Pitch range, intensity, and vocal fry in non-native and native English focus intonation

Alex Hong-Lun Yeung, Hyunah Baek, Chikako Takahashi, Joseph Duncan, Sharon Benedette, Jiwon Hwang, and Ellen Broselow

Citation: Proc. Mtgs. Acoust. **36**, 060004 (2019); doi: 10.1121/2.0001047

View online: https://doi.org/10.1121/2.0001047

View Table of Contents: https://asa.scitation.org/toc/pma/36/1

Published by the Acoustical Society of America

ASA
POMA
Proceedings
of Meetings
on Acoustics
Turn Your ASA Presentations
and Posters into Published Papers!

Volume 36 http://acousticalsociety.org/

177th Meeting of the Acoustical Society of America

Louisville, Kentucky

13-17 May 2019

Speech Communication: Paper 3pSC6

Pitch range, intensity, and vocal fry in non-native and native English focus intonation

Alex Hong-Lun Yeung, Hyunah Baek, Chikako Takahashi, Joseph Duncan and Sharon Benedette
Department of Linguistics, Stony Brook University, Stony Brook, NY, 11790; alex.yeung@stonybrook.edu; hyunah.baek@stonybrook.edu; chikako.takahashi@stonybrook.edu; joseph.duncan@stonybrook.edu; sharon.benedett@stonybrook.edu

Jiwon Hwang
Department of Asian & Asian American Studies, Stony Brook University, Stony Brook, NY, 11794; jiwon.hwang@stonybrook.edu

Ellen Broselow
Department of Linguistics, Stony Brook University, Stony Brook, NY, 11790; ellen.broselow@stonybrook.edu

This study investigates the production of English contrastive focus by 26 L1 Mandarin speakers (MS) and 21 native English speakers (ES). The participants completed an interactive game in which they directed experimenters to decorate objects, producing sentences with contrasted adjective or noun (e.g., Andy wants an orange diamond on his towel and a NAVY diamond/orange OVAL on Mindy's towel). Our results show that while both groups were similar in their use of intensity, the MS used less lengthening of focused words than did ES. While both groups showed a pitch peak on focused adjectives, they differed in the noun-focus condition: while the MS exhibited what is normally considered canonical pitch prosody, with peaks on both focused adjectives and focused nouns, many of the ES' productions lacked the expected pitch peak on a focused noun. The ES' divergence from textbook focus prosody reflected their use of an innovative intonation pattern, especially prominent among young native English speakers (Wolk et al. 2012), which is characterized by pitch declination and an increase in vocal fry throughout the intonational phrase. We conclude that this intonation pattern restricted the ES's ability to use pitch rises to mark focus on phrase-final nouns.

Published by the Acoustical Society of America

1. INTRODUCTION

Since prosodic information is often crucial in communicative functions such as signaling new vs. given information, being able to understand and use prosody in a given language is important in successful communication. Studies of second language (L2) prosody have uncovered various ways in which non-native patterns differ from native speaker norms, such as the use of pitch, intensity, and duration cues and the magnitude and location of pitch/intensity rises and falls (e.g., Chen et al. 2001, Chen et al. 2015, Kao et al. 2016). The goal of the present study is to compare the production of English contrastive focus by native speakers of English, an intonation language, and Mandarin, a language in which the primary function of pitch is to signal lexical contrasts.

Contrastive focus in English is most commonly realized as association of a pitch peak (generally analyzed as a L+H* pitch accent, Pierrehumbert & Hirschberg 1990), with the stressed syllable of the focused element, which is also realized with increased intensity and duration. In Mandarin, prosodic realization of contrast is constrained by the necessity to preserve lexical tones, and therefore the use of acoustic cues varies depending on the tonal context (e.g., Xu 1999, Chen 2010, Ouyang & Kaiser 2015, Wang et al. to appear). Since English speakers presumably have greater freedom than Mandarin speakers to use pitch in signaling information structure, we expected that speakers whose first language (L1) is Mandarin might show less reliance on pitch than native speakers of English in producing contrastive focus in English. However, using an interactive task used to elicit sentences involving contrastive focus from native English speakers and from Mandarin speakers living in the US, we found that in some contexts, the Mandarin speakers actually adhered more closely to canonical English focus pitch patterns than did the native speakers. We argue that the English speakers' deviations from textbook focus prosody reflect an innovative but increasingly common sentence intonation associated with marked declination in pitch, often accompanied by vocal fry, toward the ends of phrases (Wolk et al. 2012, Abdelli-Beruh 2014), and that this intonation melody is antithetical to realization of a pitch peak on phrase-final focused elements.

