

## Native and non-native speaker processing and production of contrastive focus prosody

Chikako Takahashi, Sophia Kao, Hyunah Baek, Alex HL Yeung, Jiwon Hwang, & Ellen Broselow\*

**Abstract.** Several studies have found that the presence of L+H\* accent on a contrastive adjective assists native-speaking listeners in narrowing the referent of the noun following the adjective (e.g., Ito & Speer 2008, Weber et al. 2006). Our study addresses two questions: whether non-native speakers use prosodic cues in processing, as previous studies have shown for native speakers, and whether there is a relationship between the use of prosodic cues in processing and in production. Twenty-one Mandarin speakers living in the US and twenty-one native English speakers participated in two tasks investigating their processing and production of prosodic cues to contrastive focus. In the processing task, participants responded to the same recorded instruction containing an accented adjective in different contexts, in which the adjective was either contrastive (and therefore appropriately accented) or was repeated and followed by a contrasting noun, making focus accent on the adjective inappropriate. In the production task, participants guided an experimenter to place colored objects on a whiteboard, with some contexts designed to elicit contrastive focus. Overall results indicate that the Mandarin speakers made use of prosodic cues in both processing and production, although their focus prosody production differed from that of native speakers in several respects. Comparison of the results in the two experiments did not find strong correlations between processing and production. These results suggest that there is considerable heterogeneity even among native speakers in the use of prosodic cues in processing and production, and even those who do not use prosodic cues in processing may use them in production.

**Keywords.** focus prosody; English; Mandarin; contrastive focus; second language prosody

1. **Introduction.** Many languages, including English, use prosodic cues of pitch, intensity, and duration to focus specific information (e.g., Ito & Speer 2008). In studies examining processing in English and in German, the felicitous use of prosodic cues to contrastive focus was found to facilitate faster recognition and better retention (e.g., Ito & Speer 2008, Sedivy et al. 1995, Weber et al. 2006). For example, in Ito and Speer's (2008) eye-tracking study investigating the responses of native speakers of English to tree-decorating commands such as *First, hang the blue ball. Now, hang the green ball*, felicitous use of pitch accent L+H\* on contrasting adjectives led to an increase in early fixations to the target object, indicating that participants were able to utilize the prosodic information in conducting a visual search.

---

1\* Many thanks to Kevin Henderson and Sharon Benedett for stimulus recording. We also thank Marie Huffman and the LSA Annual Meeting audience for valuable comments and suggestions. This material is based upon work supported by NSF under Grant # IBSS-1519908. Chikako Takahashi ([chikako.takahashi@stonybrook.edu](mailto:chikako.takahashi@stonybrook.edu)), Sophia Kao ([sophia.kao@stonybrook.edu](mailto:sophia.kao@stonybrook.edu)), Hyunah Baek ([hyunah.baek@stonybrook.edu](mailto:hyunah.baek@stonybrook.edu)), Alex HL Yeung ([alex.yeung@stonybrook.edu](mailto:alex.yeung@stonybrook.edu)), Jiwon Hwang ([jiwon.hwang@stonybrook.edu](mailto:jiwon.hwang@stonybrook.edu)), & Ellen Broselow ([ellen.broselow@stonybrook.edu](mailto:ellen.broselow@stonybrook.edu)): Stony Brook University.

Because languages differ in the means used to signal focus (see Büring 2009 for a typology of focus realization), use of prosodic cues is frequently problematic for second language (L2) learners of languages such as English, for whom non-nativelike use of prosody has been linked to decreased intelligibility of L2 speech (e.g., Anderson-Hsieh & Koehler 1988, Hahn 2004, Munro & Derwing 1995, Sereno et al. 2016). While earlier work has revealed that non-native speakers' use of prosody in production differs from that of native speakers in factors such as the realization of pitch peak, intensity, and duration (e.g., Chen et al. 2013, Graham & Post 2017, Kao et al. 2016), there is a paucity of studies that examine non-native speakers' use of prosodic cues in processing as well as studies investigating the processing-production link in native and non-native speakers. Our study examines whether or not Mandarin L1 speakers use contrastive focus prosody in processing English as native English speakers do. We further investigate the relationship between the use of prosodic cues in production and processing in both language groups.

We will first briefly summarize the characteristics of English and Mandarin focus prosody, and review previous studies conducted on Mandarin L2 learners' use of English focus prosody, which show several differences between native and non-native speakers' realization of English focus. Section 3 presents the results of two experiments, one examining native English and Mandarin speakers' processing of contrastive focus prosody, and one in which participants of both groups produced sentences with focus prosody, along with a comparison of the results of the production and the processing experiments. The general discussion and conclusion follow in Section 4.

