The Realization and Perception of Narrow Focus in English Sentences by Cantonese EFL learners

Chong Cao* and Aijun Li†
*Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
E-mail: caochong2013@126.com
F-†Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
E-mail: liaj@cass.org..cn Tel: +86-10-85195377

Abstract— **Previous studies have investigated the acquisition of narrow focus by Chinese speakers of English from various dialect regions. While few research paid attention to the Cantonese EFL (i.e., English as a Foreign Language) learners. Given the similarities and differences in intonation patterns between Cantonese and English, the present study aims to provide an empirical research to address the acquisition of narrow focus by Cantonese EFL learners from the perspective of production and perception. Production results showed that: all Cantonese EFL learners applied the falling boundary tone to realize the imperative intonation. But a dominant number of learners produced the target sentences with the nucleus assigned at words other than the narrow focus words designed. And among the sentences whose nucleus fell on the narrow focus words, the nuclear accent was mainly marked by H*, followed by L+H*. In terms of perception, Cantonese EFL learners could identify the accented words accurately. Specifically, the perception accuracy of nuclear accent was about 94% (i.e., 94% for H* pitch accent, and 93.75% for L+H*), and there was no significant differences in the perception accuracy of different types of pitch accents. Comparison of results in the two experiments did not find strong correlations between the production and perception. Namely, although Cantonese EFL learners could perceive L+H* pitch accent, they did not employ L+H* dominantly in speech production.**

I. INTRODUCTION

Focus usually refers to the most informative part of an utterance, and is received the most prosodic prominence [1] [2]. It can be either broad or narrow. Broad focus means that a whole constituent (e.g., a noun phrase) or a whole sentence could be focused, while narrow focus usually refers to that one individual word in an utterance was assigned the focus [3]. This study will take the narrow focused sentences as its interest.

A substantial body of research has examined the prosodic realization of narrow focus in English. With respect to the phonetics, there was an increase in duration, fundamental frequency (i.e., f0) and intensity in focused constituents, as well as the post-focus compression (henceforth PFC) after the focused ones [4] [5] [6]. With respect to the phonology, it was commonly assumed that narrow focus was marked with a pitch accent [7]. For instance, focus in yes-no questions was associated with a low tone (i.e., L* pitch accent) or a low-rising tone (i.e., L*+H pitch accent) [8], while focus in declarative and imperative sentences was associated with a L+H* or a H* pitch accent [9].

Research on the acquisition of second language (i.e., L2) has suggested the effects of linguistic experience [10] [11] [12]. Chinese has been classified into seven groups of related languages, namely northern dialect, Wu, Min, Xiang, Gan, Kejia dialects and Cantonese [13]. Given the diversity of Chinese dialects as well as the great variance in phonetic inventories and intonation systems [14], there is growing interest in recruiting subjects with the same language background in the research of L2 acquisition of prosody focus. For instance, Jia and Li [15] found that English learners from Wu dialect regions exhibited problems on accent realization. Specifically, they could not realize f0 raising on the focused word and f0 compression after the accented syllables. Lian et al. [16] investigated the phonetic realization of focus in yes-no sentences produced by English learners from northern dialect, and found that compared with native speakers whose f0 contours were suppressed at the on-focus positions, learners showed expanded f0 contours. It is noteworthy that there have been few investigations of realization of narrow focus by Cantonese EFL learners. In this study, Chinese EFL learners speaking Cantonese dialect are taken as the subjects and their prosodic realization of narrow focus in English imperative sentences is examined.

As for Cantonese, a tonal language, contains six lexical tones which are commonly labeled Tones 1 to 6 [17]. Previous research has demonstrated that the intonation pattern of Cantonese was similar to English. Generally speaking, declarative sentences and imperative sentences have a falling f0 contour, while yes-no questions have a rising f0 contour [18] [19] . However, the two languages differed in how they realized prosodic focus. As mentioned above, English has on-focus expansion (i.e., increased f0, duration and intensity in focused constituents) and PFC (i.e., lower f0 and weaker intensity after the focused ones). While in Cantonese, the most reliable cue to narrow focus was duration, followed by intensity, f0 was a relatively weak cue [20]. In addition, PFC was not observed in Cantonese [21]. Given the similarities and differences in intonation patterns between Cantonese and English, this study aims to find out the acquisition performance of prosody focus of Cantonese learners of English.