2. METHODS

A. Materials and Procedures

Participants completed an interactive task in which they guided the experimenter to decorate a whiteboard based on pictures on a computer screen that was visible only to the participant. The participant could not see the experimenter's whiteboard until the task had been completed. To set the context for the task, participants first read the following paragraph aloud to the experimenter:

> *Andy and Mindy are twins. They are getting ready to go off to college. Their mom bought a lot of new things for them to take: towels, blankets, mittens, sweaters, and a dorm refrigerator. Everything their mom bought is very plain, so they want to decorate their new things. First, Andy wants to decorate both blankets. Pick up the whiteboards with pictures of blankets. Now you are going to help Andy decorate the blankets with items from the basket. I will tell you which items to use. You can't look at my screen, and I can't look at your whiteboards, but we can ask questions. At the end, if you have followed instructions correctly, we will both get candy. Are you ready?*

Participants then gave instructions based on the slides on their screens. The first slide introduced a set of three objects for the experimenter to place on the table (e.g., *Pick up a yellow diamond, an orange diamond, and a navy diamond from the basket*, Fig. 1). The following slide then demonstrated placement of two of these three objects on the whiteboard, eliciting an instruction containing two adjective-noun combinations. Participants were asked to follow a specific format in their instructions, illustrated by *Andy wants the **yellow diamond** on his towel and the **orange diamond** on Mindy's towel*. In this example, the target phrase *orange diamond* was expected to be realized with contrastive focus on the adjective *orange*, which contrasts with the color of the previously mentioned object (the yellow diamond), as well as with the other remaining object on the table (the navy diamond). Each adjective and each noun in the target phrase was a bisyllabic trochee containing only voiced segments (adjectives: *orange/navy/yellow*; nouns: *arrow/diamond/oval*).

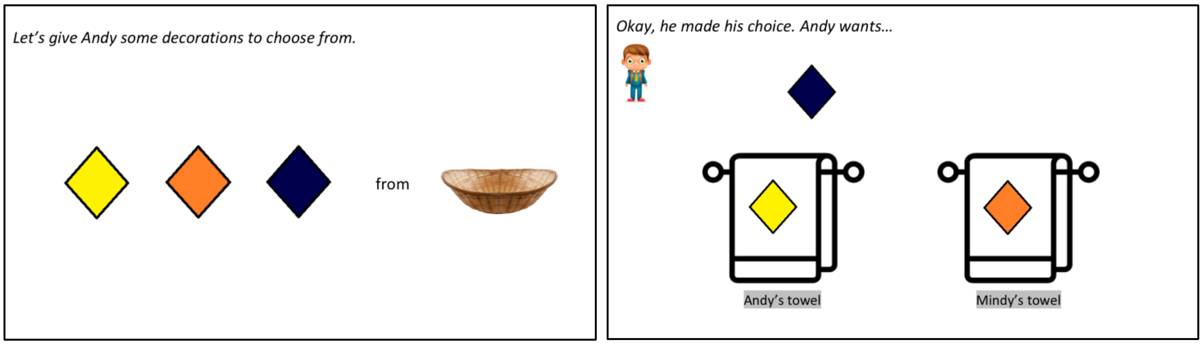
Let's give Andy some decorations to choose from.
from
Okay, he made his choice. Andy wants...
Andy's towel
Mindy's towel

Figure 1. Example slides for production task.

Each participant produced 33 target phrases: 12 in the ADJ-focus condition, in which the adjective was contrastive (e.g., *Andy wants the yellow diamond on his towel and the orange diamond on Mindy's towel*); another 12 in the NOUN-focus condition, in which the noun was contrastive (e.g., *Andy wants the yellow diamond on his towel and the yellow arrow on Mindy's towel*); and the remaining 9 in the neutral condition, in which neither the adjective nor noun was contrastive (e.g., *Please show me the yellow arrow*). Sessions took place in the Phonetics Lab at Stony Brook University, and participants' productions were recorded throughout the task using a Zoom H6 digital recorder and a SM10A-CN dynamic headworn microphone.

B. Participants

Two groups of speakers participated in the experiment: 21 native speakers of English (ES; 15 females, 6 males; mean age = 19.9 yrs) and 26 native speakers of Mandarin (MS; 12 females, 14 males; mean age = 25.5 yrs). All participants completed a short questionnaire about their language background. ES participants were undergraduate students at Stony Brook University who reported English as their native language. MS participants were graduate students at the same university who reported Mandarin as their first language. At the time of their participation, the MS had been living in the US for 19.7 months on average (range: 10-34 months). All MS participants took the Versant English Speaking Test (Pearson Education 2010); their average score was 58.7 out of 80 (range: 42-72).

C. Acoustic Measurements

The experiment yielded recordings of 1551 target phrases (47 participants * 33 phrases). Forty-two phrases were excluded due to production errors. The remaining 1509 phrases were hand-segmented in Praat (Boersma & Weenink 2019), as shown in Fig. 2. After the segmentation, three acoustic measurements associated with focus realization were obtained: F0, intensity, and duration. F0 contours were extracted using pYAAPT (available at http://bjbschmitt.github.io/AMFMdecompy/pYAAPT.html.), a ported version for Python from YAAPT (Yet Another Algorithm for Pitch Tracking) which was originally written as Matlab functions (Zahorian & Hu 2008). The minimal and maximal F0 boundaries were estimated automatically and visually for each speaker in order to optimize the individual parameter. Computed F0 contours were then time-normalized. Time-normalized intensity contours were extracted and the duration of each syllable was measured using ProsodyPro, a Praat script for prosodic analysis (Xu 2013). The domain of time-normalization was a syllable divided into 10 points. Thus, point 1 in each syllable corresponds to 1/10 of the syllable. Trials in the neutral condition were excluded from further analyses. Raw values of F0, intensity and duration were converted to z-scores within speaker for statistical analyses and data visualization.

