

Too little, too late: A longitudinal study of English corrective focus by Mandarin speakers

Alex Hong-Lun Yeung, Hyunah Baek, Chikako Takahashi, Stephen Buttner,
Jiwon Hwang & Ellen Broselow*

Abstract. This study tracks the production of English corrective focus by Mandarin speakers (MS) living in the US over a two-year period. We show that the MS differed from English speakers (ES) in the alignment of the corrective focus pitch accent: while ES productions typically showed a pitch peak on the stressed syllable, followed by an abrupt fall, the pitch rise and fall for MS was later and less steep. While the MS productions became more English-like over time in some respects, the failure to correctly align pitch accent persisted over time. We argue that this misalignment of pitch peak cannot be attributed to a lack of sensitivity to English stress, but rather represents a common failure to master the complex timing patterns involved in synchronizing pitch, intensity, and duration cues with segmental structure in a second language.

Keywords. prosody; corrective focus; second language production; English; Mandarin

1. Introduction. English prosody plays an important role in signaling information structure. Learners of English have difficulty in producing native-like prosodic patterns, and this failure to master target language prosody has been linked to decreased intelligibility (e.g., Anderson-Hsieh & Koehler 1988, Hahn 2004, Munro & Derwing 1995, Sereno et al. 2016).

A study by Kao et al. (2016) revealed that L1-Mandarin speakers differed from L1-English speakers in their production of English corrective focus prosody, most notably in the location of the pitch rise and fall that is characteristic of this contour. Kao et al. found that in words with penultimate stress spoken with corrective focus, the timing of the stress peak differed in the native vs. non-native speaker productions: the English speaker showed a pitch peak coinciding with the main stressed syllable of the focused word and followed by a sharp drop in pitch in the post-stress syllable; the Mandarin speakers showed a later pitch peak, with high pitch often maintained throughout the post-stress syllable to the end of the word.

This paper reports on a follow up to the Kao et al. study investigating the hypothesis that this late alignment of the pitch peak reflects learners' lack of sensitivity to English lexical stress—due either to difficulty in perceiving the position of lexical stress, or to failure to understand the role of stress in the association of pitch accents with focused elements. We also investigated whether these same Mandarin speakers' productions changed in the course of two years spent in an English-speaking environment. We found that for the Mandarin speakers (as for the English speakers), the position of the pitch peak was in fact dependent on the position of the stressed syllable, indicating that the Mandarin speakers were able to perceive stress and were sensitive to the relationship between focus and stress. However, regardless of stress position, the Mandarin

* This material is based upon work supported by NSF under Grant # IBSS-1519908. Special thanks to Stephanie Chen and Sharron Huang for their help with segmentation. Authors: Alex Hong-Lun Yeung (alex.yeung@stonybrook.edu), Hyunah Baek (hyunah.baek@stonybrook.edu), Chikako Takahashi (chikako.takahashi@stonybrook.edu), Stephen Buttner (stephen.buttner@stonybrook.edu), Jiwon Hwang (jiwon.hwang@stonybrook.edu) & Ellen Broselow (ellen.broselow@stonybrook.edu), Stony Brook University.

speakers' pitch peaks and falls occurred later than for English speakers, and this delay did not change appreciably over time, although the overall shape of the pitch contours of the Mandarin speakers became more similar to those of the English speakers. We argue that this misalignment of pitch peak, which is found across various L1-L2 combinations, reflects a failure to master the complex timing patterns involved in synchronizing pitch, intensity, and duration cues with segmental structure in a second language.

