Im, S. & Ahn, H. (2020). Perception of prosodic boundaries by native English speakers and Korean learners of English. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 218–228). Ames, IA: Iowa State University.

PERCEPTION OF PROSODIC BOUNDARIES BY NATIVE ENGLISH SPEAKERS AND KOREAN LEARNERS OF ENGLISH

Suyeon Im, Hanyang University
Hyunkee Ahn, Seoul National University

This study investigates how native English speakers and Korean learners of English perceive prosodic boundaries as a function of phonological features and acoustic cues in clear speech in English. Korean has more boundary types and tones than English. Due to their complex L1 prosody, Korean learners of English in this study were expected to have few difficulties in perceiving English boundaries while relying mostly on pitch among acoustic cues. Results showed that Korean learners of English rated prosodic boundaries similarly with native English speakers. Both groups of speakers were found to rely on F0 and duration, but Korean learners of English tended to use F0 in their boundary judgments to a greater extent than did native English speakers. Our findings present evidence of the effects of L1 prosody on perception of L2 prosodic boundaries.

INTRODUCTION

Speakers produce prosodic boundaries in the course of speech. Prosodic boundaries can be considered as the locations where speakers make a pause to breathe in or to come up with messages in long speeches. In Autosegmental-Metrical theory (Liberman, 1975; Pierrehumbert, 1980), there are two types of prosodic boundaries, Intermediate Phrases (ip) and Intonational Phrases (IP) in American English (English, henceforth), which may correspond to syntactic boundaries such as phrases and clauses, respectively. If a speaker pauses after the word "like" in the utterance, "I felt like I was stuck in a rut," the entire utterance constitutes one IP. Two smaller prosodic phrases, "I felt like" and "I was stuck in a rut," consist of two ips. Prosodic boundaries are associated with boundary tones, more specifically high (H) tones or low (L) tones: H- and L- for ip boundaries and H%, L%, HL%, and LH% for IP boundaries (Veilleux et al., 2006). These boundary tones convey the semantic or pragmatic meaning of an utterance in discourse context (e.g., H% for a question). Also, prosodic boundaries are characterized by acoustic cues (Beckman & Pierrehumbert, 1986; Shattuck-Hufnagel & Turk, 1998, among many others). Higher level prosodic boundaries (e.g., IP) tend to be correlated with greater pitch changes and final lengthening than lower level prosodic boundaries (e.g., ip) (Wang & Hirschberg, 1995). In comparison with English, Korean has more in terms of number and shape of prosodic boundaries. There are three prosodic domains, Accentual Phrases (ap), Intermediate Phrases, and Intonational Phrases, corresponding to words, phrases, and clauses, respectively (Jun, 2006). Korean has another prosodic domain (i.e., ap) smaller than ip, when compared with English. Also, Korean has richer inventories of IP boundary tones than English (H%, L%, HL%, LH%, HLH%, LHL%, HLHL%, LHLH%, and LHLHL%), which are mainly cued by F0 (Jun, 2000).

L1 prosody may influence the perception of prosodic boundaries in L2. According to the Markedness Theory (White, 1987), speakers with more marked phonological systems in their L1 than L2 will be more likely to acquire the L2 phonological system than speakers with less marked phonological systems in their L1 than L2. Considering its greater number and shape of prosodic

boundaries, the Korean prosodic system may be considered as more *marked* than the English one. Therefore, it is possible that Korean learners of English will be likely to perceive English prosodic boundaries similarly with native English speakers. On the other hand, the cue-weighting theory proposes that listeners weigh acoustic cues differently in their perceptions of speech. In addition, L2 listeners rely on the most familiar and useful acoustic cues from their prior experiences in L1 while they perceive new sounds in L2 (Ingvalson et al., 2012; Iverson et al., 2003, among many others). Korean learners of English may rely on F0 in their perceptions of English prosodic boundaries to a greater extent than can native English speakers because they tend to utilize F0 to make distinctions of IP boundary tones in their L1 Korean.

There are few empirical studies on how prosodic cues are adopted in speech segmentation by native English speakers and Korean native speakers (Kim et al., 2012; Kim & Cho, 2009; Tremblay et al., 2016). For instance, Kim et al. (2012) investigated how Korean native speakers and Dutch native speakers use F0 rise (LH%) and final lengthening in segmentation of an artificial language. Results showed that both groups of speakers relied on final lengthening, while only the Korean native speakers immediately used F0 rise in the final position of a word. This suggests that final lengthening can be considered as more universal than F0 in speech segmentation, since it was adopted by both groups of speakers at initial exposure. Tremblay et al. (2016) examined how English learners of French and Korean learners of French perceived word boundaries in French. It was assumed that French and Korean, but not English, have similar uses of F0 rise for ap boundaries. Korean learners of French were found to have more difficulties in using the F0 rise than English learners of French. This implies that the similarity between Korean and French hindered Korean learners of French from segmenting word boundaries in French appropriately. Although these previous studies provide insights on how native English speakers and native Korean speakers use prosodic cues in word segmentation, it is still unknown how Korean learners of English, compared with native English speakers, perceive prosodic boundaries larger than words in an utterance and also how they judge prosodic boundaries as a function of not only acoustic cues but also phonological features.

