English Focus Prosody Processing and Production by Mandarin Speakers

Chikako Takahashi, Sophia Kao, Hyunah Baek, Alex H.L. Yeung, Marie Huffman, Ellen Broselow, Jiwon Hwang

Stony Brook University

Research questions

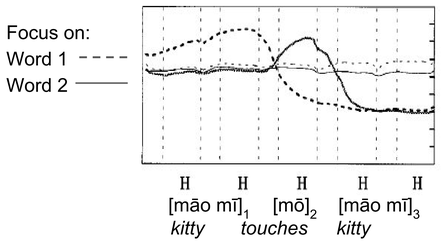
- Do native and non-native speakers produce and process prosodic focus cues in English similarly?
- Is there a relationship between speakers' processing and production of focus prosody?

Background

Production of sentence focus

Both English and Mandarin use prosody to signal focus.

(Cooper et al. 1985; Xu 1999)



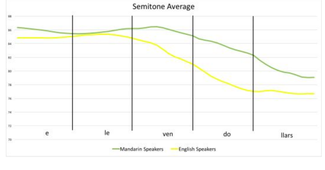
(Xu 1999)

However, in production of corrective focus, Mandarin speakers

- misaligned the pitch peak and failed to utilize intensity cues,
- which affected the perceived naturalness of their English focus intonation.

(Kao et al. 2016)

Q: *The price of the bus ticket is twenty dollars, right?*
A: *No, the price of the bus ticket is* ***eleven dollars****.*



Processing of sentence focus

- English speakers use prosodic cues (e.g., contrastive pitch accents) during discourse comprehension (Ito & Speer 2008).
- Mandarin speakers are less successful in detecting a prominent word in English sentences, but they use similar criteria (e.g., pitch, intensity, duration) (Rosenberg et al. 2010).

Exp 1: Production of contrastive focus

Participants

- 21 native English speakers (ES)
- 21 non-native speakers of English whose L1 is Mandarin (MS)

Target Phrases: 12 ADJ + N (both σ́σ)

- ADJ (e.g.,*yellow, orange, navy)*
- Noun (e.g., *arrow, diamond, oval*)

Procedure and Analysis

- Elicited Instruction: *Put the Adj + Noun over the* ***Adj + Noun*** (Figure 1)
- ProsodyPro (Xu 2013) were used to measure pitch peak and average (F0, semitone) per syllable and word.
- 44 out of 252 phrases of ES were excluded due to upward intonation.

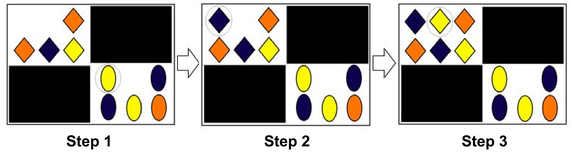


Figure 1. Production Experiment
Step 1: *Put the yellow oval over the NAVY oval, please.*
Step 2: *Put the navy diamond over the ORANGE diamond, please.*
Step 3: *Put the yellow diamond over the NAVY diamond, please.*

Results

- ES aligned pitch peak with the stressed syllable of the focused word, but MS placed the peak later, within the last syllable of the focused word (Figure 3).

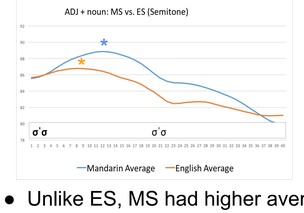


Figure 3. Time-normalized F0 contour of ADJ+noun by MS and ES

- Unlike ES, MS had higher average pitch on the final syllable of focused adj than on the initial (stressed) syllable (Figure 4).
- This indicates that they aligned the pitch peak not with the stressed syllable but with the word edge.

