Prosodic Realization of Focus in English by Bidialectal Mandarin Speakers

Jiajing Zhang, Ying Chen* and Jie Cui*
Nanjing University of Science and Technology, Nanjing, Jiangsu, China
E-mail: iris_zhang@njust.edu.cn Tel: +86-15996310199
*E-mail: ychen@njust.edu.cn Tel: +86-17714341101
*E-mail: jiecui@njust.edu.cn Tel: +86-13140745799

***Abstract*—This study was designed to explore the prosodic patterns of focus in English by bidialectal Mandarin speakers. One learner group speaks Nanjing Mandarin as first dialect (D1) and standard Mandarin as second dialect (D2), and the other learner group speaks Changchun Mandarin as D1 and standard Mandarin as D2. This paper compares their prosodic outcome of focus realization in English in a production experiment. Results indicate that both Changchun and Nanjing bidialectal speakers produced clear in-focus expansion of duration, pitch and intensity and post-focus compression (PFC) of pitch and intensity yet were not able to acquire native-like patterns of PFC in English. Although these two groups' D1s are different dialects of Mandarin, they produced statistically similar patterns of prosodic focus in L2 English. These findings provide further support for the claim that PFC cannot be easily transferred cross-linguistically [11, 14, 15, 17, 18] despite its existence in both dialects of learners' L1 and their L2.**

I. INTRODUCTION

Prosody refers to the suprasegmental features in an utterance. Its acoustic parameters include duration, pitch (F0), intensity, etc. Focus is used to highlight a certain part of an utterance in a certain context, which in many cases, in addition to morphosyntactic means, can be realized by means of prosody [23]. Therefore, prosodic change plays an essential role in the realization of focus. Reference [20] proposed main features of prosodic change: an increase in F0, intensity, and duration in focused components, a decrease of F0 and intensity in post-focus components (post-focus compression), no consistent prosodic change in pre-focus components. Post-focus compression (PFC) was found in many languages, including Beijing Mandarin [20] and English [19], whereas PFC does not exist in some languages, such as Southern Min [11, 23] and Cantonese [14]. Some languages listed above are tonal languages and some are non-tonal, suggesting that whether a language is tonal seems not to determine whether it is a PFC language.

Previous studies regarding the phonetic realization of focus in various languages/dialects, mainly concluded that PFC could only be eliminated through language contact, but no language can obtain PFC through contact. Thus, PFC is not easy to transfer cross-linguistically [11, 14, 15, 17, 18].

Based on previous findings in the research of prosodic focus and L2 speech, this study was designed to explore the prosodic realization of focus in English by two groups of bidialectal Mandarin speakers. One learner group speaks Nanjing Mandarin as first dialect (D1) and standard Mandarin as second dialect (D2), and the other learner group speaks Changchun Mandarin as D1 and standard Mandarin as D2. Both Nanjing Mandarin and Changchun Mandarin are subdialects of Mandarin Chinese. They are both tonal but with different tonal inventories. Therefore, Mandarin is these two groups' L1 and they speak standard Mandarin as D2 in spite of different D1. References [16, 20, 21] found that standard Mandarin have PFC. A pilot study prior to the present experiment found that Nanjing Mandarin and Changchun Mandarin also have PFC. Therefore, the primary purpose of this study was to investigate how these bidialectal Mandarin learners realize English prosodic focus and whether L1, a PFC language with two dialects, facilitate the learners producing PFC in English. The following research questions are addressed:

(1) Do Changchun bidialectal and Nanjing bidialectal Mandarin learners of English produce significant in-focus expansion of duration, F0 and intensity in their L2 English?

(2) Do they produce significant post-focus compression (PFC) of F0 and intensity in L2 English?

(3) Do the prosodic focus patterns in Changchun bidialectal and Nanjing bidialectal learners' production of English differ from one another?

(4) Do Changchun bidialectal and Nanjing bidialectal learners produce native-like prosodic focus in L2 English?