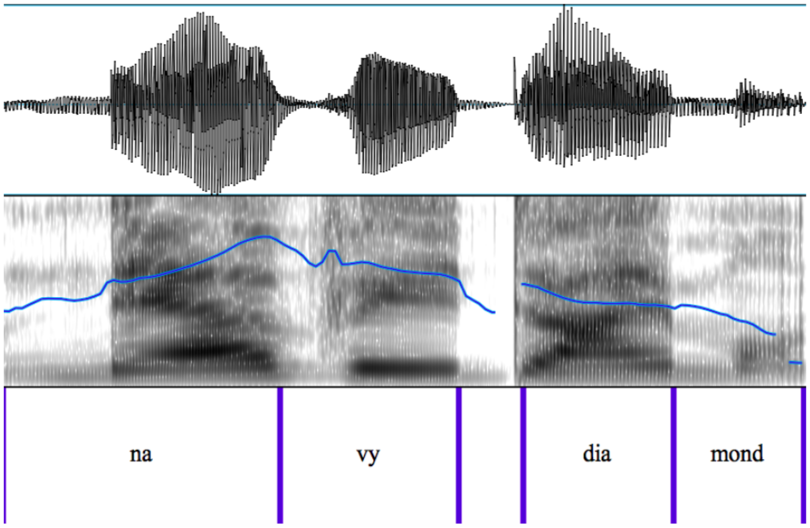

Figure 2. Example of a segmented target phrase.

3. RESULTS

A. PITCH

We first compared the average F0 contours of the two groups for the target phrases. One difference between the two groups appeared in the overall F0 range, which was wider for the Mandarin speakers than for the English speakers, with steeper rises and falls for MS productions. This difference is expected given the Mandarin L1 pattern of F0 range expansion associated with focus (Xu 1999, Chen & Braun 2006, Chen & Gussenhoven 2008, Ouyang & Kaiser 2015).

A second difference between the groups, involving the relationship between F0 peak and location of focus, was more surprising. In general, the Mandarin speakers exhibited the expected pitch peak on the focused word in both ADJ-focus and NOUN-focus conditions, consistent with descriptions of English focus as involving association of L+H* pitch accent with the focused element. In contrast, English speakers showed the expected pitch peak on a focused adjective, but not on a focused noun (Fig. 3 Left). Fig. 3 Right compares the magnitude and direction of change in F0 given the presence/absence of contrastive focus, computed by subtracting the maximum F0 value of the noun from the adjective maximum F0 in each phrase. Positive values in these variables indicate higher F0 in the adjective than in the noun, as expected for ADJ-focus, while negative values indicate higher F0 in the noun, expected for NOUN-focus conditions. Both groups showed the expected positive values for ADJ-focus, but only the Mandarin speakers had the expected negative value for the NOUN-focus condition – the English speakers actually had higher pitch on the pre-focus adjective than on the focused noun.

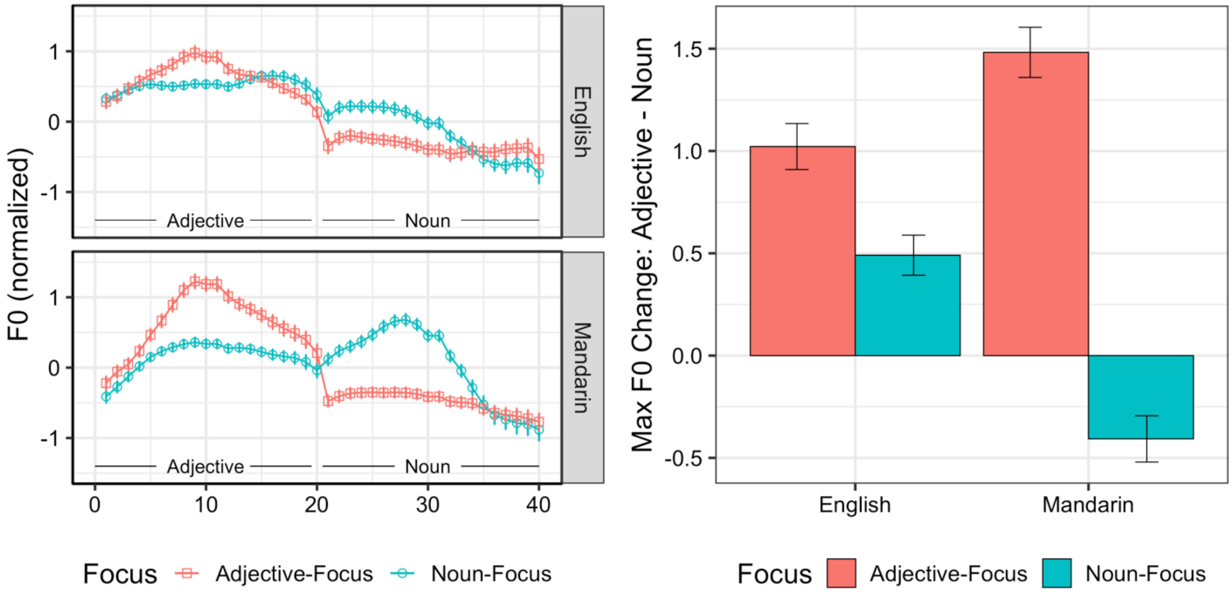

Figure 3. Average normalized F0 contour on normalized time (left) and average max F0 change from adjective to noun (right). Error bars represent 95% confidence intervals. y-axis indicate normalized F0 in z-score.

