COMPARISON OF ENGLISH NARROW FOCUS PRODUCTION BY L1 ENGLISH, BEIJING AND TAIWAN MANDARIN SPEAKERS

Tanya Visceglia [1], *Chao-yu Su* [2] *and Chiu-yu Tseng*[2]

[1] Education Center for Humanities and Social Sciences, National Yang Ming University, Taipei, Taiwan
[2] Phonetics Lab, Institute of Linguistics, Academia Sinica, Taipei, Taiwan

ABSTRACT

L1 English and two varieties of L1 Mandarin English speech data were extracted from the Taiwan AESOP corpus (Asian English Speech cOrpus Project) for the purpose of investigating differences in the realization of English narrow focus by L1 speakers of North American English, Taiwan Mandarin and Beijing Mandarin. Results show the combined effect of two patterns of L2 focus production: general underdifferentiation of on-focus and post-focus contrasts, which was exhibited by both L2 speaker groups, and transfer of L1-specific prosodic features, which can be argued to represent the source of difference between the two L2 groups. Overall, on-focus/post-focus contrasts in mean F0, amplitude and pitch range were realized most robustly by L1 English speakers. L1 Taiwan Mandarin speakers produced a smaller increase in mean F0 and amplitude for on-focus constituents and much smaller decrease in mean F0 and amplitude on post-focus constituents than L1 English speakers did, whereas Beijing Mandarin speakers produced no increase in mean F0 in on-focus constituents, and the smallest decrease in mean F0 on post-focus constituents, but a 35% higher post-focus compression of intensity than Taiwan Mandarin speakers did. Notably, both L2 speaker groups failed to produce post-focus compression of pitch range, which has been shown to be a highly salient cue to the presence of focus in English.

Index Terms*:* L2 English, prosodic focus, Beijing Mandarin, Taiwan Mandarin, post-focus compression (PFC)

1. INTRODUCTION

The Asian English Speech cOrpus Project (AESOP), a multi-national research effort, was designed to collect and compare L2 English speech corpora in order to derive a set of core properties common to all varieties of Asian English, as well as to discover features that are particular to individual dialects [1]. The data presented here represent part of the ongoing research conducted by the TWNAESOP research team, which was formed to develop a systematic understanding of the acoustic characteristics in L2 Taiwan English speech. It should be emphasized here that the major research goal of AESOP is not to normalize Asian Englishes to any particular ENL standard, but instead to catalog and ultimately predict similarities and differences among the varieties of English found across Asia. It is hoped that our collective findings will contribute to the further development of English speech tools and interfaces such that these applications can be better tailored to accommodate Asian users.

Moreover, current research has refuted the idea that L2 speech necessarily becomes less intelligible as a result of being different from native pronunciation. Many studies have demonstrated weak or no correlation between global accent ratings and level of overall intelligibility [2]. Thus, our analysis of Taiwan English pronunciation is not as much concerned with accentedness, defined as how different a speaker's pronunciation is perceived to be from that of the L1 community, as it is with intelligibility, defined as how well the speaker's intended message is understood, and comprehensibility, defined as perceived level of difficulty in following the speaker's intended meaning [3].

Among the suprasegmental features which have been found to correlate with comprehensibility in L2 speech is non-targetlike realization of the prosodic cues to information structure in continuous speech. Prosodic focus serves to highlight aspects of information structure in speech by distinguishing between given versus new information, or signaling contrastive interpretation. Realization of prosodic focus is a challenge for L2 speakers, and failure to sufficiently emphasize focused information has been demonstrated to reduce L2 speakers' level of comprehensibility [4] That is to say, L2 speakers' failure to realize focus may contribute to listeners' difficulty in extracting their intended meaning or in following their discourse structure [5].

Recent research investigating cross-linguistic differences in the acoustic cues used to realize focus provides a foundation for comparison of L1 and L2 focus production, allowing us to determine whether L2 differences can be attributed to transfer of first language prosodic strategies, or whether some L2-universal prosodic constraints, such as the tendency of L2 speakers to produce

shorter phrase groupings, may also contribute to differences between L1 and L2 realization of focus [6].