**2. Mandarin and English Focus Prosody.** Any English constituent may be focused to indicate contrastive information. Focus is generally exhibited by in-focus expansion, characterized by expanded F0 and intensity as well as longer duration of the stressed syllable in the focused word, and by post-focus compression, which manifests as a substantial decline in F0 and intensity. This decline generally begins in the stressed syllable of the focused word, with sustained compression lasting through the end of the utterance.

At least some dialects of Mandarin have been reported to signal focus prosody similarly to English, by means of in-focus expansion and post-focus compression (Jin 1996, Xu 1999, Xu et al. 2012). However, previous studies have reported differences between the production of focus prosody by native English and L2 English Mandarin speakers (Chen et al. 2014, Chen 2015, Kao et al. 2016, Wu & Chung 2011). For example, in a study of Mandarin and English speakers' production of corrective focus in English by Kao et al. (2016), both English speakers and Mandarin speakers exhibited in-focus expansion and post-focus compression, although the Mandarin speakers differed from the English speakers in the alignment of their F0 peak and the magnitude of their intensity drop, showing a later F0 peak than that of native English speakers and a smaller intensity drop between the stressed syllable of the focused word and the following syllable. Naturalness ratings by English native speakers confirmed a relationship between the perceived naturalness of Mandarin speaker's production and their misalignment of the F0 peak as well as their use of intensity cues.

While non-nativelike use of prosodic cues in producing focus prosody has been reported in many studies, the relationship between non-native production and processing of prosodic cues is not clear. In order to determine if Mandarin speakers use English contrastive focus prosody in processing, and to investigate the link between the processing and production, we conducted two experiments involving the processing and production of contrastive focus prosody.

3. **Experiments.** Twenty-one Mandarin speakers with English as a second language (MS; 11 males, 10 females) participated in the study as an experimental group. All had been attending a graduate program at Stony Brook University for 6-18 months at the time of their participation, and none had lived in an English-speaking country before beginning their graduate study. Twenty-one native speakers of English (ES; 7 males, 14 females), all undergraduate students at Stony Brook University, also participated in the same tasks as a control group.

The participants first completed the production task, followed by the processing task. Both together took approximately one hour. In the production task, the participants directed the experimenter to place colored objects on a whiteboard, with contexts designed to elicit contrastive focus. In the processing task, participants responded to auditory instructions spoken with felicitous or infelicitous prosody. The production task was administered first in order to avoid the auditory stimuli heard during the processing task influencing the participants' performance in the production task. This paper, however, will introduce the processing task followed by the production task below, as our research questions begin with a comparison of processing by native and non-native speakers. Finally, the participants took a language background survey. Those in the experimental group completed a separate online survey outside the session, and those in the control group filled in a survey sheet before they left the session.

3.1.PROCESSING. The processing experiment was designed to investigate the use of focus prosody in processing English by the two groups of speakers. Given the previous studies of English speakers' use of focus prosody in processing (Ito & Speer 2008, Sedivy et al. 1999), we expected that felicitous contrastive prosody on the adjective would reduce the response time (RT) in the English speaker group. The goal of the experiment was to determine whether such an effect was also present in the Mandarin speaker group.

3.1.1. METHODS. The participants took part in an item identification task in which each participant was seated in front of a laptop computer. In each trial, the participant was instructed to click on an item on the screen. All items were described with an adjective denoting colors or patterns (e.g., *ivory*, *purple*, *flowered*, *dotted*) and a noun (e.g., *mittens*, *necklace*, *sweater*). All adjectives and nouns were disyllabic trochees. The participants were first shown a screen with three items and heard the first instruction, e.g., *Click on the purple mittens*.

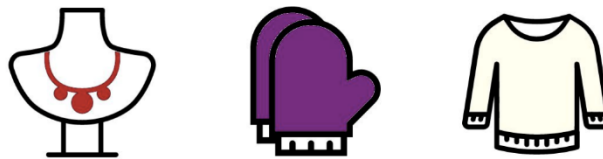


Figure 1. Screen for First Instruction (*Click on the purple mittens.*)

After pressing the correct key, participants were shown a new screen with the previous item in the center and four new items on the corners, and then heard the second instruction, e.g., *Now click on the scarlet mittens*.