Several quantitative studies have investigated the

978-616-590-477-3 ©2022 APSIPA

acquisition performance on intonation by Cantonese EFL learners. For example, Setter [22] examined the duration of stressed and unstressed syllables produced by Cantonese speakers of English, and reported that duration of unstressed syllables was almost 50% longer than that of native speakers. Besides, Cantonese English learners • ' speech was always perceived as flat in intonation, probably due to more stressed syllables, less f0 fluctuations and narrower pitch range than native speakers. Szeto [23] verified two characteristics of intonation in Cantonese English, namely rigid pitch assignment at syllabic level and rapid or monotonous pitch movement between successive syllables.

Noticeably, previous studies on the acquisition of L2 prosody focus by Cantonese EFL learners have been interested in phonetic analysis, such as average f0, the location of f0 peak, duration and the magnitude of intensity drop, while largely overlooked the phonological representation. Besides, most research was limited to the realization, while neglecting the perception as well as the relationship between the production and perception.

Therefore, more work is needed to address these under-investigated aspects in the existing literature. This study expanded on previous research in two ways. Firstly, this research aims to demonstrate the phonological representation of prosody focus by Cantonese EFL learners. Secondly, the study extended the body of research from investigating production only to investigating both the production and perception, as well as the relationship between production and perception.

The present research specifically addresses three questions:

(1) How is the narrow focus realized phonologically by Cantonese EFL learners?

(2) Could Cantonese EFL learners perceive the narrow focus?

(3) What is the production-perception link in Cantonese EFL learners?

II. PRODUCTION EXPERIMENT

A. Participants

Forty Cantonese EFL learners were paid to participate in the experiment, and all were undergraduates at Universities in Guangzhou city. All participants spoke Cantonese as their first language (i.e., L1), and have spoken English (i.e., L2) for over five years. All of them passed College English Test (i.e., CET-6), so their English proficiency could be regarded as at a similar level. None of them reported to have hearing or language impairments.

B. Materials

Target phrases consisted of noun phrases (henceforth NPs) containing one of eight adjectives describing colors (i.e., black, white, brown, gray, blue, green, red, and pink) combined with one noun (e.g., ball, tree, star). Both the adjectives and nouns were monosyllables. The target phrases were produced in a carrier sentence as shown below. The adjective in the target NPs was a contrast to another adjective in the context NPs, while the noun was kept the same (e.g., put the green ball over the brown ball), marking the adjective in the target NP as the contrastive focus.

Put the (adjective) (noun) over the (adjective) (noun)
CONTEXT NP TARGET NP

Participants sat in front of a computer screen, and saw slides presenting the exact position of colored pictures, as illustrated in Fig. 1. The numbers represent the group order. Participants produced a total of 6 sentences based on Fig. 1, of which 3 were target sentences, as shown below.

(1) *Put the green ball over the brown ball*. Corresponding to the first group of pictures;

(2) *Put the black tree over the pink tree*. Corresponding to the third group of pictures;

(3) *Put the brown key over the green key*. Corresponding to the fifth group of pictures.

The remaining three sentences were served as fillers. In order to prevent participants from developing expectations that sentences containing the same nouns were likely to be targets, fillers that consisted of different colored pictures were constructed.

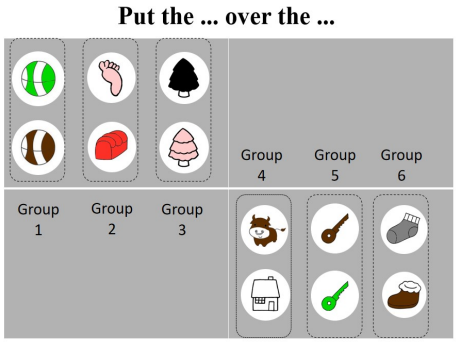

Fig. 1 An example slide for the production task.

The experimenter also saw slides which showed the same pictures as participants' slides through another computer screen. But pictures in the experimenter's slides were placed disorderly. Participants were required to use the carrier sentence to describe the color and shape of pictures, and to instruct the experimenter to move the pictures making the slide match the participants' slide.