II. METHODS

A. Participants

Ten Changchun (CC) Mandarin learners of English and ten Nanjing (NJ) Mandarin learners of English (five males and five females in each group) were recruited from Nanjing University of Science and Technology (NJUST). They are all undergraduate students, age from 18 to 24, born and raised in either Nanjing or Changchun. All participants studied English since elementary school and had passed China College English Test Band 4 (CET-4) by the time of the experiment.

B. Stimuli

The stimuli were borrowed from [15] and listed in Table I. There are two types of focus location in the target sentences — no-focus location for neutral focus, and focused locations for initial, medial and final foci. Each focused location contains five sentences that only differ in in-focus words, which vary in lexical stress. Each participant was instructed to use the same sentence to answer two prompt questions in English, one for neutral focus, the other for narrow focus. To elicit narrow focus of each sentence, different prompt questions were raised to educe focus in different locations: initial focus (the subject), medial focus (the verb) and final focus (the object).

TABLE I
PROMPT QUESTIONS AND ANSWERS, ADAPTED FROM [15]

Neutral focus	Question	What's the news?
	Answer	See initial, medial, final focus sentences below
Initial focus	Question	Who may marry Ray?
	Answer	Lee / Nina / Melanie / Marie / Ramona may marry Ray.
Medial focus	Question	What may Lee do to Norman?
	Answer	Lee may leave / marry / nominate / remind / remember Norman.
Final focus	Question	Who may Ray marry?
	Answer	Ray may marry Lee / Nina / Melanie / Marie / Ramona.

C. Recording

Recording took place in the soundproof booth in Language Cognition and Speech Science Lab at NJUST. The stimuli were presented in PowerPoint slides in three different predetermined orders. Each participant was instructed to click and listen to the prompt questions, which were recorded by a female native speaker of American English, and answer the questions with the target sentences provided with the written form in the slides. A total of 270 target sentences were produced by each participant (2 focus types × 3 focus locations × 5 sentences × 3 different orders × 3 repetitions in each order). A Marantz professional solid state recorder PMD661 and a Shure professional unidirectional head-worn dynamic microphone SM10A-CN were used for recording. The stimuli were recorded with a sampling rate of 44,100HZ and saved in an SD card. The production experiment was self-paced and lasted approximately 30 minutes for each participant.

D. Analysis

Data were analyzed by Praat version 5.3.65 [8] and ProsodyPro version 5.5.2 [22] with the pitch setting of 50 to 250 Hz for male participants, 100 to 350 Hz for female participants. Acoustic measures were conducted on the mean of the second repetition of three orders according to the convention of L2 speech research. The acoustic parameters in the current paper include mean duration, mean F0, mean intensity and time-normalized F0 at ten even-interval points in each syllable. The ten even-interval F0 values was extracted to track the time-normalized F0 trajectory for a direct observation of pitch change as a function of focus.

To examine the prosodic change of focus, the differential value of duration, F0 and intensity between the neutral focus and the initial, medial, final foci respectively were calculated by subtracting the mean value of each component in the no-focus sentence from the corresponding mean value in the focused sentences (initial, medial, and final respectively). Note that the mean F0 was converted from Hertz to semitones [st = 12 log2(F_0), where reference level is 1 Hz] because pitch in speech operates on a logarithmic scale just as in music [5, 6].

Therefore, pre-focus change was calculated on the mean value of the two syllables of "Lee may" in the medial-focus sentences and the four syllables of "Ray may marry" in the final-focus sentences minus their counterparts in the neutral-focus sentences. The in-focus change was computed on the prosodic values of the stressed syllable in each focused word minus its no-focus counterparts. Post-focus change was calculated on the mean value of the four syllables of "may marry Ray" in the initial-focus sentences and the two syllables of "Norman" in the medial-focus sentences minus that of the neutral-focus sentences.