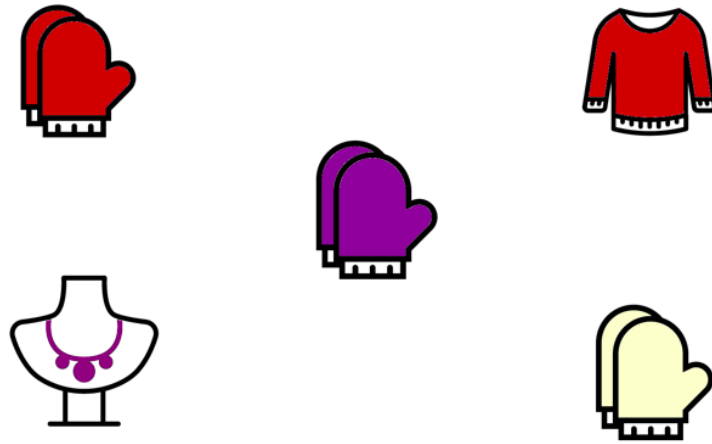


Figure 2. Screen for Second Instruction (*Now click on the scarlet mittens.*)

The participants were instructed to always use the index finger of their right hand to press the corresponding key. When the participant clicked on an item, the trial was completed and a new trial began. The experiment began with a practice block in the presence of an experimenter, who answered any questions. After the practice trials, the experimenter left the room.

In each trial, the first instruction provided the context for the second instruction. The context determined whether the prosodic focus (L+H\*) on the adjective was felicitous or infelicitous, as illustrated in table 1. This design allowed us to compare response times to the same stimulus (with prosodic focus on the adjective) in two different contexts. The stimuli were recorded by a male native speakers of English. The L+H\* on the focused adjective was confirmed through acoustic analysis of pitch and intensity drop from the adjective to the following noun. The entire experiment was divided into four sets. Each set included 24 randomized stimuli and was presented in PsychoPy (Peirce 2007).

Conditions	Instruction 1	Instruction 2	Contrast
<b>Felicitous prosody</b>	<i>Click on the purple mittens.</i>	<i>Now click on the SCARLET mittens</i>	<i>purple – scarlet</i>
<b>Infelicitous prosody</b>	<i>Click on the scarlet necklace.</i>	<i>Now click on the SCARLET mittens.</i>	<i>necklace – mittens</i>

Table 1: Felicitous and Infelicitous Conditions

Response time (RT) was measured from the offset of the noun in the second instruction to the time of the key press. The collected RTs were log-transformed for statistical analysis. One English speaker and one Mandarin speaker were excluded as outliers, with Tukey’s method using interquartile range.

3.1.2. RESULTS AND DISCUSSION. As shown in figure 3 below, ESs were in general faster than MSs in both conditions. In both language groups, mean RT for the felicitous condition was shorter than for the infelicitous condition. We analyzed effects of Condition on RT with linear mixed effects-models using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Development Core Team 2006). As fixed effects, we included Language group, Condition

and the interaction term. As random effects, we entered intercepts for Subject and Item and by-subject and by-item random slopes for the effect of Condition. We used likelihood ratio tests to compare this model to nested models with each term removed. Language group significantly influenced RT ( $\chi^2(1) = 24.31, p < .001$ ) with ESs faster than MSs ( $\beta = -46.29, SE = 8.32$ ). Condition also significantly affected RT ( $\chi^2(1) = 4.77, p = .029$ ), with shorter RT in felicitous condition than in infelicitous condition ( $\beta = -6.56, SE = 2.92$ ). However, the interaction did not reach significance ( $\chi^2(1) = 0.28, p = 0.59$ ), suggesting that the effect of Condition did not depend on the language group.