2. Background. In English, corrective focus is typically realized by an increase in pitch, intensity, and duration of the main stressed syllable of the focused word, followed by a substantial decline in F0 and intensity immediately following that syllable (Pierrehumbert & Hirschberg 1990, Xu & Xu 2005). Although focus prosody in Mandarin may also involve expanded pitch range, greater intensity, and longer duration, Mandarin focus prosody differs from English focus prosody in critical ways. First, in Mandarin, the choice of acoustic cues and the pattern of F0 movements is crucially dependent on the properties of the lexical tone, involving expansion of the pitch range rather than a specific pitch pattern. For example, focus may be realized on the high level Tone 1 as a raising of the pitch, while focus on the falling-rising Tone 3 is manifested as lowered rather than raised pitch (Xu 1999, Ouyang & Kaiser 2015, Wang 2020). Both corrective and contrastive focus involve F0 expansion and increased duration, with the magnitude of these factors somewhat greater for corrective focus (Chen & Gussenhoven 2008, Ouyang & Kaiser 2015), and intensity is reportedly used as a cue for corrective but not new information focus (Ouyang & Kaiser 2015). Compression of post-focus pitch is dependent on tonal context and may be lacking in certain combinations of focus and post-focus tones (Chen 2010). Not surprisingly given the differences between the prosody of the two languages, previous work indicates that Mandarin learners of English may have difficulty producing English-like focus prosody (Chen et al. 2001, Chen 2015). Research has shown, furthermore, that post-focus compression does not transfer across languages in general (Chen et al. 2014, Wu & Chung 2011, Xu 2011).

Kao et al. (2016), a study comparing native and non-native productions of English corrective focus, found differences between the two groups. Sentences involving corrective focus were elicited from 18 native speakers of English and 18 L1-Mandarin speakers, all graduate students at a US university who had satisfied the university's requirement of a minimum TOEFL score of 90. This study revealed that the L1-Mandarin speakers tended to show a later pitch peak in focused words than the English speakers, coupled with a failure to consistently drop the pitch of the post-stress syllable.

One possible explanation of this pattern is that Mandarin speakers were less sensitive to lexical stress than English speakers. While the existence of word stress in Mandarin is controversial, it is clear that lexical stress, if it exists in Mandarin, plays a less important role than in English, opening the possibility that Mandarin speakers might be 'stress-deaf' in the terminology of Peperkamp and Dupoux (2002), who claimed that speakers for whom lexical stress is not contrastive are unable to hear stress differences in foreign language structures. The hypothesis that Mandarin speakers cannot correctly perceive English lexical stress is called into question by Lin et al. (2014), who report comparable accuracy rates in distinguishing pairs like *PERmit-perMIT*. However, Ou (2016) points out that work reporting comparable accuracy rates for Mandarin and English speakers in the perception of English stress typically presents stimuli with a declarative contour, in which stress and pitch peak coincide, and suggests that Mandarin speakers may simply be conflating stress and pitch peak rather than actually perceiving stress in an English-like manner. Ou found that when Mandarin speakers were presented in an ABX task

with words in which the stressed syllable carried a low pitch (because spoken with a yes-no question contour L* H-H%), even experienced learners had difficulty distinguishing iambic and trochaic stress, and inexperienced learners were unable to do so. Mandarin speakers' difficulties with English stress are not limited to perception; for example, Zhang, Nissen, and Francis (2008) show that Mandarin speakers did not consistently manipulate acoustic cues to stress in an English native-like way.

The Kao et al. study considered only words with a single stress pattern, penultimate stress. The goals of the current study were to determine (i) whether the location of the corrective focus pitch peak in the Mandarin speakers' productions varied with stress position--that is, whether the pitch peak came later in words with antepenultimate vs. penultimate stress; (ii) whether the location of the pitch peak in Mandarin speakers' productions differed from that of English speakers in the two stress contexts; and (iii) whether differences between native and non-native speaker realization of corrective focus persisted over the course of two years in an English-speaking context. We considered the following hypotheses:

- (1) **Stress-Insensitivity Hypothesis:** The Mandarin speakers align a pitch accent with the right edge (rather than the stressed syllable) of a focused word, due to either (i) a failure to correctly perceive lexical stress in English ('stress deafness' in the sense of Peperkamp & Dupoux 2002), or (ii) a failure to realize the role played by stress in anchoring pitch accents in English. In either case, the Mandarin speakers' late alignment of the focus pitch peak could be seen as a reflection of the native language strategy of expanding the pitch range of the entire focused word.
- (2) **Mistiming Hypothesis:** The Mandarin speakers do, like English speakers, recognize that the corrective focus pitch accent to a stressed syllable is anchored to a stressed syllable. They differ from English speakers only in the timing of that alignment.

The Stress-Insensitivity Hypothesis would predict that Mandarin speakers should show a pitch peak on (or near) the final syllable of the focused word, regardless of stress position, while the Mistiming Hypothesis predicts that the location of the pitch peak should be tied to the position of stress, though not necessarily realized with the same timing used by English speakers. We first compare Mandarin and English speakers' productions of corrective focus on words in the context of penultimate stress and consider how the Mandarin speakers' productions varied over time. We then compare the production of corrective focus in the context of antepenultimate stress.