Data of native American English speaker [15] with the same stimuli are adopted to compare with the production of NJ and CC bidialectal Chinese learners to examine whether the learners' production was native-like.

III. RESULTS

A. Time-normalized F0 Contours

Time-normalized F0 of all the target sentences were first plotted. Figures 1-3 display the time-normalized F0 contours with three focus locations (initial, medial, final) vs. their no-focus counterparts produced by speaker groups (CC and NJ learners of English). Each curve represents an average of each target sentence produced by the ten speakers of each group. The focused sentences are represented by dash curves, while the no-focus sentences by solid curves. Syllable boundaries are signified by vertical lines. Under each focus location, the number of syllables in the pre-focus and post-focus components are the same, and that in focused words differs due to the word stress placement. In order to observe PFC, the F0 contours are right-aligned in initial and medial focus location, whereas those in the final focus are left-aligned to observe pre-focus change.

Figures 1 and 2 display the time-normalized F0 contours in the initial focus location. The curves indicate that both CC speakers and NJ speakers expanded F0 in the focused words and compressed F0 in the post-focus constituents (thus produced PFC). Although NJ learners seemed to have more in-focus expansion and less post-focus compression than CC learners, their realization of PFC pattern was similar with each other. Besides, the in-focus expansion in F0 started from the stressed syllable in the in-focus word and ended with the unstressed syllable after that stressed syllable. This unstressed syllable was exactly where PFC began [15]. The contours show that this pattern of prosodic-focus realization exists in both CC and NJ speakers' English production, for instance, PFC started in "na" syllable in "Nina," "la" syllable in "Melanie," and "na" syllable in "Ramona."

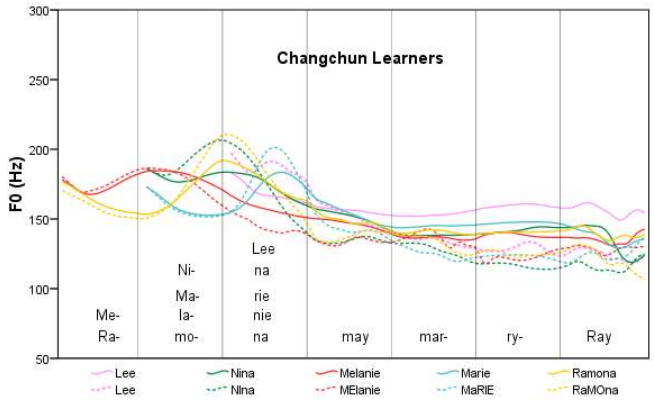

Fig. 1 Time-normalized F0 contours (Hz) in initial focus location by Changchun learners

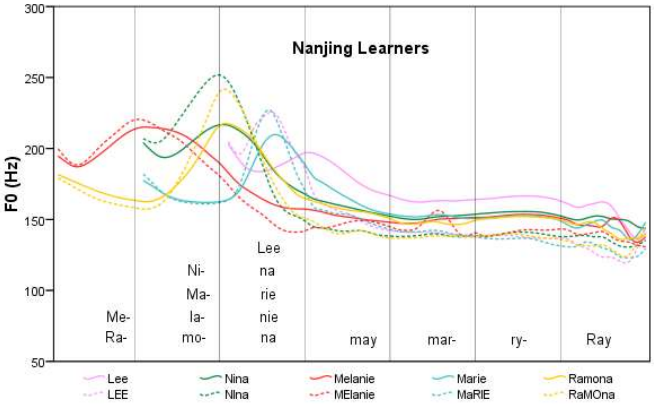

Fig. 2 Time-normalized F0 contours (Hz) in initial focus location by Nanjing learners

Figures 3 and 4 illustrate that both CC and NJ learners produced higher F0 in the focused syllable than that in the unfocused sentence, and there was more post-focus reduction than pre-focus reduction. The two learner groups demonstrate similar patterns of medial focus though NJ learners produced more noticeable in-focus expansion and PFC.