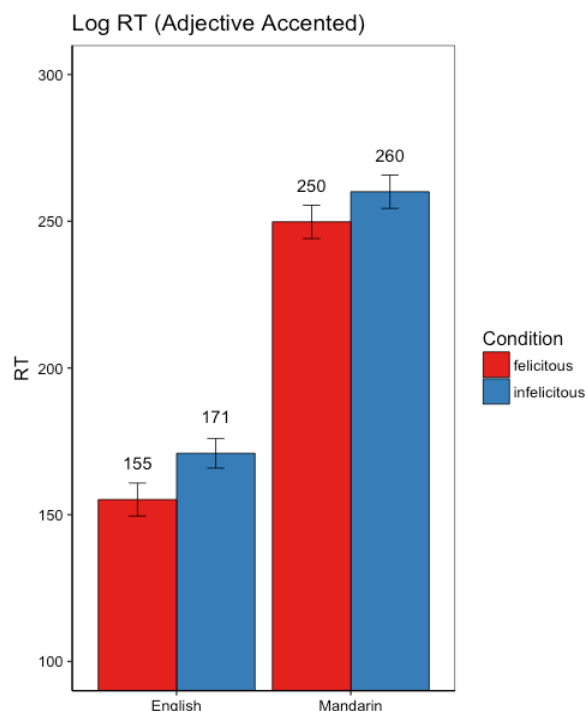


Figure 3. Average RTs for each condition in English and Mandarin group

Figure 4 below shows the individual RT variation in the two conditions. Each data point represents a mean RT score for each individual, with a line connecting the same participant's results in felicitous and infelicitous conditions. Despite the overall shorter RTs in the felicitous condition, as illustrated in figure 3, around a quarter of participants in each group failed to show the expected difference in RTs.

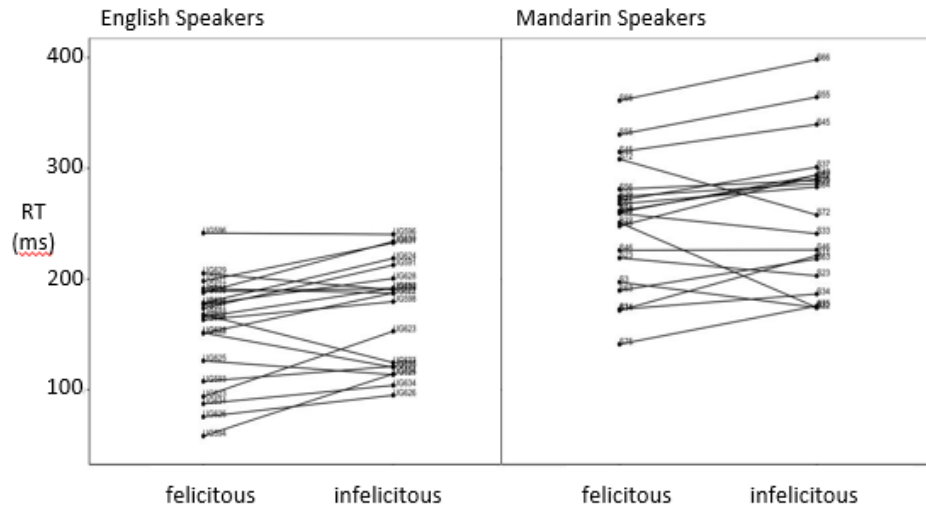


Figure 4. Individual Variation: each data point representing a mean RT score for each individual

Our results suggest that these Mandarin speakers were in general able to use focus prosody similarly to native speakers. That is, felicitous contrastive focus prosody facilitated information processing. However, the results also demonstrate a large degree of variability in the use of prosodic cues in processing, even among the native speakers of English. This variability could be related to individual differences in using focus prosody in that those who were not sensitive to focus prosody in processing might not exhibit canonical use of prosodic cues in production. In order to investigate the relationship between the processing and production of contrastive focus prosody, our production study examined the same participants' use of prosodic cues in a production experiment.

3.2.PRODUCTION. Based on previous studies reporting that non-native speakers of English differ in the production of prosodic cues (Kao et al. 2016 for Mandarin speakers producing sentences, Graham & Post 2018 for Mandarin, Spanish, and Japanese speakers producing words), we predict differences in use of prosodic cues, such as alignment of pitch peak and extent of intensity drop, between English and Mandarin speakers. Our goal is to investigate the relationship between individual participants' production data and their processing data, in order to determine whether use of prosodic cues in processing correlates with native-like production of focus prosody.

3.2.1. METHODS. Target phrases in the production experiment consisted of NPs containing one of three adjectives representing colors (*navy*, *orange*, and *yellow*) paired with one of three nouns referring to shapes (*arrow*, *diamond*, and *oval*). The targets contained only voiced sounds and both adjectives and nouns were disyllabic trochees. The combination of three adjectives and three nouns yielded nine target noun phrases, which were produced in a carrier sentence as shown below.

(1) Target NP elicited in the production task.

*Put the (adjective) (noun) over the (adjective) (noun), please.*

CONTEXT NP

TARGET NP

In the elicited sentences, the adjectives in the context NP and the target NP were contrasting while the nouns were repeated (e.g., *put the orange diamond over the navy diamond, please*), making contrastive focus on the adjective appropriate.