It has been claimed that Beijing Mandarin, Taiwan Mandarin and English represent different strategies for realization of narrow focus: English and Beijing Mandarin exhibit on-focus f0 range/intensity expansion and post-focus f0 range/intensity compression (PFC), and Taiwan Mandarin exhibits on-focus increase in intensity and duration, but no post-focus compression of f0 range or intensity [7,8]. Thus, comparison of Taiwan and Beijing Mandarin speakers' production of focus in English would help to investigate the question of whether differences between L1 and L2 realization of narrow focus can be attributed to transfer of L1 prosodic strategies, L2-specific processing strategies, or a combination of the two.

2. METHOD

2.1 Materials

The speech data analyzed in this study represent a subset of the core phonetic experimental tasks developed by AESOP. The AESOP materials, which include sets of both read and spontaneous speech tasks, as well as a recording platform and recording protocol manual specifically designed for this project [9] were developed in a collaborative effort by AESOP teams in Taiwan, Japan and Hong Kong. The current data set consists of fifteen sets of question/answer pairs. Question sentences were designed to provide a context of given information. Answers to those questions are all in the form of declarative sentences. Each sentence contains one word in contrastive focus, which should be expanded, and one or more words in the post-focus position, which should be compressed. Each answer sentence contains a different lexical item in narrow focus; the fifteen items were chosen to represent a range of syllabicities and stress types. Participants were required to produce the answer sentences only, and to stress the word appearing in bold capital letters. An example appears below:

Background sentence:

Will 3-day delivery be fast enough?

Experimental sentence:

No. We need **OVERNIGHT** delivery.

2.2 Participants

Participants were recruited from university campuses and research institutions in Taiwan. The eight L1 English speakers (4 male, 4 female) are all native speakers of North American English. Nine of the L2 speakers (5 male, 4 female) are university students who are native speakers of Taiwan Mandarin and who have received at least ten years of English instruction. Most of them also have some knowledge of Taiwanese, though they are all strongly Mandarin dominant. The nine female speakers of Beijing Mandarin, also university students who had received at least ten years of English instruction, were recruited for participation from Ming Chuan University's cross-strait exchange program.

2.3 Procedure

Speech data were recorded in quiet rooms directly into a laptop computer, using a recording platform developed specifically for this project, with pre-loaded experimental sentences each appearing individually on a computer screen. Participants wore head-mounted Sennheiser PC155 microphones positioned 2 cm away from their mouths, and they were instructed to speak naturally at a normal rate and volume.

2.4 Data Analysis

390 English sentence tokens (120 L1 English, 135 L1 Taiwan Mandarin, 135 L1 Beijing Mandarin) were selected for analysis. They were sampled at a rate of 16kHz with a quantization of 16 bits. All data were pre-processed for segmental labeling using phone sets from the CMU electronic dictionary, which were then manually spot-checked by trained transcribers for segmental alignment. A trained transcriber additionally labeled two constituents in each sentence: on-focus (consisting of the focused word itself) and post-focus (the words following the focused item up to the next intermediate or intonation phrase break) [9].
Duration, mean F0, F0 range (both treated logarithmically), and average intensity were derived for the on-focus and post-focus constituents for the purpose of comparison between L1 and the two L2 speaker groups. To allow comparison across individual speakers and sentences, f0, duration and intensity were subjected to z-score normalization [10]. The normalization function appears below, in which $X_s(t)$ represents the original value of original feature vector in frame *t* of sentence *S* and $X'_s(t)$ represents the normalized values, while μ_s and σ_s represent the mean and standard deviation of whole frames in sentence *S*.

$$X'_s(t) = \frac{X_s(t) - \mu_s}{\sigma_s}$$

Note that the results presented in Section 3 are represented proportionally, abstracted away from their original units of measurement, in order to highlight differences in the relative salience of the acoustic contrasts produced by the three groups. The zero value in the bar graphs below represents the average value of that parameter across participants; positive and negative values correspond to measurements higher and lower than average.