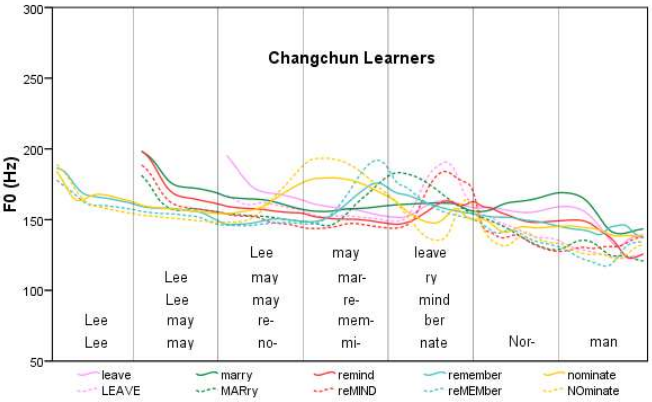

Fig. 3 Time-normalized F0 contours (Hz) in medial focus location by Changchun learners

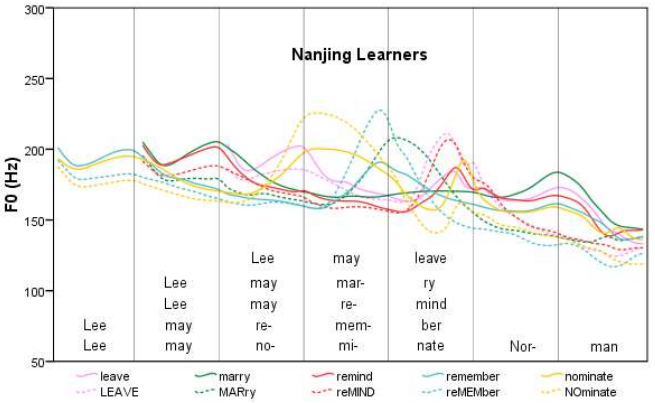

Fig. 4 Time-normalized F0 contours (Hz) in medial focus location by Nanjing learners

Figures 5 and 6 indicate that both CC and NJ learners produced similar noticeable F0 expansion in focused components but almost no F0 change in pre-focus components.

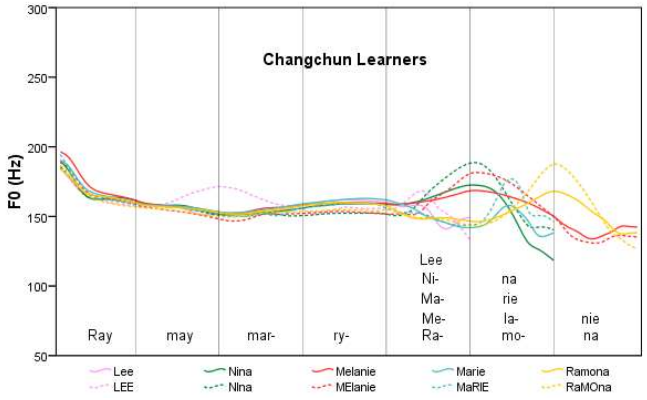

Fig. 5 Time-normalized F0 contours (Hz) in final focus location by Changchun learners

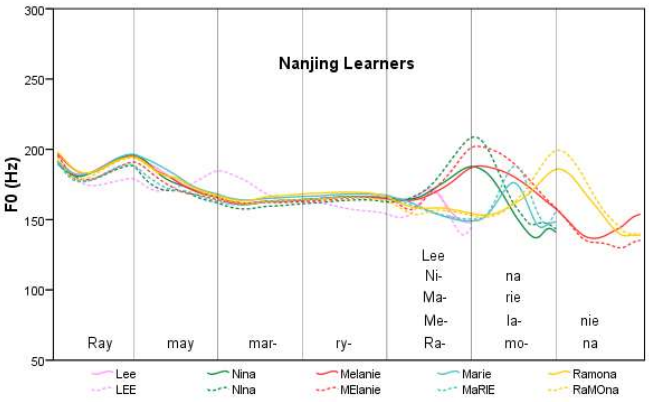

Fig. 6 Time-normalized F0 contours (Hz) in final focus location by Nanjing learners