Each participant was seated in front of a computer screen across the table from the experimenter, who had a whiteboard and a pool of colored objects to place on the whiteboard. Participants saw slides showing the correct placement of colored objects, with each slide showing one new object placement. Participants were asked to give instructions to help the experimenter make their whiteboard match the screen they saw. Figure 5 below illustrates an example slide used in the production task, eliciting the instruction *Put the orange diamond over the navy diamond, please*. After each round of six steps, the participant checked whether the experimenter's white board matched their final slide.

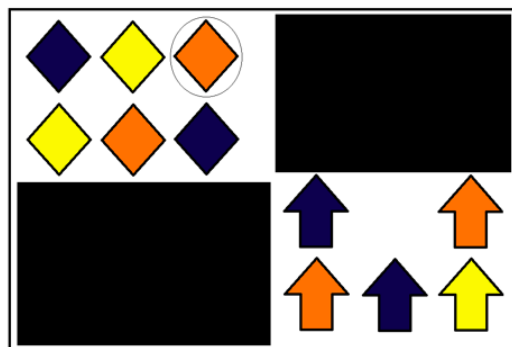


Figure 5. An example slide for the production task eliciting *Put the orange diamond over the navy diamond, please*.

The recording was conducted in a sound-treated room at Stony Brook University, using a Zoom H6 digital recorder. The production task took approximately 15 minutes. A total of 504 utterances (2 rounds  $\times$  6 instructions  $\times$  2 groups  $\times$  21 participants) was obtained.

Target NPs were hand-segmented into an adjective and a noun using Praat (Boersma & Weenink 2011). After segmentation, time-normalized pitch (semitone) and intensity (dB) contours were extracted using a Praat script (ProsodyPro, Xu 2013). Forty-four of the 252 phrases produced by the English speakers were produced with upward intonation, and these were excluded from analysis.

3.2.2. RESULTS. For each group of speakers, the average pitch contour was extracted by combining all elicitations of the target phrases. Figure 4 illustrates the average time-normalized pitch contours of the two groups, where the first two syllables are the adjectives bearing contrastive focus, and the last two syllables are the nouns. As indicated by the arrows, the pitch began to drop earlier for the ESs than for the MSs, although the MSs exhibited a greater range of pitch. This pattern was observed for both male and female participants, as shown in Figure 5, although there was little pitch movement in male ESs. To assess whether pitch peak was earlier for ESs, we fit a linear mixed-effects model on Time where the pitch peak occurred in the time-normalized pitch contour. The model included Language group as a fixed effect and Subject and Item as random intercepts. We used a likelihood ratio test to compare this model to an intercept-only model. Language group significantly affected peak latency ( $\chi^2(1) = 12.27$ ,  $p = 0.00045$ ) with an earlier pitch peak for ESs than MSs ( $\beta = 3.79$ ,  $SE = 1.01$ ).



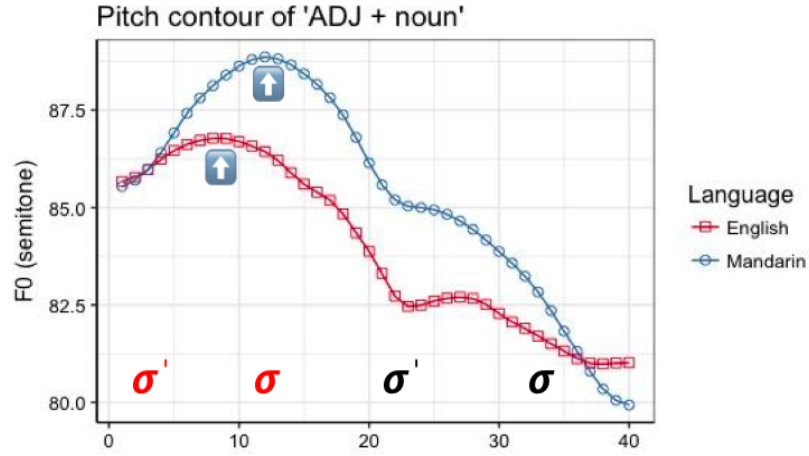


Figure 5. Time-normalized average pitch contours of ES and MS.