In the current study, we investigate how Korean learners of English judge English prosodic boundaries in a clear speech in relation to phonological features and acoustic cues, compared with native English speakers. For the phonological features, the types of prosodic boundaries (ip and IP) are annotated by linguistic experts using Tone and Break Indices (ToBI) annotation conventions (Veilleux et al., 2006). For the acoustic cues, F0 and duration are obtained for each word in the speech (see below for more details). We are interested in (1) how Korean learners of English differ from native English speakers in judging prosodic boundaries and (2) how these two groups of speakers rely on phonological and acoustic properties while making judgments of boundaries. The current study will expand our understanding of the perceptions of prosodic boundaries and cues by language learners whose L1 prosodic system is different from that of L2.

METHODS

Perception Experiments

Two groups of speakers participated in perception experiments. For the group of native English speakers, thirty-five undergraduate or graduate students (23 females and 12 males, mean age 24.3) were enrolled at a Midwestern university in the U.S. For the group of Korean learners of English, thirty undergraduate students majoring in English (26 females and 4 males, mean age 20.6) were recruited at a university in Korea. They were advanced learners of English with an average score of 820 out of 990 on the TEPS, the Test of English Proficiency developed by Seoul National University, which is convertible into TOEIC score 930 out of 990 and TOEFL score 109 out of 120 (https://www.teps.or.kr/InfoBoard/ConversionTable#).

Before the experiment started, the participants were introduced to a couple of examples in which speakers break up long stretches of speech into phrases to breathe in or to come up with messages. During the perception experiment, the participants were asked to hear the speech material twice and simultaneously mark words on the transcript if they perceived a disjuncture between the word and the following word. The speech material was a TED talk entitled "Try something new for thirty days" (https://www.ted.com/talks/matt_cutts_try_something_new_for_30_days). The TED talk contained 361 words (length 2' 25") and was delivered by a male speaker of American English in a clear and engaging style. The transcript was presented without capital letters and punctuation on a computer screen through an online interface, Language Markup and Experiment Design Software (Mahrt, 2013), as shown in Figure 1.

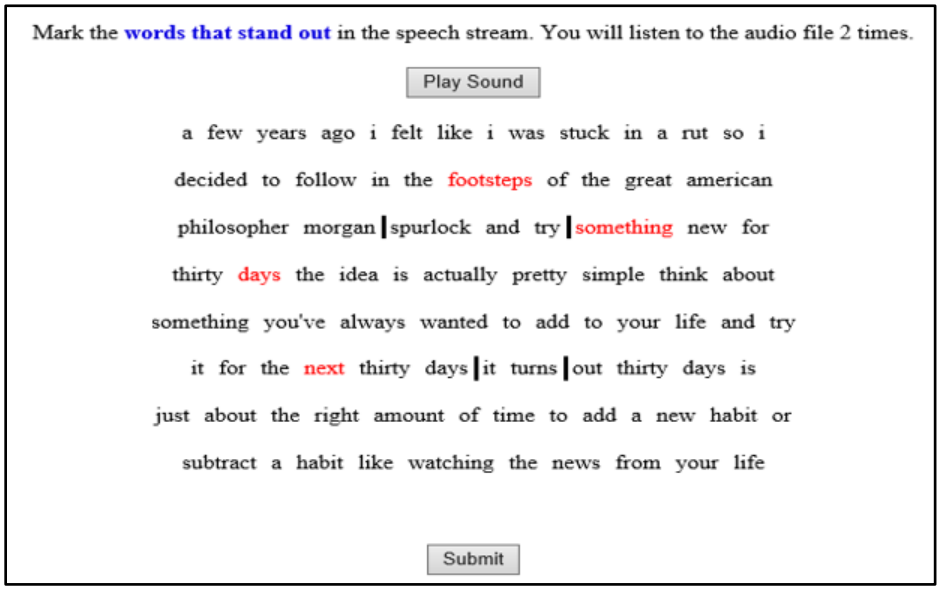

Figure 1. Screenshot of perception experiments. The black bars between words indicate the prosodic boundaries perceived by a participant. The red words are the ones perceived as prominent by the same participant, which are not analyzed in the current study.

Prosodic Cues

For the phonological features, boundary types were labeled by two experts using ToBI annotation conventions (Veilleux et al., 2006) in the same speech material as the one used for the perception experiments. Six boundary types were observed (H-, L-, H%, L%, HL%, and LH%) and were categorized into two types, intermediate phrases (H- and L-) and intonational phrases (H%, L%, HL%, and LH%). In addition to ip and IP boundaries, word boundaries were added to the analysis, which included all the words carrying neither ip nor IP boundaries in the TED talk.