3. Corrective focus with penultimate stress. We conducted a longitudinal study of 57 L1-Mandarin speakers' production of English focus prosody. Participants completed five recording sessions over a period of two years: during the first month of their residence in the US (Time 1), at the end of their first semester (Time 2), at the end of their second semester (Time 3), at the end of their third semester (Time 4), and at the end of their fourth semester (Time 5). The Mandarin speaking participants in the current study also took a test of English proficiency (the Versant English Test²) at the beginning of each recording session.

3.1. PARTICIPANTS. Fifty-seven Mandarin speakers (MS; 19 female, 38 male) took part in the experiment. All were students in graduate programs at Stony Brook University, with a mean age of 24.6 years (range: 20-38). They had begun their study of English at a mean age of 9.4 years

² Versant English Test scores "represent the ability to understand spoken English and speak it intelligibly at a native-like conversational pace on everyday topics" and range from 20 to 80 (Pearson Education 2019).

(range 4-16). In addition, 18 undergraduate native speakers of English (ES; 11 female, 7 male) participated in one session as a control group.

3.2. MATERIALS AND PROCEDURE. Eight short paragraphs were created to set up contexts to elicit corrective focus. In each session, the participants were first shown a passage on a computer screen, as in Figure 1, and asked to read it aloud.

1. One advantage of attending Stony Brook is that it is close to New York City.

2. There are several ways to travel to New York City from our campus.

3. If you don't want to worry about traffic on the roads, you can take the train.

4. But the price of a train ticket is twenty dollars, while a bus ticket is only eleven dollars.




Figure 1. Reading passage for *eleven dollars*

When the participants had finished reading the passage aloud, the experimenter asked two or three questions related to information from the passage. The last question was intended to elicit corrective focus intonation. The experimenters' questions and expected responses from the participants for the passage in Figure 1 are shown below; the target phrase containing corrective focus is marked in bold.

- (3) Experimenter: *Can you take the bus to New York City?*
Expected response: *Yes, you can take the bus to New York City.*
Experimenter: *What is the price of a bus ticket?*
Expected response: *The price of a bus ticket is eleven dollars.*
Experimenter: *Did you say the price of a bus ticket is fifteen dollars?*
(produced with broad focus, i.e., pitch peak on *dollars*.)
Expected response: *No, I said the price of a bus ticket is **ELEVEN** dollars.*

The experiment was conducted in a sound-treated room on the Stony Brook University campus. All elicited utterances were recorded using a Zoom H6 digital recorder and a SM10A-CN dynamic head-mounted microphone at a sampling rate of 44.1 kHz. The production task took approximately 15 minutes.

3.3. ANALYSIS. Following the recording, the target phrases designed to carry focus prosody were hand-segmented into syllables in Praat (Boersma & Weenink 2019). Then time-normalized pitch contours in each syllable were generated using pYAAPT, a python script for fundamental frequency tracking (Zahorian & Hu 2008), and intensity data were generated using ProsodyPro, a Praat script developed for large-scale analysis of speech prosody (Xu 2013). While pitch and intensity were measured in Hz and dB, respectively, both were converted into z scores for each utterance in order to control for individual variation. All statistical analyses were conducted with R (R Core Team 2019), using R package lme4 (Bates et al. 2015). The *p*-values were calculated using the lmerTest package (Kuznetsova et al. 2017).

3.4. RESULTS. Figure 2 shows the mean pitch contours of one of the target phrases with penultimate stress, *eleven dollars*, produced by the ES control group and by the MS shortly after their arrival in the US (Time 1). The contours are time-normalized so that each unit on the x axis

represents 1/10 of a syllable. As expected, the ES contour shows a pitch peak aligned with the stressed syllable (*le*) of the focused word, followed by a significant pitch drop on the post-stress syllable (*ven*). In contrast, the MS contour at Time 1 shows a plateau beginning near the right edge of the stressed syllable (*le*) and continuing through the following syllable (*ven*), with a pitch drop beginning well into the final (post-stress) syllable.