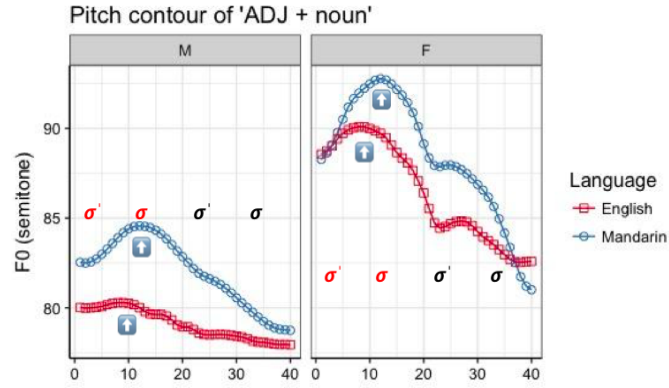


Figure 6. Time-normalized average pitch contours of ES and MS by gender.

Figure 6 shows time-normalized intensity contours for each group. As indicated by the arrows, the intensity drop from the stressed syllable of the focused word to the following syllable was larger for ESs than for MSs. We fit a linear mixed-effects model on the intensity difference from the first peak to the second peak with Language group as a fixed effect and with Subject and Item as random intercepts. We used a likelihood ratio test to compare the model to an intercept-only model. Language group affected intensity drop ( $\chi^2(1) = 8.48, p = 0.0036$ ) with ESs having a larger intensity drop than MSs ( $\beta = -1.39, SE = 0.45$ ). The same pattern was found in both male and female groups, as shown in Figure 7.



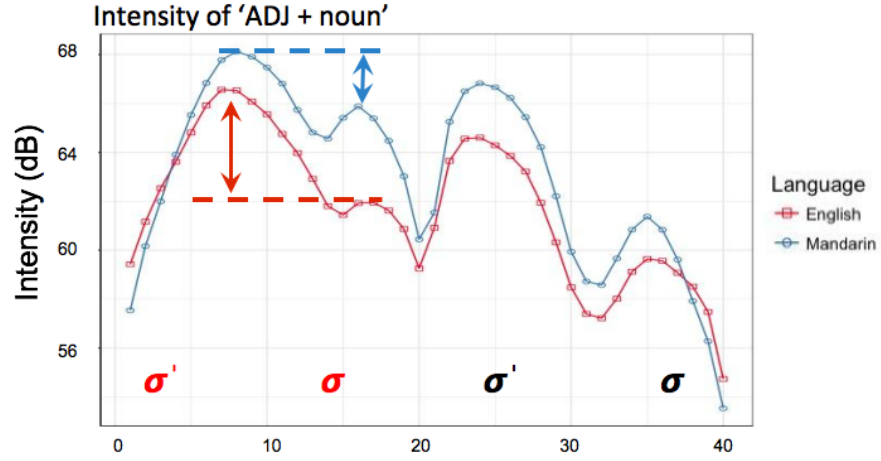


Figure 7. Time-normalized average intensity contours of ES and MS.

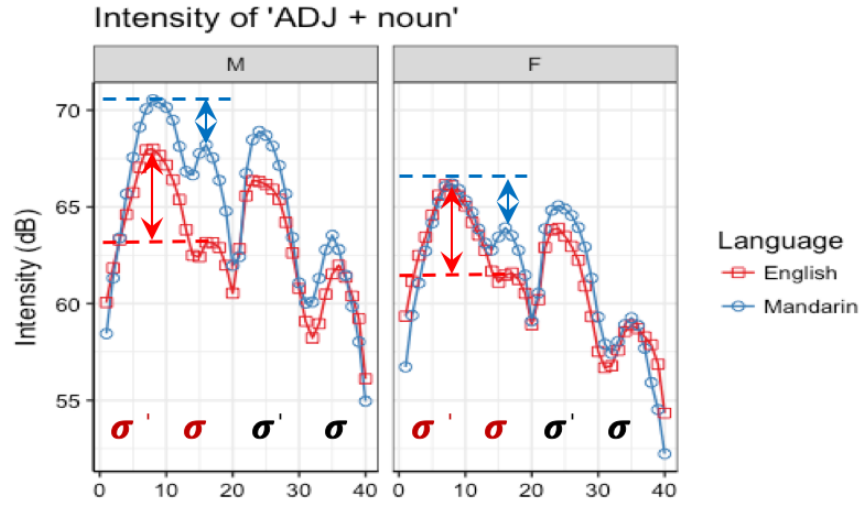


Figure 8. Time-normalized average intensity contours of ES and MS by gender.

Figure 9 shows the time-normalized alignment of pitch and intensity contours. For the ESs, the pitch peak and the intensity peak were aligned, while for the MSs the pitch peak came later than the intensity peak.

