	on	t=0.842 Sig.=0.462	t= Sig.=	t=-1.911 Sig.=0.152	t=1.398 Sig.=0.257
	post	t=0.239 Sig.=0.826	t= Sig.=	t=-0.588 Sig.=0.598	t=0.355 Sig.=0.746
F	pre	t=0.582 Sig.=0.601	t=-1.137 Sig.=0.338	t=-0.399 Sig.=0.716	t=-3.746* **Sig.=0.033**
	on	t=2.034 Sig.=0.135	t=0.649 Sig.=0.563	t=-0.937 Sig.=0.418	t=-1.366 Sig.=0.265
Beijing					
		1S	2S_1	2S_2	3S
I	on	t=1.556 Sig.=0.142	t=0.653 Sig.=0.524	t=0.092 Sig.=0.928	t=-1.695 Sig.=0.112
	post	t=-3.142** **Sig.=0.007**	t=-2.313* **Sig.=0.036**	t=-0.607 Sig.=0.554	t=-3.036** **Sig.=0.009**
F	pre	t=-1.038 Sig.=0.317	t=-0.188 Sig.=0.854	t=-1.022 Sig.=0.324	t=-1.692 Sig.=0.115
	on	t=1.062 Sig.=0.306	t=0.123 Sig.=0.904	t=1.007 Sig.=0.331	t=-1.692 Sig.=0.115

The statistic results demonstrate that for subjects speaking Hangzhou, Jinan and Zhenjiang dialect, their F$_0$ contours don't exhibit significant variations across the focused and unfocused conditions, in all focus positions and for all compositions of the focused words, except for only three instances of significant differences (Hangzhou, bi-syllabic word, post-focus compression; Jinan, tri-syllabic word, on-focus expansion; Zhenjiang, tri-syllabic word, pre-focus compression). However, for subjects speaking Beijing dialect, the presences of focus lay great influence on the post-focus regions with the F$_0$ contours significantly compressed (mono-syllabic word, t=-3.142, Sig.＜0.01; bi-syllabic word initially stressed, t=-2.313, Sig. ＜ 0.05; tri-syllabic word, t=-3.036, Sig. ＜ 0.01), whereas no significant changes on neither the pre-focus regions nor the on-focus word is observed. Considering the possibility that maxF$_0$ and minF$_0$ may be changed with mean F$_0$ remains. Paired-Sample t-test on maxF$_0$ and minF$_0$ was conducted and the results show no significant changes for all the four dialect speakers over the two focus conditions.

It is therefore evident from the statistical analysis that F$_0$ is not the most effective exhibitor of the presence of focus. Considering that duration also can be a cue of focus, a series of Paired Sample t-test is performed again to explore whether the parameter would be employed by the dialect speakers to signal focus in discourse. Measurements were conducted regarding of two focus positions (sentence-initial, sentence-final) and the regions before, on and after focus, respectively. Variables included are duration of the syllables bearing lexically stress in the on-focus words, and that of all syllables in the pre-focus and post-focus regions.

Table 3. Results of Paired Sample t-test on duration

Hangzhou		1S	2S_1	2S_2	3S
I	on	t=-0.879 Sig.=.444	t=-0.501 Sig.=0.651	t=-0.899 Sig.=0.435	t=-0.542 Sig.=0.626
	post	t=-2.691 Sig.=0.074	t=-5.443 **Sig.=0.012**	t=-0.633 Sig.=0.572	T=-0.552 Sig.=0.619
F	pre	t=-3.081 Sig.=0.054	t=-0.492 Sig.=0.656	t=-3.572 **Sig.=0.037**	t=-4.806 **Sig.=0.017**
	on	t=0.950 Sig.=0.412	t=0.446 Sig.=0.686	t=-3.802 **Sig.=0.032**	t=0.401 Sig.=0.715
Jinan		1S	2S	2S	3S
I	on	t=-0.813 Sig.=0.476		t=-1.368 Sig.=0.265	t=-1.313 Sig.=0.281
	post	t=-0.512 Sig.=0.644		t=-0.276 Sig.=0.800	t=0.278 Sig.=0.799
F	pre	t=-1.496 Sig.=0.232	t=-3.832 **Sig.=0.031**	t=1.331 Sig.=0.275	t=-1.769 Sig.=0.175
	on	t=0.629 Sig.=0.574	t=-0.329 Sig.=0.764	t=1.435 Sig.=0.247	t=-0.153 Sig.=0.888
Zhenjiang		1S	2S_1	2S_2	3S
I	on	t=-1.667 Sig.=0.194	t= Sig.=	t=3.059 Sig.=0.055	t=-2.643 Sig.=0.077
	post	t=-1.098 Sig.=0.352	t= Sig.=	t=-1.856 Sig.=0.160	t=-0.963 Sig.=0.407
F	pre	t=-2.050 Sig.=0.133	t=-0.325 Sig.=0.766	t=1.797 Sig.=0.170	t=-0.694 **Sig.=0.538**
	on	t=-1.706 Sig.=0.187	t=-0.961 Sig.=0.407	t=-1.275 Sig.=0.292	t=0.733 Sig.=0.517

Table 3 demonstrates that for Jinan and Zhenjiang speakers, duration is not significantly different from a focused condition to an unfocused one except for two instances in all focus positions and all focused word compositions. Comparatively, the effect of focus on duration is more systematic for Hangzhou subjects. For the bi-syllabic word with lexical stress on the second syllable and the tri-syllabic word stressed on the first syllable, the duration of pre-focus words is significantly shorter when the sentence is in the focused condition. And in such condition, for the mono-syllabic word, duration of pre-focus syllables is marginally shorter (t=-3.081, Sig.=0.054). The on-focus position, however, is not observed with extension of duration in all focus positions and all focus compositions for all three dialects.