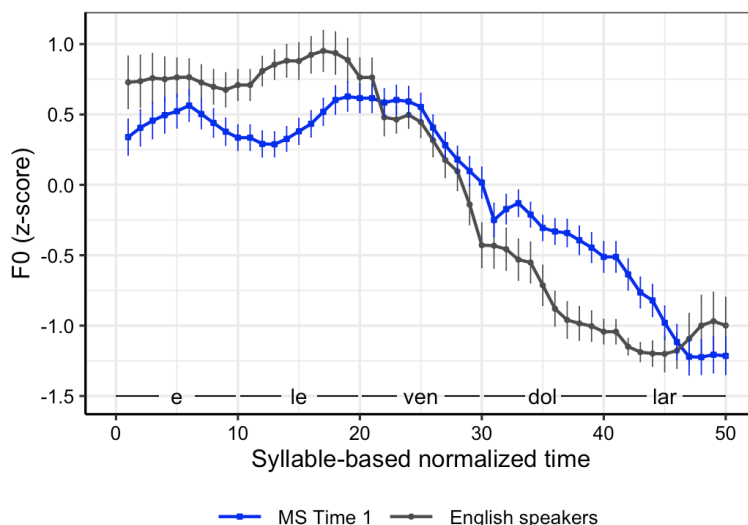


Figure 2. Pitch contours of *eleven dollars*: MS Time 1 and ES

Figure 3 shows the mean pitch contour of the same target phrase produced by MS after arrival (Time 1), at the end of the first semester (Time 2), and at the end of the fourth semester (Time 5). In order to test the F0 peak latency difference between the two groups, the latency of max F0 between 10 and 30 on the normalized time window, which corresponds to the syllables (*le*) and (*ven*), was extracted and analyzed using simple linear regression with Language Group as a predictor. Language Group was associated with a significant difference in F0 peak latency ($F(1, 73) = 8.88, p < .01$). The MS F0 peak showed an average delay of 5.17, which represents about one half syllable on the normalized time scale ($SE = 1.73$). As Figure 3 shows, this late alignment of the pitch drop (relative to the ES productions) is manifested at all time points. The F0 peak latency was analyzed within the MS group with Time as a predictor and Participant as random intercept using mixed-effects linear regression. Time did not affect the location of the pitch alignment. Thus, comparing the MS productions at three time points with the productions of the English control, we see a difference from the English native speakers in the location of the pitch peak which persists over the two-year period.

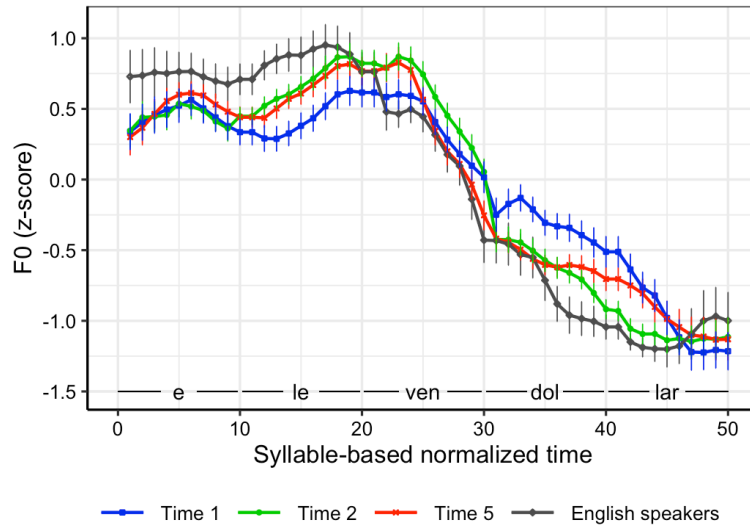


Figure 3. Pitch contours of *eleven dollars*

The ES and MS contours at Time 1 differed in a second respect: the magnitude of the pitch drop from the stressed syllable of the focused word (*le*) to the stressed syllable of the post-focus word (*dol*). While the MS contours show a relatively smaller pitch drop than the ES contour at all time points, the MS drop from *le* to *dol* was closer to that of the native speaker pattern at later time points (Time 2 and 5) than at Time 1. The difference in mean F0 from *le* to *dol* was computed and analyzed in a simple linear regression model with Language as a predictor. The pitch drop was significantly smaller in the MS pitch contour at Time 1 than in the ES pitch contour ($F(1, 73) = 8.87, p < .01$). A mixed-effects linear regression model with Time as a predictor and Participant as random intercept was fit with reference level of Time 1 for MS. Mean F0 difference in F0 z-score from the syllable (*le*) to the syllable (*dol*) increased by 0.56 at Time 2 ($SE = .16, p < .001$) and by 0.45 at Time 5 ($SE = .15, p < .01$) compared to Time 1. Thus there was a change over time in the magnitude of the post-peak drop, but not in the timing of the peak.