Each participant was required to give instructions according to 20 slides and produced a total of 104 sentences, including 52 target sentences. A total of 4160 sentences were collected, including 2080 target sentences (40 participants × 52 target sentences = 2080). However, there were 10 participants' data were damaged and were discarded, resulting in a total of 1560 sentences (30 participants × 52 target sentences = 1560) for acoustic analyses.

C. Recording

To making sure that participants understood all names of pictures in English, they were given pictures and the corresponding nouns a week in advance to familiarize themselves with the words. The recording was carried out in

a sound-attenuated booth, with a "xPerception" software, an Lexicon sound card, and a Shure SM 58 microphone. Sentences were recorded with a sampling rate of 22050 Hz and 16-bit accuracy rate in mono channel. We saved all recordings as *.wav* files on a computer for further analysis. The production task took approximately 35 minutes.

D. Annotation and data extraction

The recordings were automatically segmented at first by an automatic segmentation software. And the corresponding boundaries were manually checked and corrected. The pitch accents of target words (i.e., adjectives in the target NPs) were phonologically annotated according to a labeling system ToBI (i.e., Tone and Break Indices), which used five types of pitch accents to describe the intonation, namely, H*, L*, L+H*, L*+H, and H+!H*. The annotation was made up of four tiers, as illustrated in Fig. 2, the first two tiers were determined by using an automatic segmentation software and manual correction, the last two tiers were annotated by three experienced annotators manually.

I) Word tier: the orthographic tier.

II) Phone tier: the transcription of phonemes.

III) Phonological tier: the phonological description of target word.

IV) Break Index tier: marketing the positions of intermediate phrase (i.e., 3) and intonation phrase (i.e., 4).

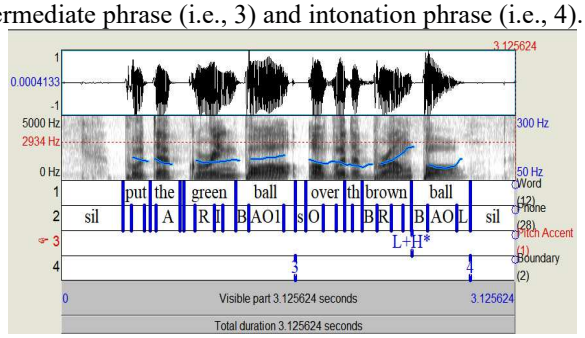

Fig. 2 An annotation sample for one of target sentences "Put the green ball over the brown ball".

F0 of the voiced part of syllables was extracted by a Praat script at 11 equal proportional intervals. To exclude the physiological influences of factors as gender and age, the original f0 data were normalized based on a z-score formula:

$$z = (x - \mu)/\ \sigma \qquad 1)$$

Where "x" indicates the raw f0 value, "μ" and " σ" refer to the mean value and standard deviation of f0 values for the speaker, respectively, "z" is the normalized f0 value.

E. Results

First of all, we found that all learners applied the falling boundary tone to realize the imperative intonation, which was the same as native English speakers. Then, the intonation patterns at the narrow focus words (i.e., target words within the target NPs) produced by Cantonese EFL learners were analyzed. Before analysis, the accent position of each sentence was closely examined in case that participants may produce the sentences with the nucleus assigned at words other than the target words, and those failed were excluded from phonological analysis. It turned out that a dominant number of Cantonese EFL learners produced target sentences with the nucleus assigned at words other than the narrow focus words. Only 18.21% of target sentences showed the appropriate nucleus position.

Table 1: Pitch accent type on target words employed by Cantonese EFL learners

Pitch Accent	Number
H*	222
L+H*	62
Total	284

Table 1 shows that Cantonese EFL learners adopted two types of pitch accent to realized the narrow focus, namely, H* and L+H* accent type. The dominant number of target sentences were realized as H*, while only a small number of sentences were marked with L+H*.