4. DISCUSSION AND CONCLUSIONS

According to the statistical analysis, the acoustic differences between on-focus and neutral focus in second language English produced by Beijing speakers are evident: F_0 of the post-focus regions is significantly lower with focus than without, whereas no significant changes on the pre-focus or on-focus words are observed. However, no systematic F_0 variations from the focused and un-focused conditions are observed for Hangzhou, Jinan and Zhenjiang speakers. The exhibition of post-focus compression (PFC) in Beijing subjects' English utterances may be transferred from their native language of Mandarin Chinese, in which PFC is also identified [2]. However, for Hangzhou, Jinan and Zhenjiang learners of English, no post-focus compression is observed in their production of second language English. The conclusion that Zhenjiang speakers have no post-focus compression is contradictory with the findings in [7], which claims the existence of PFC in Zhenjiang speakers' English production. Conclusions drawn in the latter study is not so convincing as they are in the present study considering that the conclusion is drawn from visually inspection of the F_0 contours without statistical support and the values used to plot the graphs are fundamental frequency in Hz rather than in semitones. The absence of significant F_0 variations in the English produced by speakers with Hangzhou, Jinan and Zhenjiang dialects is probably due to that such changes on F_0 with and without focus may be also absent in the dialects. It also can be that such F_0 variations exist in dialects but are not easily transferred into second language English as it is for non-Chinese-heritage American learners' Mandarin [6]. Further studies are needed to explore this possibility.

Considering that acoustic parameters other than F_0 may correlate with focus in learners' English utterances, syllable duration before, at and after the focus was examined and the statistic results show that for HZ speakers, duration is the main acoustic correlates of prosodic focus and it is decreased significantly in the pre-focus regions for nearly all word compositions that have been included. While no such significant changes on duration exhibits in neither Jinan nor Zhenjiang subjects' utterance. This finding suggests one possibility that in Hangzhou dialect, prosodic focus in discourse is realized by shortening the duration of syllable preceding the focus. Studies are expected to verify this hypothesis. While for both Jinan and Zhenjiang dialect, neither duration nor F_0 is identified as the correlate of focus. However, here we present one plausible explanation for why Hangzhou

speakers realize focus through varying duration. There are seven types of tones in Hangzhou dialect. If described in five degrees with 1 representing the lowest point and 5 the highest, they are 435, 213, 53, 24, 13, 4 and 2 [8]. Noticeably, two of them, 435 and 213, fall into the category of *shangheng*. According existing studies, *shangsheng* tones are the longest among the four tone types, *yinping, yangping, shangsheng* and *qusheng*. Considering that the effect of focus is to make the great greater and the small smaller, as exemplified by the pitch range expansion on the on-focus words and compression in the post-focus regions in English, it is reasonable to hypothesize that syllables bearing *shangsheng* in dialects will be prolonged if put under focus, while the syllables around the focus will be shortened. To verify this, further studies are expected.

5. CONCLUSION

This study compared how Chinese learners with different dialect backgrounds realize focus in second language English. Parameters examined were mean F_0, maxF_0, minF_0 and duration. The results of statistical analysis demonstrate that F_0 is the closest correlate with focus in Beijing speakers' performance, mainly going through a process of going down after focus, without significant changes in the pre-focus and on-focus position. The statistics also shows that duration in the pre-focus regions is significantly reduced in Hangzhou speakers' utterance. However, none of the two parameters is significantly varied with the presence of focus in neither Jinan speakers nor Zhenjiang speakers' performance.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Xu, Y., "Phonetic Realization of Focus in English Declarative Intonation". *Journal of Phonetics, 33,* 159-197, 2005.

[2] Xu, Y., "Effects of Tone and Focus on the Formation and Alignment of F_0 Contours". *Journal of Phonetics, 27,* 55-105, 1999.

[3] Thorsen, N. G. (1978). An acoustical investigation of Danish intonation. Journal of Phonetics, 6, 151-175.

[4] Wu, W. L. & Xu, Y. "Prosodic Focus in Hong Kong Cantonese without Post-focus Compression.", Speechprosody, Chicago, 2010.

[5] Chen, S-w., Wang, B. & Xu, Y., "Closely related languages, different ways of realizing focus.", Interspeech, Brighton, UK: 1007-1010, 2009.

[6] Chen, Y., Guion-Anderson, S., Xu, Y. "Post-Focus Compression in Second Language Mandarin", Speechprosody, 2012.

[7] Hu, N., Jia, Y., Liu, B., "Phonetic and Phonological Realization of Narrow Focus in English Declarative Sentences by Zhenjiang EFL Learners", Speechprosody, 2012.

[8] Wang, Q., "The phonological system of Hangzhou dialect". Journal of Tsinghua University, 14, 1999.

[This paper was published in OC-COCOSDA,2012.]