ES and MS also differed in their use of intensity. Figure 4 shows the maximum intensity of each syllable of *eleven dollars*. Both ES and MS had intensity peaks on the stressed syllable of the focused word (*le*), suggesting that the MS were indeed aware of the location of lexical stress. However, the intensity difference between the stressed syllable (*le*) and the post-stress syllable (*ven*) was larger for ES than for MS ($F(1, 13) = 14.14, p < .001$), and the MS showed no change in the intensity drop from *le* to *ven* over time.

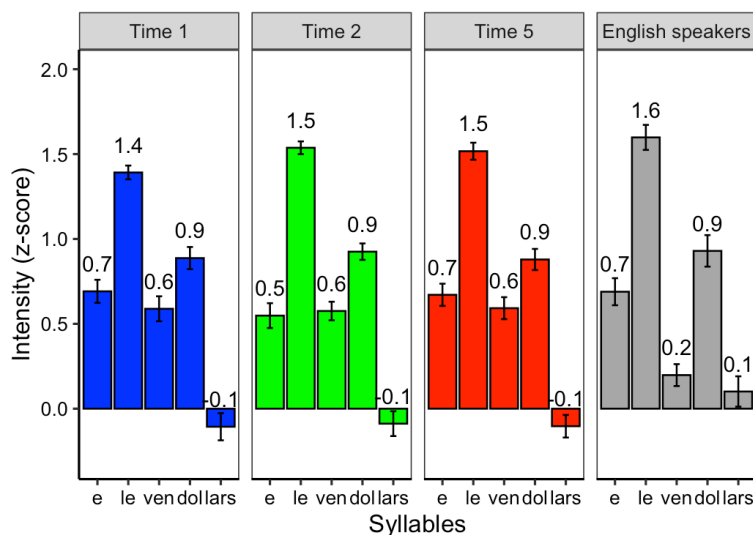


Figure 4. Syllable maximum intensity of *eleven dollars*

To summarize, we found that in the context of penultimate stress, the pitch peak came later in the focused word for MS than for ES, and the MS pattern showed little change. The MS also showed no change in their intensity patterns, which showed (like the ES productions) a clear intensity peak on the stressed syllable, but a smaller difference between the stressed and post-stressed syllables than that found for the ES. The MS did show some change over time in the magnitude of the pitch drop from the focused word (*eleven*) to the post-focus word (*dollars*), which made their later productions more similar to those of the ES.

4. Corrective focus with antepenultimate stress. To test the hypothesis that the Mandarin speakers failed to recognize the relationship between stress and pitch peaks, we analyzed productions containing the target phrase *ordinal number*, in which the focused word *ordinal* carries stress on the antepenultimate syllable (in contrast with penultimately stressed *eleven*). If the Mandarin speakers' failure to align the pitch peak with the stressed syllable was due to an inability to correctly perceive stress and/or tendency to produce the entire focused word with a raised pitch, we would expect them to exhibit a pitch peak toward the end of the focused word *ordinal*, as they did for *eleven*. If, on the other hand, the MS correctly perceived the position of stress, and if the location of the pitch peak was contingent on the location of the stressed syllable, they should exhibit an earlier pitch peak and pitch drop in *ordinal* compared to *eleven*, due to the difference in stress.

4.1. PARTICIPANTS, MATERIALS AND PROCEDURES. The target phrase *ordinal number* was elicited from a subset of the MS (21 speakers) who participated in the *eleven dollars* recording. These speakers, along with 29 ES (undergraduates at Stony Brook University), read the paragraph in Figure 5, followed by the dialogue in (4).