The unexpected failure of the English speakers to consistently use F0 to signal noun focus motivated further investigation of the use of other cues, as well as of inter- and intra-speaker variability in cue use.

B. INTENSITY RESULTS

While there was a noticeable difference between the two language groups in the use of pitch in the noun-focus condition, there was no significant group difference in the use of intensity across focus conditions. Comparing the maximum intensity of the two syllables within the focused word, both Mandarin and English speakers exhibited a similar high intensity on the stressed syllable, followed by a drop in intensity (Fig. 4 Left). Subtracting the maximum intensity of the noun from the maximum intensity of the adjective, both groups showed the expected positive values (intensity decrease) from the focused adjective to the post-focus noun as well as the expected negative values (intensity increase) from the pre-focus adjective to the focused noun (Fig. 4 Right). Thus, both language groups were consistent in realizing focused elements with increased intensity in both ADJ-focus and NOUN-focus conditions. This result is somewhat surprising given the finding of Ouyang & Kaiser (2015: 68) that in their Mandarin speakers' productions, "intensity cues only appear for corrective focus, not for new-information focus."

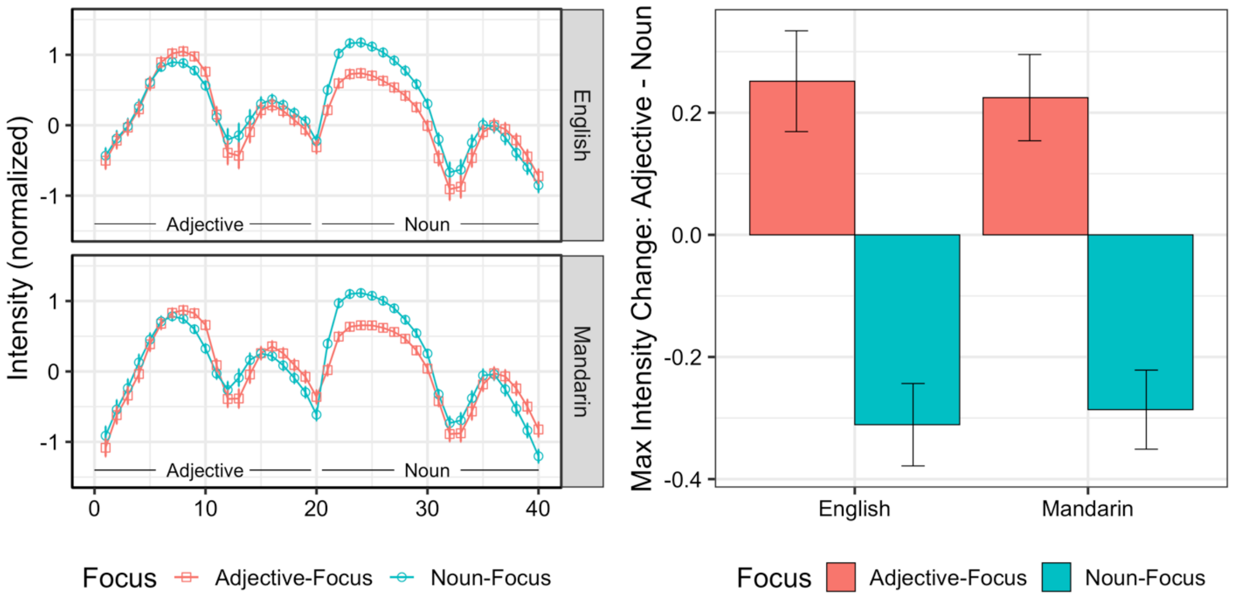

Figure 4. Average normalized intensity contour on normalized time (left) and average maximum intensity change from adjective to noun (right). Error bars represent 95% confidence intervals. y-axis indicate normalized intensity in z-score.

Thus far we have focused on the group averages for use of F0 and intensity cues to focus. However, the F0 contours of the English speakers in particular showed a good deal of variation (both intra- and inter-speaker). Fig. 5 illustrates the average F0 target phrase contours for two English speakers, one showing a peak on the focused word, and the other lacking a peak on a focused noun.