B. Focus Change in Mean F0, Intensity, and Duration

Repeated measures ANOVAs were applied to analyze the mean F0, intensity, and duration change with a within-subjects factor—focus condition (three levels: pre-focus, in-focus, and post-focus) and a between-subject factor—learner group (two levels: Changchun and Nanjing).

The repeated measures ANOVA on mean F0 indicates no interaction between condition and group and the main effect of learner group also showed no significance. Only the main effect of focus condition ($F(1.42, 25.60) = 134.677, p < 0.001$)[1] was highly significant. To compare the degree of using F0 change to code focus at sentential level between CC and NJ learners, the magnitude on F0 differentials was computed. It respectively subtracted pre-focus change and post-focus change from in-focus change and was examined by independent-samples *t*-tests. Results show no significant difference between CC and NJ learners in the F0 magnitude between in-focus and pre-focus changes and between in-focus and post-focus changes.

To explore whether the similar mean F0 production by these two groups was native-like, native American English speakers' production [15] of the same stimuli was adopted to independent-samples *t*-tests to compare the difference in focus change of each focus condition between every two groups (among CC learners, NJ learners, and native speakers). Since there were three comparisons between each two groups, a Bonferroni correction was applied to adjust the threshold of *p*-value to 0.0167 (0.05/3 the same below) though this adjustment was considered too conservative [13]. As the ANOVA results predicted, the *t*-tests results show no statistical significance in all three focus condition changes between CC and NJ learners. A noticeable difference in post-focus change between CC learners and native speakers ($t(10.31) = 3.140, p = 0.010$)[2] was detected. A marginal significance in post-focus items between NJ learners and native speakers ($t(11.17) = 2.779$, $p = 0.018$) was also observed. The *t*-tests of the magnitude on F0 change demonstrate that the magnitude between in-focus and pre-focus changes and between in-focus and post-focus changes were significant between CC learners and native speakers ($t(18) = -4.015$, $p = 0.001$; $t(10.32) = -3.365$, $p = 0.007$), and between NJ learners and native speakers ($t(18) = -3.472$, $p = 0.003$; $t(10.92) = -2.985$, $p = 0.013$).

Figure 7 displays the mean F0 with standard error bar of these three groups. CC and NJ learners share a similar prosodic pattern, while native speakers show much larger in-focus expansion and post-focus compression than the other two groups. Undoubtedly the magnitude of F0 between in-focus and pre-focus changes and between in-focus and post-focus changes of native speakers was also larger than that of CC and NJ learners.