1. Some English learners may find it hard to understand the difference between cardinal and ordinal numbers.
2. Cardinal numbers are used for counting, such as the four in “four apples”.
3. When we talk about dates and the orders of things, we use ordinal numbers.
4. Ordinal numbers have a suffix, usually 'th', after the number.
5. An example of an ordinal number is the fourth in “the fourth of July”.




Figure 5. Reading passage for *ordinal number*

(4) Experimenter: *When we want to count something, should we use cardinal or ordinal numbers?*

Expected response: *We should use cardinal numbers.*

Experimenter: *“Fourth” is a cardinal number, right?*

Expected response: *No, “Fourth” is an **ORDINAL** number.*

The pitch and intensity of *ordinal number* was analyzed in the same way described in section 3.3. If the late pitch peak alignment in *eleven* was due to Mandarin speakers’ insensitivity to stress position, we would expect to find a similar pitch contour for both *eleven* and *ordinal*, with a peak near the right edge of each word.

4.2. RESULTS. For both ES and MS, the location of pitch and intensity peaks was clearly different for *ordinal* vs. *eleven*. The pitch contours of the phrase *ordinal number* produced by MS and ES are shown in Figure 6. For MS, the pitch peak was located near the boundary between the stressed syllable (*or*) and the following syllable (*di*) of the focused word. However, although the position of the Mandarin speakers’ pitch peak shifted according to the position of stress in the focused word, the pitch peak (computed as max F0 between 1 and 20 on the normalized time, corresponding to the first two syllables) in the Mandarin speakers’ productions at Time 1 was still later than the peak in the native speaker productions ($F(1, 48) = 13.62, p < .001$). This misalignment of the focus pitch peak persisted at all time points.

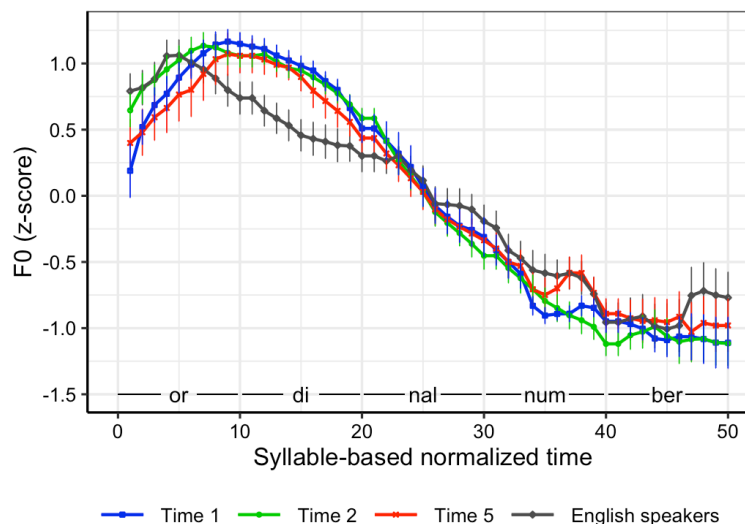


Figure 6. Pitch contours of *ordinal number*

For both groups, the location of intensity peaks also shifted as stress shifted. Figure 7 shows the maximum intensity of each syllable of *ordinal number*. Both ES and MS had intensity peaks on the stressed syllable of the focused word (*or*), which clearly suggests that the MS were aware of the position of lexical stress. This pattern did not change over time.