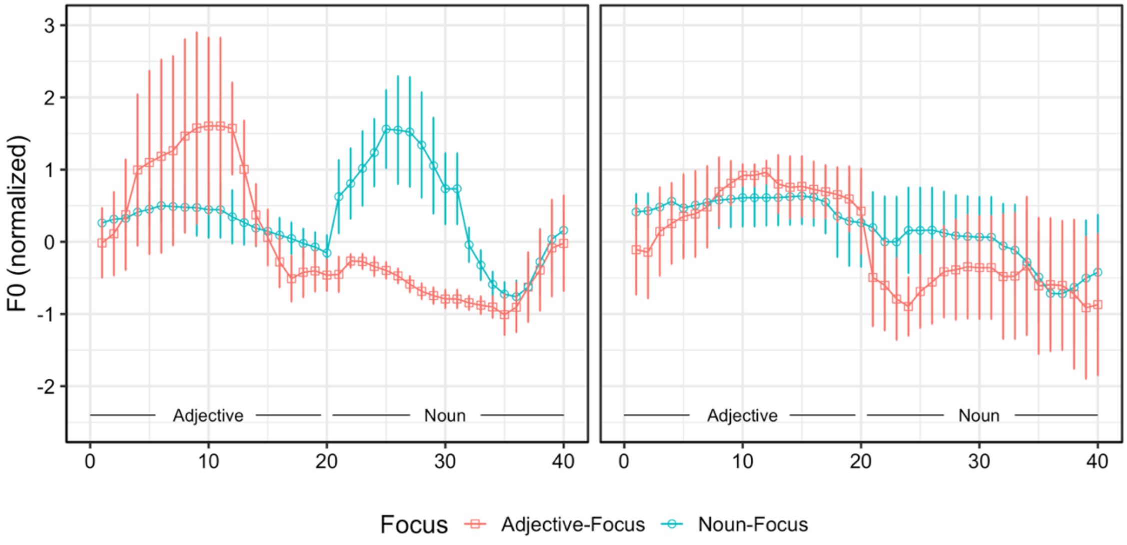

Figure 5. Averaged F0 contour of two English speakers: canonical NOUN-focus intonation (left) and a non-canonical NOUN-focus pattern (right). Error bars represent 95% confidence intervals.

Variation in the use of conventional pitch contours was found even within speakers, as illustrated by Fig. 6 showing two NOUN-focus tokens produced by a single English speaker.

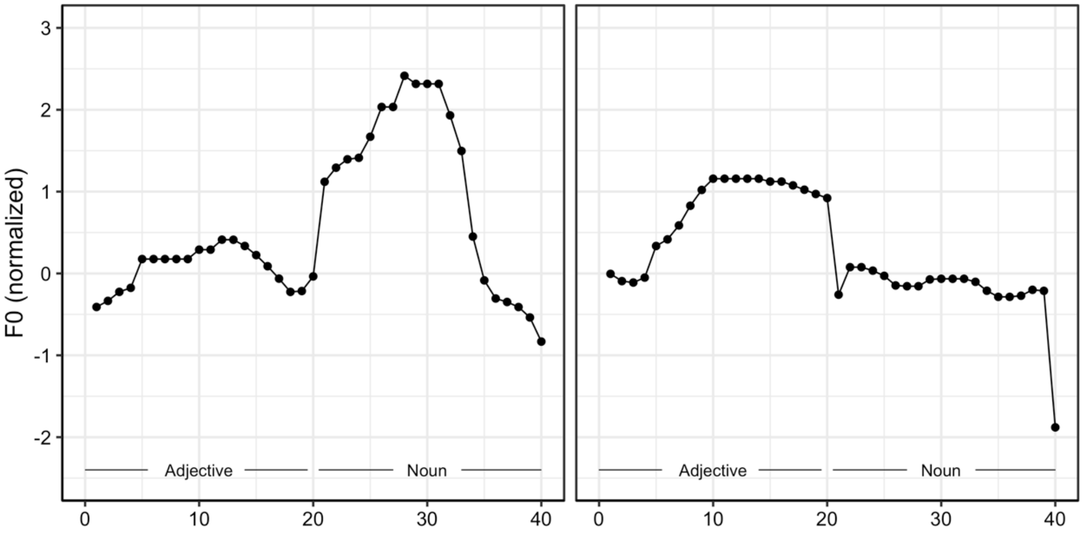

Figure 6. F0 contour of two tokens from a single English speaker: canonical NOUN-focus intonation (left: yellow OVAL) and non-canonical NOUN-focus intonation (right: yellow ARROW).

Noting the English speakers' inconsistency in the use of pitch to signal NOUN-focus, we considered the interplay of the use of pitch and intensity cues in individual tokens. Fig. 7 arranges tokens according to cue use, computed as the difference between focused and unfocused words within the target noun phrase in max F0 (x axis) and max intensity (y axis). Positive values reflect the canonical use of the cues (i.e., higher F0 and intensity in focused words than in unfocused words). The upper right quadrants of Fig. 7 contain tokens in which the focused word has both higher F0 and higher intensity; the upper left and lower right quadrants contain tokens with only one focus cue, and tokens in the lower left quadrant lack canonical use of either cue. Consistent with the group averages, the English data show a clear difference in the use of F0 for ADJ-focus vs. NOUN-focus tokens, while the Mandarin data show far greater overlap between the tokens from the two focus conditions.