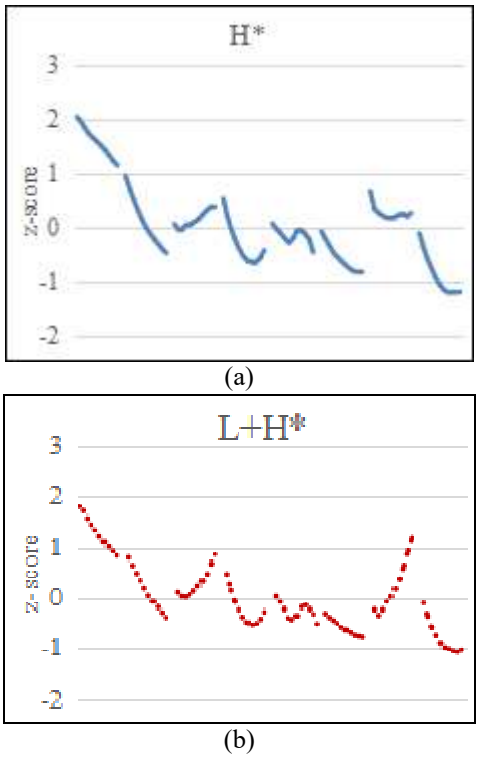

Fig. 3 Mean f0 contours of target sentences.

The average time normalized f0 contours of target sentences were shown in Fig. 3 (a-b). The X-axis represents the materials read by speakers (e.g., *Put the green ball over the brown ball*), the Y-axis represents the average normalized z-score value of f0. The blue solid curves represent pitch contour of target sentences whose narrow focus words were realized as H*, and the red dotted lines for L+H*. F0 contours of narrow focus words are framed in black. The gaps in the curves represent syllable boundaries.

Through visual inspection of Fig. 3(a), the difference between target word and its surrounding words lies in the pitch register. The narrow focus word demonstrates a higher pitch register. Through visual inspection of Fig. 3(b), the pitch contour of target word is raised obviously. As for the constituents after the focused word (i.e., the noun within the target NP), it can been seen from Fig. 3 that both groups show obvious f0 lowering independent of phonological realization of target words.

III. Perception Experiment

A. Participants

Forty participants of the production experiment also participated the perception experiment.

B. Materials

Experimental stimuli consisted of 20 declarative sentences with a narrow focus. Two accent types were tested, namely, L+H* and H*. Each accent type accounted for half number of the total experiment stimuli. In order to avoid the monotony of the experiment, another 10 declarative sentences with broad focus were also designed.

The stimuli were recorded by a prosodically trained male speaker of American English at 22050 Hz sampling rate. The speaker was required to produce target sentences in a natural way after reading a leading sentence, and to produce pitch contours of target words on request. For example:

The narrow focus word was marked with L+H* accent type:

Leading sentence: Bill played the guitar in a class at school. What about Tina?

Target sentence: Tina played the ***DRUM*** in a class at school.

The bold and italic word " drum" was the target word. The pitch contour of the illustrative sentence was shown in Fig. 4.

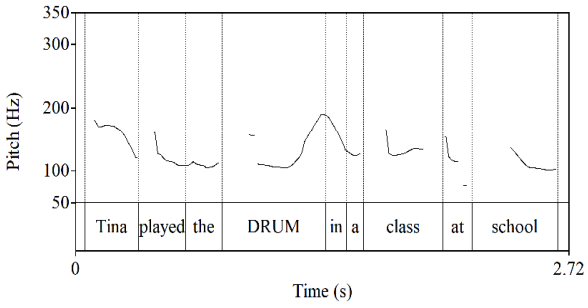

Fig. 4 The example sentence with L+H* pitch accent.

The narrow focus word was marked with H* accent type:

Leading sentence: Barbara teaches geography in the university. What about Susan?

Target sentence: Susan teaches ***PHYSICS*** in the university.

The bold and italic word " physics" was the target word. The pitch contour of the illustrative sentence was shown in Fig. 5.

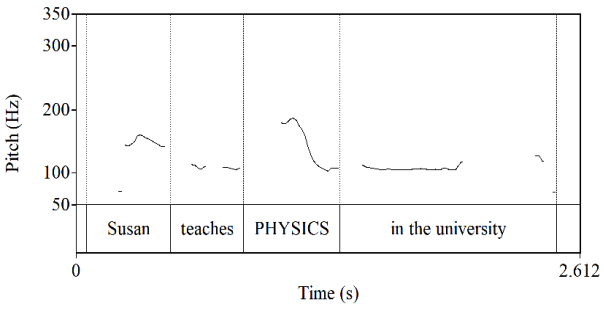

Fig. 5 The example sentence with H* pitch accent.