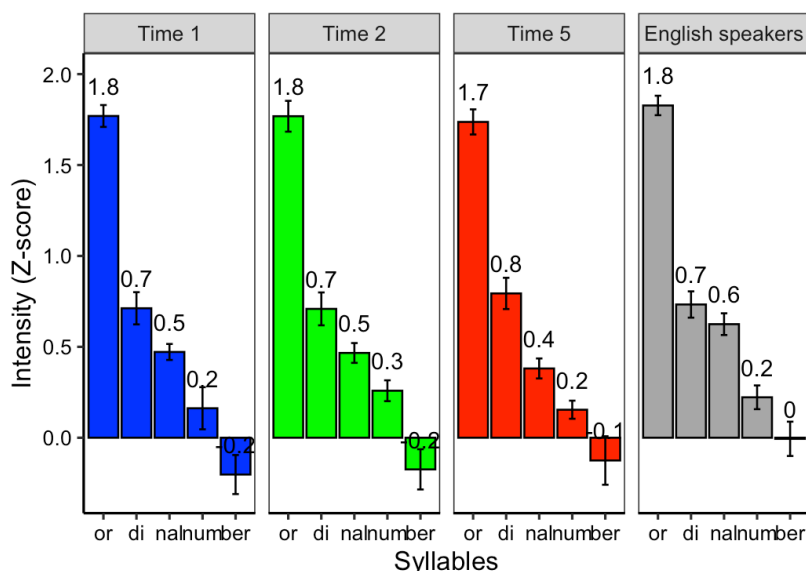


Figure 7. Syllable maximum intensity of *ordinal number*

5. Discussion & conclusion. Comparison of the location of the pitch peaks and drops in the Mandarin speakers' productions of words with penultimate vs. antepenultimate stress clearly indicates that the Mandarin speakers were sensitive to the position of the English lexical stress, and this is confirmed by the fact that the stressed syllables in each word type showed higher intensity than the non-stressed syllables. Furthermore, while the high pitch extended beyond the stressed syllable in both penultimate and antepenultimate stress words, it did not necessarily extend through two unstressed syllables (in *ordinal*); and it is not therefore the case that the Mandarin speakers assumed that the pitch accent aligns with the right edge of the word rather than with the stressed syllable, or that the entire focused word must be produced with raised pitch. We can therefore reject the hypothesis that the delay in the Mandarin speakers' pitch peaks (as compared with the productions of the English speakers) was a function of a lack of sensitivity to stress. Where the Mandarin speakers deviated from native speakers was not in a failure to connect the pitch peak with stress but rather in the timing of that connection: the Mandarin peak came after the stressed syllable, for both penultimate and antepenultimate stress positions. This non-native alignment of the pitch peak did not change over the course of two years, despite the fact that the participants' scores on the Versant test of English proficiency did show an upward trend during that period: Time 1 average 53.4 (range: 41-68) to Time 2 average 55.7 (range: 43-68), and to Time 5 average 58.2 (range: 42-73).

One aspect of the Mandarin productions which did change over time was the magnitude of the pitch drop from the stressed syllable of the focused word to the stressed syllable of the unfocused word, which became closer to that of the English speakers. Thus, while the timing of the pitch peak continued to show a delay compared to the corresponding peak in the English speakers' productions, the overall shape of the Mandarin speakers' rise-fall contour moved closer to that of native speakers over time.

A possibility we have not yet considered is that the Mandarin speakers actually use some different pitch accent than English speakers for corrective focus. However, Mandarin speakers' failure to master native-like pitch accent alignment is not limited to corrective focus; Lu and Kim (2016) found a similar phenomenon in L1-Mandarin productions of English list intonation. Nor is this misalignment limited to Mandarin speakers. Differences between native and non-native alignment of pitch accents has been noted among L2 speakers in a variety of languages: in L2 English by L1 Spanish and L1 Japanese speakers (Graham & Post 2018); in L2 Greek by L1 Dutch speakers (Mennen 2004); and in L2 Dutch by L1 Northern Chinese speakers (He et al. 2011). Given the widespread nature of pitch accent misalignment in a range of L1-L2 combinations, it seems unlikely that the misalignment is (at least solely) an effect of transfer from the native language. We conclude, instead, that the persistent failure to correctly align the focus pitch accent with the stressed syllable reflects a failure to fully master the complex orchestration and synchronization of the prosodic and segmental structure of a new language.