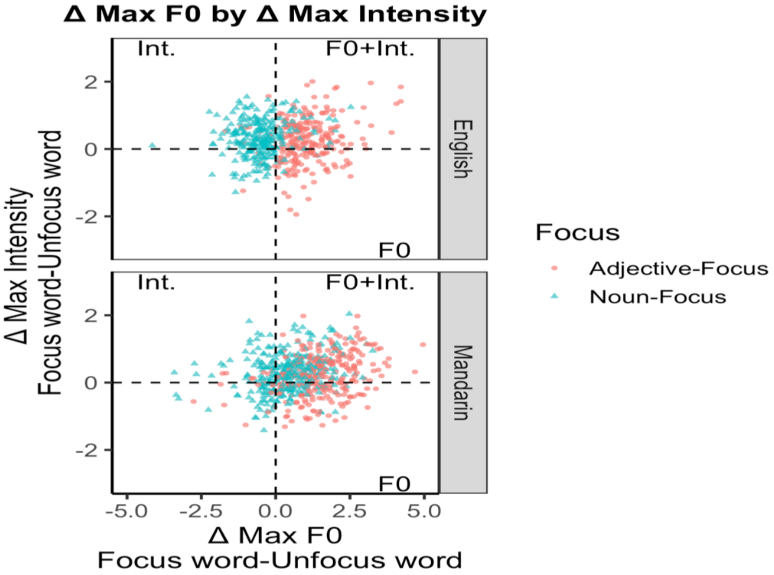

Figure 7. Use of pitch and intensity cues: maximum intensity difference (Focused - Unfocused word) and maximum F0 difference (Focused - Unfocused word).

C. DURATION RESULTS

An additional cue to English focus is duration, and group differences emerged in use of this cue as well – though again, only in the NOUN-focus condition. While for pitch, it was the English speakers who departed from the canonical pattern, for duration it was the Mandarin speakers who failed to use this cue to mark the focused noun. Despite phrase-specific variances in duration, both English and Mandarin speakers, on average, produced longer adjectives than nouns when the adjective was focused, showing positive values in Fig. 8 Right. For NOUN-focus, however, while the focused noun was longer than the adjective for English speakers (negative values in Fig. 8 Right), the nonfocused adjectives were still longer than the focused nouns in the Mandarin productions.

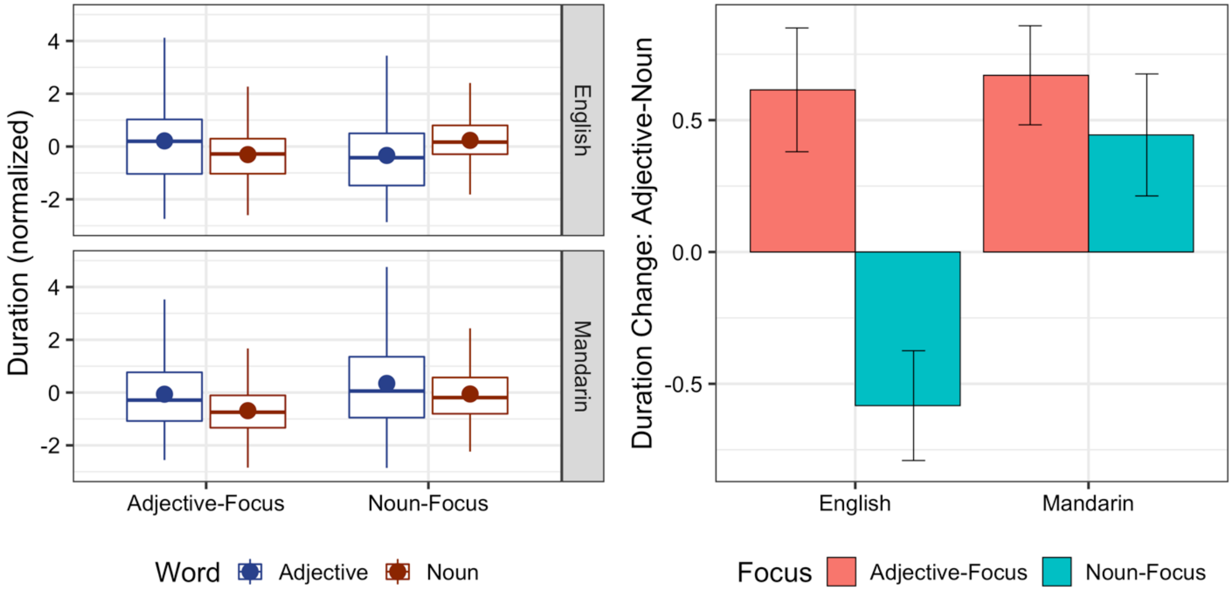

Figure 8. Normalized word duration boxplot with points representing means and horizontal lines representing medians (left); average duration change from adjective to noun with error bars representing 95% confidence intervals (right). The upper and lower end of the whisker in the box plot represent the highest and lowest values in 1.5 * IQR respectively.