3. RESULTS

3.1 Duration

Figure 1 below shows differences in on-focus expansion and post-focus compression of duration. Both L1 English speakers and Taiwan Mandarin speakers appear to expand duration in post-focus position (Taiwan Mandarin speakers to a greater extent), whereas Beijing Mandarin speakers produce on-focus and post-focus constituents of approximately equal duration. Implications will be discussed in Section 4.

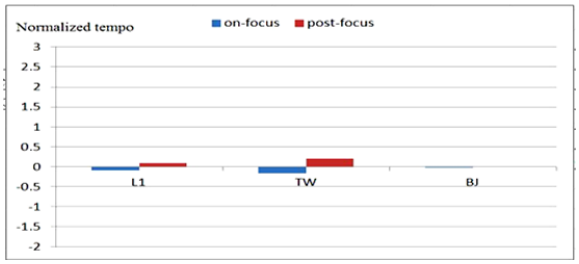

Figure 1: *Temporal comparison of on-focus and post-focus constituents for L1, TW and BJ speakers*

Table 1. *Temporal comparison of on-focus and post-focus constituents for L1, TW and BJ speakers*

	L1	TW	BJ
On-focus	-0.09	-0.16	-0.009
Post-focus	0.09	0.2	0.0024

3.2 Mean F0

Comparison of derived mean F0 between focused and post-focus constituents is shown in Figure 2. Although L1 and both L2 speaker groups both maintain mean F0 contrasts, the difference produced by L1 speakers is much larger. We also see a difference between the two L2 speaker groups: Taiwan Mandarin speakers produce both on-focus F0 raising and post-focus lowering, whereas Beijing Mandarin speakers exhibit only a slight lowering in the post-focus constituent. An interpretation will be offered in Section 4.

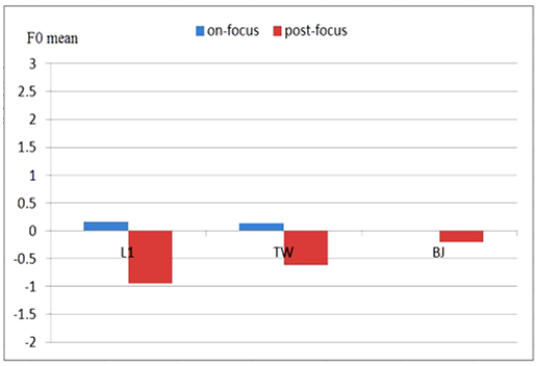

Figure 2: *Mean F0 comparison of on-focus and post-focus constituents by L1, TW and BJ speakers.*

Table 2. *Mean F0 comparison between on-focus and post-focus constituents for L1 and L2 speakers*

	L1	TW	BJ
On-focus	0.17	0.13	0.0098
Post-focus	-0.94	-0.62	-0.2003

3.4 F0 Range

Measurement of on-focus F0 range expansion and post-focus compression across speaker groups reveals that post-focus compression is realized strongly by L1 speakers, but is entirely absent for both L2 speaker groups. Possible interpretations will be discussed in Section 4.

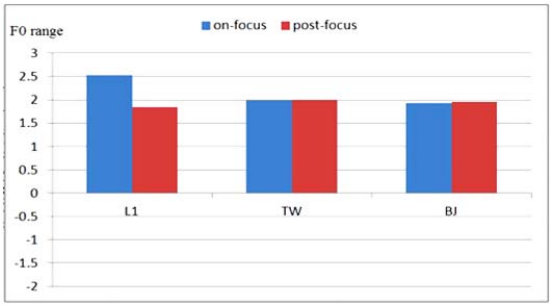

Figure 3: *F0 range comparison between on-focus and post-focus constituents for L1, TW and BJ speakers.*

Table 3. *F0 range comparison of on-focus and post-focus constituents for L1, BJ and TW speakers*

	L1	TW	BJ
On-focus	2.53	1.99	1.93
Post-focus	1.84	2	1.96

3.3 Intensity

Comparison of intensity contrasts between on-focus and post-focus areas (see Table 4) reveals that an intensity contrast is produced by all three speaker groups. Again, L1 speakers produced the largest contrast, and we see a clear L1 effect across the two L2 speaker groups: Beijing Mandarin speakers realize the intensity contrast much more robustly than Taiwan Mandarin speakers do (L1 on-focus 63% higher than post-focus; Beijing Mandarin 55% higher; Taiwan Mandarin 20% higher). Section 4 will discuss possible interpretations for this finding.