C. Procedure

The perception experiment was conducted in a quite room. 30 sentences were played to the participants in random order through headphones. The text of the sentence being played was presented on the computer screen. Participants were required to click on the word(s) with the mouse that were accented they thought. If they thought there was no accented words in the sentence, they had to choose the punctuation at the end of the sentence, which aimed to prevent participants from making their choices before the sentence was completely played. Once a response was collected, participants clicked on the " OK" button at the bottom of the computer screen, then the next stimulus was presented automatically.

D. Results

As long as participants chose the narrow focus word of each sentence as the accented word, "1" point would be counted. On the contrary, "0" point would be counted. For each participant, the maximum score of each type of pitch accent was 10 points. Finally, the scores of the same pitch accent type of all participants were accumulated. The results were shown in Fig. 6. The X-axis represents the type of pitch accent, the Y-axis represents the total points for each type of pitch accent.

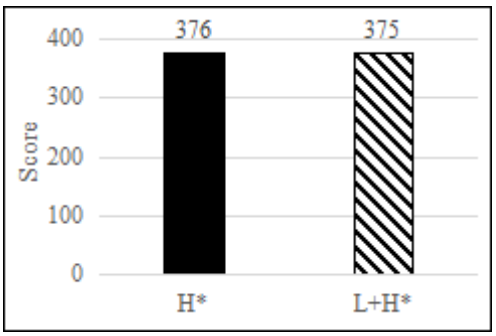

Fig. 6 The perception score of L+H* and H* type.

As to the perception of H* accent type, Cantonese speakers obtained a total of 376 points, averaged 94% correct identification accuracy across 400 sentences (10 target sentences marked with H* × 40 participants = 400) , ranging from the highest accuracy rate 100% to the lowest accuracy rate 90%. As to the perception of L+H* accent type, Cantonese speakers obtained a total of 375 points, averaged 93.75% correct identification accuracy across 400 sentences (10 target sentences marked with L+H* × 40 participants = 400), ranging from the highest accuracy rate 100% to the lowest accuracy rate 90%.

Generally, participants performed the perception task quite accurately. As indexed by the perception accuracy, paired t-tests showed no significant differences between two accent types ($p > 0.05$).

IV. Discussion and Conclusions

Two experiments were designed to investigate the research questions concerning the phonological realization and perception of narrow focus by Cantonese EFL learners, as well as the relationship between the production and perception.

In the production task, from a perspective of phonology, the pitch accent patterns of narrow focus adopted by Cantonese speakers were H* and L+H* which were also employed by native English speakers. But these two types differed from each other in the frequency of occurrence compared with native English speakers. Specifically, the accent type L+H* was hypothesized to be the most appropriate marker of the contrastive focus for native English speakers, while the accent type H* was less acceptable [9]. In contrast, the dominant type employed by Cantonese EFL learners was H* rather than L+H*. As illustrated in the introduction, Cantonese speakers rarely used the acoustic correlate f0 to perform the narrow focus in their L1 production, instead, they preferred to use the duration to realize the narrow focus. In this case, it could be inferred that the phonological realization of narrow focus in L2 (i.e., English) may not be affected by L1 (i.e., Cantonese).

In the perception experiment, Cantonese EFL learners could distinguish accent type H* accurately, as well as the accent type L+H*. Moreover, participants exhibited no differences in the response accuracy of H* and L+H*, which indicated that Cantonese speakers could make use of f0 cues to distinguish accented words during the perception process of narrow focus. Comparison of the results in the two experiments suggested that learners who used prosodic cues in perception may not use them in production, which further suggested that there was no strong correlation between the production and perception.

In summary, the research investigated the realization and perception of narrow focus in English sentences by Cantonese EFL learners. The major findings of this study were as follows: with respect to the production, all Cantonese EFL learners employed the falling boundary tone to realize the imperative intonation. But a dominant number of Cantonese EFL learners failed to realize the nuclear accent on the narrow focus words, only 18.21% of target sentences showed the appropriate nucleus position. And among these sentences, the types of pitch accent of narrow focus employed by Cantonese speakers were H* and L+H*, and the former was the dominant one. In terms of perception, Cantonese speakers perceived accented words accurately, no matter the accented words were marked by H* or L+H* pitch accent. The study did not find strong correlations between the production and perception of narrow focus.