D. STATISTICAL ANALYSES

We analyzed the three acoustic measurements with generalized linear mixed-effects models, using the lmer4 package (Bates et al. 2015) in R (R Core Team 2016). Models were fitted for each dependent variable (max F0/intensity in adjective minus max F0/intensity in noun and adjective duration minus noun duration) with language group (English vs. Mandarin) and focus position (ADJ vs. NOUN) as fixed effects. For the random effect structure, by-participant and by-item random intercepts, by-participant and by-item random slopes for the effect of focus position and by-item random slopes for the effect of language group were included. The predictors were dummy-coded, and the reference level in all models was ADJ-focus and English speakers. If a model failed to converge, we simplified the model by removing random effect terms and adopted the one with the maximal random effect structure that converged. The *p*-values were calculated using the lmerTest package (Kuznetsova et al. 2016). Conditional R^2 variance, the variance explained by fixed and random factors, was obtained using the MuMIn R package (Johnson 2014).

Table 1. Linear mixed-effects models for max F0 change, max intensity change and duration change from adjective to noun.

	Model 1: Max F0 Change				Model 2: Max Intensity Change				Model 3: Duration Change			
	R^2= 0.62				R^2=0.59				R^2=0.45			
Predictor	B	SE	t	p	B	SE	t	p	B	SE	t	p
(Intercept)	**1.92**	**0.15**	**6.77**	**<0.001**	**0.34**	**0.13**	**2.61**	**<0.05**	0.36	0.37	0.97	0.35
Focus position: Noun-Focus	**-0.54**	**0.22**	**-2.48**	**<0.05**	**-0.64**	**0.10**	**-6.43**	**<0.001**	**-0.87**	**0.17**	**-5.21**	**<0.001**
Language group: Mandarin	**0.46**	**0.19**	**2.31**	**<0.05**	-0.001	0.12	-0.01	0.99	0.01	0.24	0.04	0.97
Focus position: Noun-Focus * Language group: Mandarin	**-2.51**	**0.80**	**-3.15**	**<0.01**	0.02	0.13	0.17	0.87	**1.02**	**0.22**	**4.56**	**<0.001**

The results of the models confirm that Mandarin and English speakers used F0 and duration cues differently when the noun was focused, as the coefficient of the interaction term (Focus position*Language group) in Models 1 and 3 was significant. In contrast, the two groups did not differ in their use of intensity cues – both groups had higher intensity in focused words than in unfocused words, whether the words were adjectives or nouns. Model 1 also shows that in the ADJ-focus condition, the Mandarin speakers had a significantly larger F0 fall from adjective to noun than the English speakers, confirming the observation that Mandarin speakers used a wider range of F0.

4. DISCUSSION & CONCLUSION

The present study compared L1 Mandarin and L1 English speakers' use of prosodic cues in the production of English contrastive focus. While both language groups exhibited similar use of prosodic cues when the adjective was focused, the two groups differed in their use of duration and pitch in the NOUN-focus condition. The direction of difference in the use of pitch cues was unexpected: it was actually the non-native speakers who adhered most closely to the textbook descriptions of English focus prosody, with a pitch peak on the focused element in both ADJ-focus and NOUN-focus conditions. In contrast, the English speakers frequently relied on intensity and duration to signal noun focus. Closer examination of individual productions revealed that many of the L1-English speakers' productions exhibited creakiness, especially towards the end of the target phrase (the position of the focused noun). Their productions reflect an innovative intonation pattern noted among younger speakers of American English, in which vocal fry increases toward final position (Yuasa 2010, Wolk et al. 2012, Abdelli-Beruh et al. 2014). While young speakers' use of creakiness and vocal fry has been much discussed in recent literature, our study is the first to uncover how this new intonation pattern might influence focus intonation. We assume that the steady declination of pitch in this innovative pattern is antithetical to the realization of a pitch peak on a focused element at the end of the phrase, so for the users of this innovative pattern, non-pitch cues are primary markers for the NOUN-focus condition. Since more than 90 percent of our L1-Mandarin participants had taken at least one semester of English instruction since arriving in the US, and since the importance of suprasegmentals has been increasingly emphasized in second language instruction, the use of the more conservative standard pattern by Mandarin speakers most likely reflects explicit instruction in contrastive intonation prosody.

Abstracting away from the differences between ADJ-focus and NOUN-focus, our study revealed that while English speakers were able to use increased pitch, duration, and intensity to signal focus, the Mandarin speakers seemed to rely on pitch and intensity rather than duration. The

Mandarin speakers' ability to use pitch cues to signal focus is perhaps related to the fact that expansion of the pitch range of a lexical tone contour is a feature of focus realization in Mandarin (e.g., Xu 1999, Chen & Braun 2006, Chen & Gussenhoven 2008, Lee et al. 2015, Ouyang & Kaiser 2015, Wang et al. to appear). The Mandarin speakers' failure to reliably use duration to signal focus is less expected, since expanded duration has been argued to be associated with focus in Mandarin (Xu 1999, Chen & Braun 2006, Chen & Gussenhoven 2008, Lee et al. 2015, Ouyang & Kaiser 2015), though the importance of duration relative to pitch-related factors depends on the lexical tone (Wang et al. to appear). The absence of a duration effect in our study may reflect the fact that in NOUN-focus instructions (where the adjective was repeated but the noun was contrastive), some Mandarin speakers showed a tendency to elongate the adjective, presumably while planning for production of the non-repeated noun. However, a similar difference between native and L1-Mandarin use of duration in focus was found in earlier studies as well. Chen et al. (2001) found that while L1-Mandarin speakers were able to manipulate F0 to mark English sentence stress, they produced focused words with shorter durations than did English speakers. Similarly, Chen et al. (2015) found that L1-Mandarin college freshmen showed a slightly lower duration increase in focused elements than either native English speakers or L1-Mandarin college seniors, suggesting that the use of duration to signal focus became more English-like with increased exposure to English. Both Chen et al. (2001) and Chen et al. (2015) reported English-like use of intensity for their Mandarin speakers' productions of sentence stress, consistent with our findings. This use of intensity is somewhat surprising, given that studies of focus realization in Mandarin show little or no reliance on intensity; Ouyang and Kaiser (2015) report that "intensity cues only appear for corrective focus, not for new-information focus" and Wang et al. (to appear) report that in their study of corrective focus, "Intensity did not play a role" except when the focused word was Tone 3. However, since intensity and F0 tend to be correlated in the realization of lexical tones in Mandarin (Whalen & Xu 1992), it may simply be that Mandarin speakers view the increase in intensity as an automatic correlate of the increased pitch of the English contrastive focus rather than an independently manipulable cue to focus.