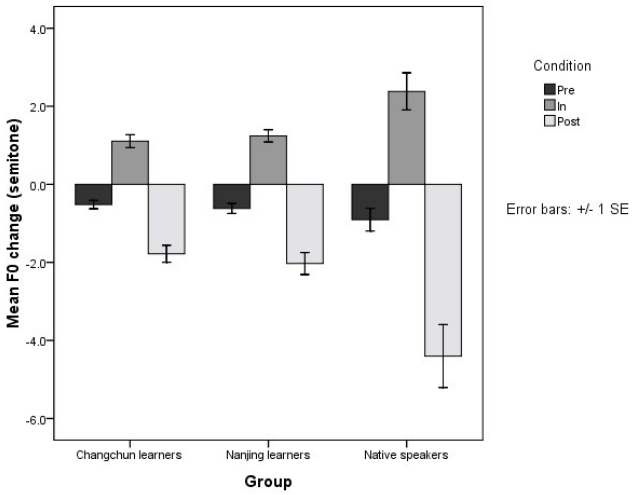

Fig. 7 Mean F0 change (semitone) by focus condition and participant group

Another repeated measures ANOVA on mean intensity in CC and NJ groups indicates no interaction of focus condition and group, no main effect of group and only significance in the main effect of condition ($F(2, 36) = 112.160$, $p < 0.001$). The independent-samples *t*-tests between each two groups show only the difference of post-focus change was significant between the CC learners and native speakers and also between NJ learners and native speakers ($t(13.28) = 3.418$, $p = 0.004$; $t(18) = 2.915$, $p = 0.009$). Furthermore, *t*-tests of magnitude between in-focus and post-focus differentials was found significantly different between CC learners and native speakers ($t(13.46) = -3.040$, $p = 0.009$), and a marginal significance in that between NJ learners and native speakers ($t(18) = -2.601$, $p = 0.018$).

Again, CC and NJ learners used almost the same pattern to code focus in intensity, but they all differed from native speakers' production. Figure 8 confirms these results that native speaker used more PFC of intensity to realize prosodic focus and the pre-focus and in-focus change showed almost no noticeable difference among the three groups.

[1] If the Mauchly's Test of Sphericity is satisfied, the variance is equal, and then results were reported under the premise of equal variance; if not, adjusted results were reported when variance is unequal.

[2] If the Levene's Test for Equality of Variances is satisfied in independent-samples *t*-test, and then results were reported under the premise of equal variance; if not, adjusted results were reported when variance is unequal.

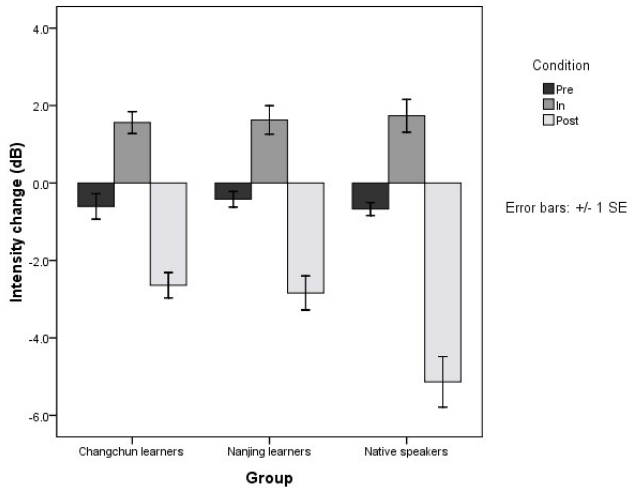

Fig. 8 Intensity change (dB) by focus condition and participant group

The repeated measures ANOVA in mean duration demonstrates similar results as in mean F0 and intensity, only the main effect of focus condition was significant between CC and NJ learners ($F(1.2, 21.83) = 50.966$, $p < 0.001$). The independent-samples *t*-tests show a marginal significance in post-focus change between CC learners and native speakers ($t(18) = 2.377$, $p = 0.029$), and between NJ learners and native speakers ($t(18) = 2.151$, $p = 0.045$). However, no significant difference was discovered in these three groups in the F0 magnitude between in-focus and pre-focus changes and between in-focus and post-focus changes.

Therefore, no significant difference existed in the duration change of CC and NJ groups, but neither group produced the native-like prosodic pattern in post-focus components in duration. Figure 9 indicate that all groups produced a clear in-focus expansion and almost no pre-focus change. The native speakers of American English produced more post-focus change compared to the two learner groups.