Acknowledgment

This work is financially supported by National Major Social Sciences Foundation of China (15ZDB103), and China Postdoctoral Science Foundation funded project (2020T130718 and 2020M680811).

References

[1] D. Bolinger, “Contrastive accent and contrastive stress,” *Language*, vol. 37, pp.83-96, 1961.

[2] C. Gussenhoven, “Sentence accents and argument structure,” In I. M. Roca, pp. 79-106, 1992.

[3] D. R. Ladd, “Stylized intonation,” *Language*, pp. 517-540,1978.

[4] Y. Xu and C. X. Xu, “Phonetic realization of focus in English declarative intonation,” *Journal of Phonetics*, vol. 33, no.2, pp. 159-197, 2005.

[5] Y. Lee, “*Prosodic focus within and across languages*,” Doctoral dissertation, University of Pennsylvania, 2015.

[6] W. E. Cooper, S. J. Eady, and P. R. Mueller, “Acoustical Aspects of Contrastive Stress in Question-Answer Contexts,” *Journal of the Acoustical Society of America*, vol. 77, pp. 2142-2156, 1985.

[7] C. Gussenhoven, “Focus, mode and the nucleus,” *Journal of Linguistics*, pp. 377-417, 1983.

[8] M. Beckman and G. Ayers, “*Guidelines for ToBI labeling*,”. Ohio State University, 1997.

[9] D. G. Watson, M. K. Tanenhaus, and C. A. Gunlogson, “Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*,” *Cognitive Science*, vol. 32, pp. 1232-1244, 2009.

[10] J. Liu, Y. Xu, and Y. Lee, “Post-focus compression is not automatically transferred from Korean to L2 English,” *Journal of Phonetics and Speech Sciences*, vol. 11, no.2, pp. 15-21, 2019.

[11] C. Shang and Y. H. Bu, “An error analysis of English vowels based on experimental phonetics,” *Journal of Tianshui Normal University*, vol. 6, pp. 135-137, 2009.

[12] X. J. Yang, “A study on the negative transfer of English dialect in the Eastern region of Inner Mongolia,” *Journal of Language and Literature Studies*, vol. 1, pp. 64-65, 2011.

[13] B. R. Huang and X. D. Liao, “*Modern Chinese*,” Beijing: Higher Education Press, 2010.

[14] Y. Li, “Timing and Melody: *An acoustic study of rhythmic patterns of Chinese dialects*,” Ph.D. Dissertation. University of Victoria. 2015.

[15] Y. Jia and A. Li, “Phonetic Realization of Accent from Chinese English Learners in Various Dialectal Regions,” *in Proceedings of ISCSLP*, 2012.

[16] X. Lian, W. Duan, Y. Jia, and A. Roeck, “Phonetic Realization of Narrow Focus by Beijing EFL learners in English Yes-No Sentence,” *in Proceeding of O-COCOSDA*, 2013.

[17] E. Zee, "Chinese (Hong Kong Cantonese) in the international phonetic association," *Handbook of the international Phonetic Association*, pp. 58-60, Cambridge University Press. 1999.
[18] C. Y. F. Chang, "Intonation in Cantones,". Muchen: LINCOM Europa. 2003
[19] T. J. Vance, "An experimental investigation of tone and intonation in Cantonese," *Phonetica*, pp. 368-392, 1976.
[20] C. Y. Ge, "*A Study of Chinese intonation across dialects*," Ph. D. Dissertation, University of Chinese Academy of Social Sciences, 2020.
[21] Y. Xu, S. Chen, and B. Wang, "Prosodic focus with and without post-focus compression: A typological divide within the same language family?" *The Linguistic Review*, vol. 29, no. 1, pp. 131-147, 2012.
[22] M. Setter, "A comparison of speech rhythm in British and Hong Kong English," *in Proceedings of the 15th International Congress of Phonetic Sciences*, 2003.
[23] W. M. Szeto, "An Experimental Study of Intonation in Cantonese English," 2003.