Figure 4: *Intensity comparison between on-focus and post-focus constituents for L1 and L2 speakers.*

Table 4. *Intensity comparison between on-focus and post-focus for L1 and L2 speakers*

	L1	TW	BJ
On-focus	1.24	1.01	0.4526
Post-focus	0.79	0.82	0.2525

4. DISCUSSION

4.1 Duration

Taiwan Mandarin speakers tend to exhibit longer durations in both the focused and post-focus constituents than both L1 English and Beijing Mandarin speakers do. Since these data have been normalized to subtract the effect of between-group differences in speech rate, we can reasonably infer that L2 speakers' productions are longer in duration because they are less likely to reduce the unstressed syllables in on-focus words or to reduce words in the post-focus constituent. L2 speakers' failure to reduce and/or delete de-phrased and unstressed syllables is a well-known characteristic of L2 speech rhythm, particularly for speakers whose L1 is syllable- or mora-timed [11]. Thus, transfer of Taiwan Mandarin's syllable-timed template or focus marking strategy is a likely explanation for this finding, as bilingual Taiwanese/Mandarin speakers in Taiwan have also been found to exhibit post-focus lengthening in Mandarin [7]. Both Taiwan Mandarin and L1 English speakers' expansion of the post-focus constituent may also be attributable to final lengthening effects, since all post-focus constituents were located before a major or minor phrase break. Beijing Mandarin speakers, in contrast, produced equal durations in the focused and post-focus constituents. Their failure to produce final lengthening, the presence of which has been demonstrated in Mandarin [12], may be related to their difficulty integrating higher levels of prosodic information, such as boundary cues, into production of L2 speech.

4.2 Mean F0

Although both L1 English and both L2 speaker groups contrast average F0 between on-focus and post-focus constituents, L1 speakers do so to a much greater extent. Moreover, Beijing Mandarin speakers do not exhibit on-focus raising and produce a lesser degree of F0 lowering in the post-focus constituent than Taiwan Mandarin speakers do, although F0 lowering is an acoustic correlate of post-focus in Beijing, but not Taiwan Mandarin. Weak realization of F0 height contrast was also observed in an earlier study of L1 Taiwan Mandarin speakers' production of lexical stress [13]. Participants were able to acoustically differentiate stressed and unstressed syllables in English multisyllabic words when those words were presented in isolation, but not when they were embedded in higher-level prosodic contexts, i.e. in narrow-focus conditions or at sentence boundaries. We believe that the competing processing demands of simultaneously encoding segmental, lexical stress, and focus information create a processing overload, which weakens realization of the acoustic contrasts used to mark prominence at both the lexical and utterance levels.

4.3 F0 Range

Post-focus F0 range compression is entirely absent in both groups of L2 speakers' data (see Figure 3). Pitch range is almost identical in the on-focus and post-focus constituents for both L2 speaker groups; whereas L1 English speakers' pitch range in the on-focus constituent is 72.9% larger than in the post-focus constituent. For the Taiwan Mandarin speakers, we could plausibly attribute this absence to L1 prosodic transfer, as Taiwan Mandarin has been reported to exhibit no post-focus compression. However, PFC has been reported to be a feature of Beijing Mandarin [7], so Beijing speakers' failure to transfer this property is puzzling, particularly in light of the fact that they were able to positively transfer post-focus compression of amplitude. Moreover, post-focus compression has been reported to be a highly salient feature in perception of narrow focus by L1 speakers [8], so failure to produce pitch range compression may result in listeners' failure to perceive the speaker's intended focus. This observation has been informally

confirmed by pilot native listener judgments and will be the subject of more detailed investigation in future perception studies.