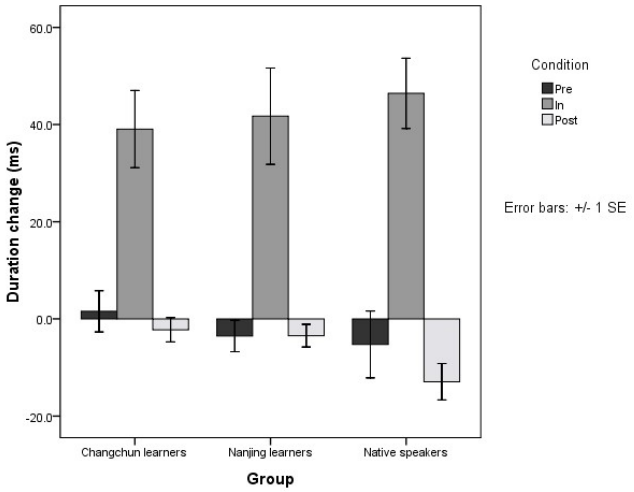

Fig. 9 Duration change (ms) by focus condition and participant group

IV. Discussion

The results of the experiment have answered the research questions. For the first and second questions, both CC and NJ bidialectal learners of English produced in-focus expansion of mean F0, intensity, and duration and post-focus compression of mean F0 and intensity. They also produced pre-focus compression of mean F0 and intensity in English, except CC learners slightly expanded on duration in their pre-focus component. This result is in line with the finding of [20] that the prosodic change in pre-focus component is not consistent.

According to the results of repeated measures ANOVAs and independent-samples *t*-tests, there was no significant difference in mean F0, intensity, and duration in the English production between CC and NJ bidialectal learners. This answers the third research question. In addition, no difference of the F0 magnitude between in-focus and pre-focus changes and between in-focus and post-focus changes in the English production between CC and NJ learners suggests that both learner groups share similar degree of acoustic cues to code focus in English though their D1s differs in tonal inventory.

Results of the comparison with native speakers' English production indicate that even though the two learner groups realized in-focus expansion and post-focus compression in English, they did not produce native-like PFC. This result provides an answer to the fourth research question and it concurs with the findings in [15] for L2 English and [20, 23] for Mandarin. Reference [15] examined the prosodic realization of monolingual Mandarin speakers with different experience of L2 English and found that even the advance Chinese learners of English with high L2 proficiency were not able to reach the native-like level in English PFC. PFC has been claimed not easy for L2 learners to transfer from one language to another [11, 14, 15, 17, 18]. In this study, all learners are bidialectal Mandarin speakers. Both their D1 and D2 are PFC languages in spite of different tonal inventories. However, even in such a bidialectal environment of L1, learners were not able to positively transfer PFC from their D1 or D2 into their L2 English to make it native-like. This finding reconfirms that PFC is not easy to transfer cross-linguistically. Furthermore, previous studies revealed that the presence/absence of lexical tone in L1 or L2 seem to have no effect on the realization of PFC [11, 14, 15, 18]. The results in the current study reconfirm that tonal system does not affect the acoustic realization of PFC regardless of tonal inventory.

Finally, prosody is an essential parameter in not only acoustic phonetics but also auditory phonetics. PFC has been proved to be a meaning-bearing prosodic feature [18]. When it comes to the speech production, PFC can be a criterion to evaluate learner's L2 speech proficiency [11, 18, 21]. That both bidialectal learner groups were not able to produce native-like PFC could be plausibly attributed to the fact that too much emphasis on the word stress in teaching English as a foreign language in China. Learners have attached more importance on the lexical-level stress than on the sentential-level prosody.

V. CONCLUSIONS

This study examined the prosodic realization of focusing English by bidialectal-Mandarin speakers from Changchun and Nanjing and explored the effect of bidialectalism in their L1 on prosodic focus in their L2. Three main findings were discovered.

First, both Changchun and Nanjing learners produced in-focus expansion and post-focus compression in their L2 English. Second, even though the two groups produced clear post-focus compression in English and the both dialects of their L1 has PFC, their PFC production in English was still not native-like, which reconfirmed that PFC is not easy to transfer. Finally, the tonal inventory in L1 seems not affect the realization of PFC in L2.