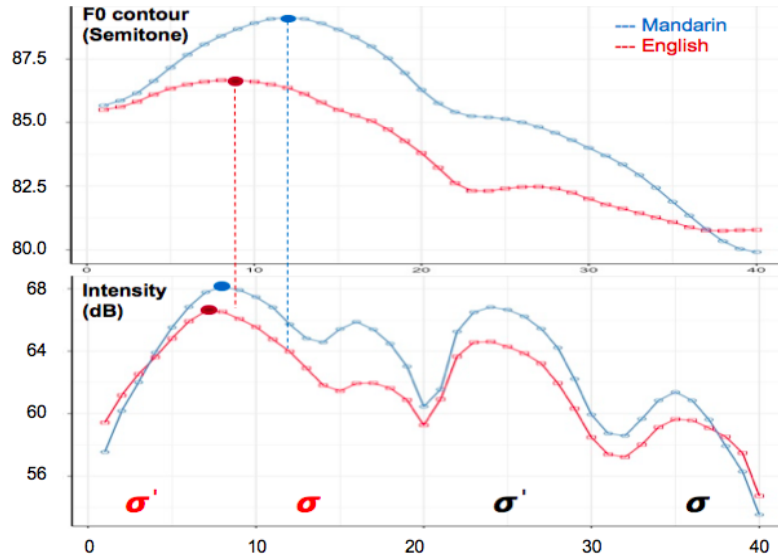


Figure 9. Time-normalized alignment of pitch and intensity contours.

Summarizing the results of the production task, the two groups differed in their realization of focus, with MSs tending to manifest a later pitch peak than ESs, with a smaller intensity drop than ESs from the stressed syllable of the focused word to the following syllable. The MS - ES differences in the production of contrastive focus mirror differences in their production of corrective focus (Kao et al. 2016).

3.3. COMPARISON OF PROCESSING AND PRODUCTION. One of the major aims of this study was to investigate the relationship between processing and production of contrastive focus prosody. In order to examine the correlation between each participant's production and processing task results, we first determined the response time score (RT score) for each participant, representing how quickly they responded in the felicitous condition compared to the infelicitous condition. The RT score was calculated by subtracting each participant's mean RT for the felicitous condition from their RT for the infelicitous condition. Therefore, a positive RT score indicated the expected use of prosodic cues, showing shorter RT for the felicitous than the infelicitous condition. We examined the relationship between RT score and the magnitude of the drop in F0 and intensity from the focused word to the following word. We also examined the correlation between the RT score and two additional factors: the participants' proficiency based on the Versant English test (Pearson Education 2010) and the perceived naturalness of the participant's production of contrastive focus as judged by an expert rater. However, no strong correlations between any of the measures were found. To investigate further, we visually inspected each participant's average F0 and intensity contour along with their RT score. While some Mandarin and English speaking participants who obtained a positive RT score also exhibited canonical use of prosodic cues in focus prosody production, some other 'good' prosodic cue users in processing exhibited monotonous prosody in production. Similar variability was found in participants who did not show much RT difference between felicitous and infelicitous prosody conditions in the processing task.

4. DISCUSSION AND CONCLUSION. The experiments conducted for the current study investigated non-native speakers' use of L2 English contrastive focus prosody in processing and its relationship with prosodic cues in English contrastive focus prosody production. In the

processing experiment, which investigated whether Mandarin speakers show effects of felicitous versus infelicitous focus prosody in their processing of English, the Mandarin speakers in general responded more quickly in the felicitous condition compared to the infelicitous one, much like English speakers did, although the effect was smaller in the Mandarin speaker group. This result indicates that at least this group of Mandarin speakers can use L2 English contrastive focus prosody in processing. This might be due to the fact that Mandarin focus prosody has a ‘shape’ similar to that of English, with in-focus expansion and post-focus compression, and that Mandarin speakers were using their knowledge of L1 focus prosody in processing. It is also possible that the results are specific to this group of Mandarin speakers, whose level of English proficiency was high enough to qualify them for graduate study in a US university. This would mean that these speakers have learned to use English focus prosody in processing.