One general finding of our study is that comparisons of native and non-native speech often take an idealized native standard as the baseline for comparison, designating any deviation from this standard by non-native speakers as a reflection of incomplete acquisition. In our study, we found that the non-native speakers' productions actually adhered more closely to this idealized canonical version of English, while native speakers showed a considerable range of variation in the realization of focus.

ACKNOWLEDGMENTS

This material is based upon work supported by NSF under Grant # IBSS-1519908. We would like to thank Stephen Buttner for his help in segmentation, and the audience at ASA Meeting at Louisville for their comments.

REFERENCES

Abdelli-Beruh, N. B., Wolk, L., and Slavin, D. (2014). "Prevalence of vocal fry in young adult male American English speakers," Journal of Voice. 28(2). DOI: 10.1016/j.jvoice.2013.08.011.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). "Fitting Linear Mixed-Effects Models Using lme4", Journal of Statistical Software 67(1). 1-48.

Boersma, P., and Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.54, retrieved 6 June 2019 from http://www.praat.org/

Chen, Y. (2010). "Post-focus compression — now you see it, now you don't". *Journal of Phonetics 38*, 517-525.

Chen, Y., and Braun, B. (2006). "Prosodic realization in information structure categories in standard Chinese," In Hoffmann, R., Mixdorff, H. (Eds.), Speech prosody 3. Dresden, Germany: TUD Press.

Chen, Y., and Gussenhoven, C. (2008). "Emphasis and tonal implementation in Standard Chinese," Journal of Phonetics 36, 724-746.

Chen, Y., Robb, M. P., Gilbert, H., R., and Lerman, J. W. (2001). "A study of sentence stress production in Mandarin speakers of American English," Journal of the Acoustical Society of America 109(4), 1681-1690.

Chen, Y., Xu, Y., and Guion-Anderson, S. (2015). "Prosodic Realization of Focus in Bilingual Production of Southern Min and Mandarin," Phonetica 71: 249–270.

Johnson, P.C.D. (2014). "Extension of Nakagawa & Schielzeth's *R*2GLMM to random slopes models," Methods in Ecology and Evolution 5, 44-946.

Kao, S., Hwang, J., Baek, H., Takahashi, C., and Broselow, E. (2016). "International teaching assistants' production of focus intonation," Proceedings of Meetings on Acoustics 26(1). 1-13. https://doi.org/10.1121/2.0000356.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models," Journal of Statistical Software 82(13). 1-26.

Lee, Y., Wang, C., Chen, S., Adda-Decker, M., Amelot, A., Nambu, S., and Liberman, M. (2015). "A cross-linguistic study of prosodic focus," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://ieeexplore.ieee.org/document/7178873

Ouyang, I. C., and Kaiser, E. (2015). "Prosody and information structure in a tone language: an investigation of Mandarin Chinese. Language," Cognition and Neuroscience 30 (1&2), 57-72.

Pearson Education. (2010). Versant English test description and validation manual. Palo Alto, CA: Author. Retrieved from www.versanttest.com/products/english.jsp.

Pierrehumbert, J., and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack, (eds) *Intentions in Communication*, Bradford Books, MIT Press, Cambridge MA. 271-311.

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Wang, T., Jun, L., Lee, Y., and Lee, Y. (2020). "The interaction of tone and prosodic focus in Mandarin Chinese," Language and Linguistics 20(2).

Whalen, D. H., and Xu, Y. (1992). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica 49(1), 25-47.

Wolk, L., Abdelli-Beruh, N. B., and Slavin, D. (2012). "Habitual use of vocal fry in Young Adult Female Speakers," Journal of Voice 26(3).

Xu, Y. (1999). "Effects of tone and focus on the formation and alignment of f0 contours," Journal of Phonetics 27. 55-105.

Xu, Y. (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.

Yuasa, I. P., (2010). "Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women?" American Speech 85(3), 315-337.

Zahorian, S. A. and Hu, H. (2008). “Spectral/Temporal Method for Robust Fundamental Frequency Tracking,” Journal of the Acoustical Society of America 123, 4559-4571.