References

- Anderson-Hsieh, Janet & Kenneth Koehler. 1988. The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning* 38(4), 561–613. <https://doi.org/10.1111/j.1467-1770.1988.tb00167.x>.
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Boersma, Paul & David Weenink. 2019. *Praat: Doing phonetics by computer* [Computer program] (Version 6.1.07). <http://www.fon.hum.uva.nl/praat/>.
- Chen, Yiya. 2010. Post-focus compression – now you see it, now you don't. *Journal of Phonetics* 38. 517–525. <https://doi.org/10.1016/j.wocn.2010.06.004>.
- Chen, Ying. 2015. Post-focus compression in English by Mandarin learners. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Chen, Yang, Michael P. Robb, Harvey R. Gilbert & Jay W. Lerman. 2001. A study of sentence stress production in Mandarin speakers of American English. *Journal of the Acoustical Society of America* 109(4). 1681–1690. <https://doi.org/10.1121/1.1356023>.
- Chen, Ying, Yi Xu, & Susan Guion-Anderson. 2014. Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica* 71(4). 249–270. <https://doi.org/10.1159/000371891>.
- Graham, Calbert & Brechtje Post. 2018. Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics* 66. 1–14. <https://doi.org/10.1016/j.wocn.2017.08.002>.
- Hahn, Laura D. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38(2). 201–223. <https://doi.org/10.2307/3588378>.
- He, Xuliang, Judith Hanssen, Vincent J. van Heuven & Carlos Gussenhoven. 2011. Phonetic implementation must be learnt: Native versus Chinese realization of focus accent in Dutch. *Proceedings of the 18th International Congress of Phonetic Sciences*. 843–846.
- Kao, Sophia, Jiwon Hwang, Hyunah Baek, Chikako Takahashi & Ellen Broselow. 2016. International teaching assistants' production of English focus marking. *Proceedings of Meetings on Acoustics* 26(1). 060006. <https://doi.org/10.1121/2.0000356>.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune Haubo. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of statistical software* 82(13). 1–26. <https://doi.org/10.18637/jss.v082.i13>.

- Lin, Candise Y., Min Wang, William J. Idsardi & Yi Xu. 2014. Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition* 17. 316–346.
- Lu, Yu-an & Mi-ran Kim. 2016. Prosody transfer in second language acquisition: Tonal alignment in the production of English pitch accent by Mandarin native speakers. *Tsing Hua Journal of Chinese Studies, New Series* 46(4), 785–816.
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32, 543–563. <https://doi.org/10.1016/j.wocn.2004.02.002>.
- Munro, Murray J. & Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45. 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>.
- Ou, Shu-Chen. 2016. Perception of English lexical stress with a marked pitch accent by native speakers of Mandarin. *Taiwan Journal of Linguistics* 14(2) 1–31.
- Ouyang, Iris C. & Elsi Kaiser. 2015. Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience* 30(1-2). 57–72. <https://doi.org/10.1080/01690965.2013.805795>.
- Pearson Education. 2019. Versant English Test description and validation report. Retrieved from www.pearson.com/english/versant/tests/english-speaking-test.html.
- Peperkamp, Sharon & Emmanuel Dupoux. 2002. A typological study of stress “deafness.” In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology* 7. 203–240. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110197105.1.203>.
- Pierrehumbert, Janet & Julia B. Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*. 271–311. Cambridge, MA: MIT Press. <https://doi.org/10.1515/9783110197105.1.203>.
- R Core Team. 2019. R: A language and development for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>.
- Sereno, Joan, Lynne Lammers & Allard Jongman. 2016. The relative contribution of segments and intonation to the perception of foreign-accented speech. *Applied Psycholinguistics* 37(2). 303–322. <https://doi.org/10.1017/s0142716414000575>.
- Wang, Ting, Jun Liu, Yong-hun Lee & Yong-cheol Lee. 2020. The interaction between tone and prosodic focus in Mandarin Chinese. *Language and Linguistics* 21(2). 331–350.
- Wu, Wing L. & Lisa Chung. 2011. Post-focus compression in English-Cantonese bilingual speakers. *Proceedings of the 17th International Congress of Phonetic Sciences* 148–151.
- Xu, Yi & Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33. 159–197. <https://doi.org/10.1016/j.wocn.2004.11.001>.
- Xu, Yi. 1999. Effect of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27. 55–105. <https://doi.org/10.1006/jpho.1999.0086>.
- Xu, Yi. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. *Proceedings of the 17th International Congress of Phonetic Sciences*. 152–155.
- Xu, Yi. 2013. ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP)*. 7–10. Retrieved from http://www.homepages.ucl.ac.uk/~uclyyix/yispapers/Xu_TRASP2013.pdf.

- Zahorian, Stephen A. & Hongbing Hu. 2008. Spectral/temporal method for robust fundamental frequency tracking. *Journal of the Acoustical Society of America* 123(6). 4559–4571. <https://doi.org/10.1121/1.2916590>.
- Zhang, Yanhong, Shawn L. Nissen & Alexander L. Francis. 2008. Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America* 123(6). 4498–4513.