The production task examined whether the same Mandarin speakers use focus prosody in their production of English contrastive focus. The results of the task indicated that the Mandarin speakers did in fact use focus prosody in their production, similarly to the English native speakers. However, their phonetic realization of focus cues differed from that of native speakers in both the alignment of the pitch peak and the extent of the intensity drop. This pattern mirrors the results of Kao et al. (2016), which showed a similar effect in Mandarin speakers’ production of corrective focus. The results from the two experiments suggest that non-native speakers can acquire L2 prosody to some extent and they can produce and process it in a similar way as native speakers do, although there may be some differences (e.g., in the phonetic realization of focus). These conclusions must be taken as tentative, however, since it is likely that the response time measure (unlike the eye tracking used in the earlier processing studies) is not sufficiently sensitive to tease apart the effects of segmental vs. prosodic information. We are currently designing an eye tracking study designed to isolate the effects of prosodic information.

A somewhat surprising finding of our study was the apparent disconnect between the production and processing of focus prosody in some speakers in both language groups. Some speakers did not exhibit the expected manifestation of focus prosody, but seemed to utilize the focus prosody in the processing task, while others showed the reverse pattern. This indicates at least some degree of independence between the use of prosody in production and processing, mirroring a similar lack of congruence between the production and perception of cues to L2 phoneme contrasts (e.g., Schertz et al. 2015).

## References

- Anderson-Hsieh, Janet & Kenneth Koehler. 1988. The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning* 38. 561-593.  
<https://doi.org/10.1111/j.1467-1770.1988.tb00167.x>.
- Bates, Douglass, 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. 2011. Praat: Doing phonetics by computer. *Ear and Hearing* 32(2). 266. <https://doi.org/10.1097/aud.0b013e31821473f7>.
- Büring, Daniel. 2009. Towards a typology of focus realization. *Information Structure*. 177-205.  
<https://doi.org/10.1093/acprof:oso/9780199570959.003.0008>.

- Chen, Yang. 2015. Post-focus compression in English by Mandarin learners. (ms.)
- Chen, Yang, Michael Robb, Harvey Gilbert, & Jay Lerman. 2001. A study of sentence stress production in Mandarin speakers of American English. *Journal of Acoustic 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>.
- Fowler, Carol A. 1995. Acoustic and kinematic correlates of contrastive stress accent in spoken English. In Fredericka Bell-Berti & Raphael J. Lawrence (eds.), *Producing speech: Contemporary issues: For Katherine Safford Harris*. 355-373. <https://doi.org/10.1515/9783110874020.145>.
- 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>.
- Ito, Kiwako & Shari R. Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58. 541-573. <https://doi.org/10.1016/j.jml.2007.06.013>.
- Jin, Shunde. 1996. *An acoustic study of sentence stress in Mandarin Chinese*. PhD Dissertation. Ohio State University. <https://doi.org/10.3724/sp.j.1016.2011.01638>.
- Kao, Sophia, Jiwon Hwang, Hyunah Baek, Chikako Takahashi, & Ellen Broselow. 2016. International teaching assistants' production of focus intonation. *Proceedings of Meetings on Acoustics* 26(1). 1-13. <https://doi.org/10.1121/2.0000356>.
- Lee, Yong-cheol, Bei Wang, Sisi Chen, Martine Adda-Decker, Angélique Amelot, Satoshi Nambu, & Mark Liberman. 2015. A crosslinguistic study of prosodic focus. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*. 4754-4758. <https://doi.org/10.1109/icassp.2015.7178873>.
- Munro, Murray J. & Tracy M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45. 73-97. <https://doi.org/10.1111/0023-8333.49.s1.8>.
- Pearson Education. 2010. *Versant English test description and validation manual*. Palo Alto, CA: Author. Retrieved from [www.versanttest.com/products/english.jsp](http://www.versanttest.com/products/english.jsp).
- Peirce, Jonathan W. 2007. PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods* 162(1-2). 8-13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>.
- Weber, Andrea, Bettina Brown, Matthew W. Crocker. 2006. Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech* 49(3). 367-392. <http://doi.org/10.1177/00238309060490030301>.
- Wu, Wing Li & Lisa Chung. (2011). Post-focus compression in English-Cantonese bilingual speakers. In *Proceedings of the 17th International Congress of Phonetic Sciences*. 148-151.
- Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27. 55-105. <http://doi.org/10.1006/jpho.1999.0086>.
- Xu, Yi. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong. 152-155.

- Xu, Yi. 2013. ProsodyPro - A Tool for Large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*. Aix-en-Provence, France. 7-10.
- Xu, Yi, Szu-wei Chen, & Bei Wang. 2012. *Prosodic focus with and without post-focus compression: A typological divide within the same language family?* (m.s.)  
<https://doi.org/10.1515/tlr-2012-0006>.