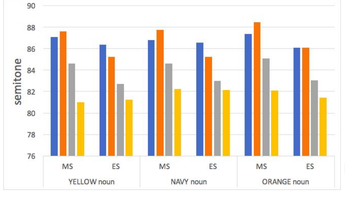


Figure 4. Average F0 per syllable of ADJ+noun by MS and ES

Exp 2: Processing of contrastive focus

Participants

- Same groups of ES and MS as in Production Experiment

Target Phrases: ADJ + N (both σ́σ)

- ADJ (e.g., *ivory, purple*, *flowered, dotted*)
- Noun (e.g., *mittens, necklace, sweater*)

Procedure

- Instruction 1: *Click on the Adj + N* (Figure 5)
- Instruction 2: *Now click on the Adj + N* (Figure 6)
  (either felicitous or infelicitous prosody; Table 1)
- For instruction 2, reaction time (RT) was measured from the offset of N to the time of the response.

Figure 5. Processing Experiment: instruction 1
(*Click on the purple mittens.*)

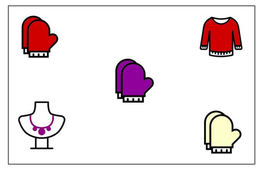


Figure 6. Processing Experiment: instruction 2
(*Now click on the scarlet mittens.*)

Table 1. Processing Experiment Conditions

Conditions	Instruction	Adj contrast
Felicitous prosody	*Click on the purple mittens. Now click on the...*	*SCARLET mittens*
Infelicitous prosody	*Click on the scarlet necklace. Now click on the ...*	*SCARLET mittens*

Results

- MS RTs significantly slower overall than ES RTs (Figure 7) although no difference in accuracy (MS=96.8%, ES=97.0%) .
- Both groups had faster (but not significant) RTs in felicitous condition than infelicitous condition.
- However, wider range of RTs for MS than ES (SD MS=133.97, ES=117.40).
- ES more likely to respond before noun offset (Figure 8).
- More early responses in felicitous vs. infelicitous condition for both ES (37 vs. 17) and MS (14 vs. 9), but 12/14 MS early felicitous responses came from one speaker.
- Facilitation effect of prosodic cues is more common for ES.

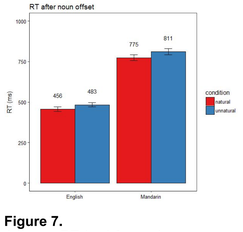


Figure 7.
Average RT (ms) for each group comparing felicitous vs. infelicitous conditions

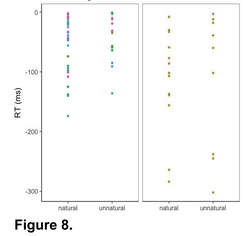


Figure 8.
RT recorded before the offset of the noun (Data point colors by subjects)

Discussion & Conclusion

Production

- The two groups differed in their realization of focus, with English speakers tending to align the pitch peak with the stressed syllable and Mandarin speakers with the right edge of the focused word.
- MS - ES differences in production of contrastive focus mirror differences in production of corrective focus (Kao et al. 2016).

Processing

- Both groups responded more quickly to instructions with felicitous vs. infelicitous prosody, although English speakers' response times were significantly faster in both conditions.
- Group differences did not reach significance (possible ceiling effects).

Production - Processing Relationship

- Although Mandarin speakers showed Mandarin-like realization of focus in their production, they could nonetheless use the English prosodic patterns in their processing.

References

- Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142-2156.
- Ito, K. & Speer, S. R. (2008). Anticipatory effects of intonation: eye movements during instructed visual search. *Journal of Memory and Language* 58, 541-573.
- Kao, S., Hwang, J., Baek, H., Takahashi, C., & Broselow, E. (2016). International teaching assistants' production of focus intonation. *Proceedings of Meetings on Acoustics 26* (1), 1-13.
- Rosenberg, Andrew, Julia Hirschberg, and Kim Manis. 2010. Perception of English prominence by native Mandarin Chinese speakers. *Speech Prosody 2010*.
- 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.

Acknowledgement

This material is based upon work supported by NSF under Grant # IBSS-1519908. Special thanks to Sharon Benedett and Kevin Henderson for the stimuli construction.