Based on the current study, future work may involve a fine-grained investigation of prosodic realization of focus in Changchun and Nanjing dialects for insight into the correlation between a tonal L1 and a non-tone L2. Furthermore, learners' L2 experience may be included as a function to the prosodic realization of focus in L2.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation of China (61573187) and the NJUST Key Bachelor's Thesis Grant.

REFERENCES

[1] B. Wang and Y. Xu, "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese," *Journal of Phonetics*, vol. 39, pp. 595-611, 2011.

[2] C. Féry and F. Kügler, "Pitch accent scaling on given, new, and focused constituents in German," *Journal of Phonetics*, vol. 36, pp. 0-703, 2008.

[3] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," Ed. *Language experience in second language speech learning: In honor of James Emil Flege.* Amsterdam: John Benjamins Publishing Company, pp. 13-34, 2007.

[4] C. T. Best, "A Direct Realist View of Cross-Language Speech Perception," Ed. *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 171-204, 1995.

[5] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scales," In *Proceedings of the 15th international congress of phonetic sciences*, Barcelona, Spain, vol. 771, pp. 774, 2003.

[6] H. Fujisaki, "Prosody, information, and modeling-with emphasis on tonal features of speech," *Workshop on Spoken Language Processing*, pp. 5-14, 2003.

[7] J. E. Flege, "Second-language speech learning: Theory, findings, and problems," Ed., *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 233-277, 1995.

[8] P. Boersma and D. Weenink, "Praat: doing phonetics by computer. Version 5.3.65. Online: http://www.fon.hum.uva.nl/praat/," 2014.

[9] R. Lado, "Linguistics across cultures: applied linguistics for Language teachers," University of Michigan Press, 1957.

[10] R. Wardhaugh, "The contrastive analysis hypothesis," *TESOL quarterly*, pp. 123-130, 1970.

[11] S. W. Chen, B. Wang and Y. Xu, "Closely related languages, different ways of realizing focus," *In Proceedings of Interspeech*, Brighton, UK, pp. 1007-1010, 2009.

[12] S. Zerbian, E. O. Aboh, K. Hartmann and M. Zimmermann, "Investigating prosodic focus marking in Northern Sotho," *Focus strategies in African languages: The interaction of focus and grammar in Niger-Congo and Afro-Asiatic*, pp. 55-79, 2007.

[13] T. V. Perneger, "What's wrong with Bonferroni adjustments," *Bmj*, *316*(7139), pp. 1236-1238, 1998.

[14] W. L. Wu and Y. Xu, "Prosodic focus in Hong Kong Cantonese without post-focus compression," *Speech Prosody 2010-Fifth International Conference*, Chicago, USA, pp.148-151, 2010.

[15] Y. Chen, "Post-focus compression in English by Mandarin learners," *Proceedings of the 18th International Congress of Phonetic Sciences*, ISBN 978-0-85261-941-4, pp. 1-5, 2015.

[16] Y. Chen, "Prosodic comparisons of two types of realization of focus in Mandarin," *Speech Prosody*, pp. 917-921, 2018.

[17] Y. Chen, S. Guion-Anderson and Y. Xu, "Post-focus compression in second language Mandarin," *Speech Prosody*, pp. 410-413, 2012.

[18] Y. Chen, Y. Xu and S. Guion-Anderson, "Prosodic realization of focus in a bilingual production of Southern Min and Mandarin," *Phonetica*, vol. 71, pp. 249-270, 2014.

[19] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics*, vol. 33, pp. 159-197, 2005.

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

[21] Y. Xu, "Post-focus compression: Cross-linguistic distribution and historical origin," *Proc. 17th ICPhS*, Hong Kong, pp. 152-155, 2011.

[22] Y. Xu, "ProsodyPro.praat. Version 5.5.2. Online: https://www.phon.ucl.ac.uk/home/yi/ProsodyPro/," 2014.

[23] Y. Xu, S. W. Chen and B. Wang, "Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?" *The Linguistic Review*, pp. 131-147, 2012.