4.4 Intensity

Intensity contrasts between on-focus and post-focus constituents were produced by all three speaker groups. Again, L1 speakers realized the contrast more robustly than either group of L2 speakers (on-focus 63% greater than post-focus). As with mean F0 and duration, an L1 effect can be observed across the Beijing and Taiwan speakers. On-focus intensity expansion has been reported to be a salient component of both Taiwan and Beijing Mandarin's prosodic realization of narrow focus. However, post-focus compression of intensity has been reported in Beijing Mandarin only [14]. As for the Beijing Mandarin speakers in this experiment, although it appears that they realized the intensity contrast much more strongly than Taiwan Mandarin speakers did (BJ on-focus 55%> post-focus; TW on-focus 20%> post-focus), the substantial amount of within-group variation observed in both L2 groups impedes interpretation of those results.

5. CONCLUSION

The data presented above suggest the combined effect of two patterns in L2 focus production: general underdifferentiation of on-focus/post-focus contrasts, which was found in both L2 speaker groups, and transfer of L1-specific prosodic features, which comprise the source of difference between the two L2 groups. Differences observed between Taiwan and Beijing speakers' production of duration and intensity in post-focus constituents provide support for the claim that Taiwan and Beijing Mandarin represent two different strategies for realization of prosodic focus [14]. However, both L2 speaker groups failed to produce post-focus compression of pitch range, although this feature is shared by Beijing Mandarin and English. Future research will examine production of these features in English by speakers of other languages which have been claimed to exhibit PFC, such as German and Mongolian, and speakers of languages without PFC, such as Cantonese, Korean and Japanese.

As for weak realization of contrasts, we believe that weakening or absence of the prosodic cues used to realize narrow focus may contribute to the perception that L2 speech is insufficiently differentiated with respect to marking of information structure. In future research, we plan to design a perception study using LPC resynthesis to investigate the relative salience of individual and combined acoustic cues to L1 and L2 listeners' perception of speakers' information structure.

7. REFERENCES

[1] Meng, H., Tseng, C., Kondo, M., Harrison, A. and Visceglia, T., "Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper" Interspeech 2009 1715-1718. Brighton.

[2] Munro, M. J. 2008. Foreign accent and speech intelligibility. In Hansen Edwards, J. G. & Zampini, M. L. (Eds.). *Phonology and Second Language Acquisition* (pp. 193-218). Amsterdam: John Benjamins.

[3] Derwing, T. M. and Munro, M. J. 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s *Studies in Second Language Acquisition*, *19*, 1-16.

[4] Warren, P., Elgort, I., and Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. Language Learning & Technology, 13(3), 87-102.

[5] Pickering, L. (2004). "The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse" *English for Specific Purposes* (23) 19-43.

[6] Ueyama, M. & Jun, S (1998) "Focus realization in Japanese English and Korean English intonation." vol. 7: 629-645, CSLI. Cambridge Univ. Press.

[7] Chen, S.W., Wang, B. and Xu, Y. (2009). Closely related languages, different ways of realizing focus. *Interspeech 2009*, Brighton, UK

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

[9] Beckman, M.E., & Pierrehumbert, J.B.(1986). Intonational structure in Japanese and English. Phonology Yearbook, 3, 255–309.

[10] Richard J. Larsen and Morris L. Marx (2000) An Introduction to Mathematical Statistics and Its Applications, Third Edition, 282.

[11] Jian, H. L. (2004). "On the syllable timing in Taiwan English" Proceedings of Speech Prosody 2004, Nara.

[12] Tseng, C. (2003) "On the role of intonation in the organization of Mandarin Chinese speech prosody", In *EUROSPEECH-2003*, 481-484.

[13] Visceglia, T., Tseng, C, Su, Z. and Huang, C. (2010) "Interaction of Lexical and Sentence Prosody in Taiwan L2 English" SLaTE Workshop, Interspeech 2010. Tokyo.

[14] Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong: 152-155.