For the acoustic cues, max F0 (Hz) and mean phone duration (ms) were measured for each word in the entire speech using ProsodyPro (Xu, 2013). The mean phone duration was obtained by dividing the entire duration of a word by the number of segments of the word. The acoustic cues were z-normalized using the means and standard deviations from all the words in the TED talk.

Analysis

Using a generalized linear mixed-effects model (GLMM), a participant's binary boundary ratings (1 for the words with a perceived boundary, and 0 for the words without a perceived boundary) were modeled as a function of the participant's L1 group (native English speakers and Korean learners of English), boundary types (word, ip, and IP), acoustic cues (max F0 and mean phone duration), interaction between L1 groups and boundary types, and interaction between L1 groups and acoustic cues. The interaction between boundary types and acoustic cues was not included due to a convergence issue. Random intercepts for participants were considered in the model. The model was run in R (R Core Team, 2019) using lme4 (Bates et al., 2015).

Also, we obtained boundary (b-) scores (Cole et al., 2014), which represent the percentage of participants' boundary markings for a word. The b-scores range from 0 to 1, where 1 indicates that 100% of the participants have marked a prosodic boundary on a word, and 0 means that 0% of the participants have done so.

RESULTS

Figure 2 shows boundary ratings by native English speakers (in the solid line) and Korean learners of English (in the dotted line) for one of the sentences analyzed in the current study, "It turns out all you have to do is write sixteen hundred sixty-seven words a day for a month." Both groups of speakers tended to mark boundaries on the words "out," "do," "day," and "month." This tendency was much greater for Korean learners of English (average b-score of the four words 0.97) than for native English speakers (average b-score 0.62). Put differently, Korean learners of English were more likely to perceive boundaries on those words in the utterance than were native English speakers.

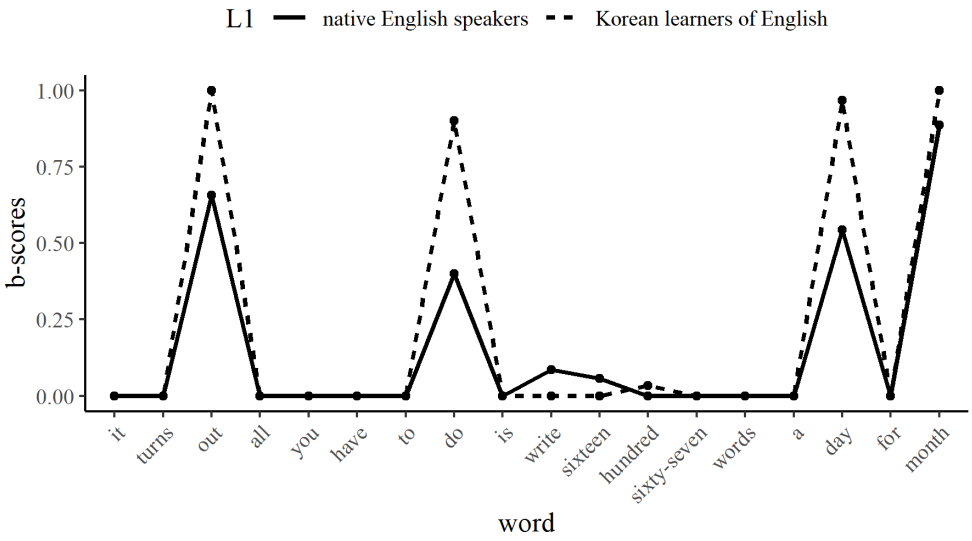

Figure 2. Boundary ratings by native English speakers (in the solid line) and Korean learners of English (in the dotted line) for a sentence.

In Table 1, the GLMM results show that the L1 groups (native English speakers and Korean learners of English) did not significantly differ from each other in perceiving prosodic boundaries. However, boundary types (ip and IP), as well as acoustic cues (max F0 and mean phone duration), were found to be significantly associated with participants' boundary ratings. These results suggest that the two groups of speakers tended to rate boundaries similarly by relying on phonological features (i.e., boundary types) and acoustic cues. Nevertheless, the GLMM results further revealed that the interaction between L1 group and IP, but not the interaction between L1 group and ip, influenced boundary ratings. In other words, Korean learners of English differed from native English speakers in the perception of IP boundaries, but not ip boundaries. Also, the interaction between L1 group and max F0, but not the interaction between L1 group and mean phone duration, was significantly associated with boundary markings. Put differently, Korean learners of English differed from native English speakers in relying on max F0, but not on mean phone duration, in their boundary judgments.

Table 1
GLMM Results on Estimated Perceived Boundaries as a Function of L1 Group, Boundary Types, Acoustic Cues, and their Interactions

Variable	est.	*SE*	*z*	*P*
(Intercept)	-5.03	0.17	-29.38	< 0.01**
Korean learners of English	0.11	0.25	0.44	0.66
ip	4.59	0.14	32.44	< 0.01**
IP	6.75	0.15	44.86	< 0.01**
F0	0.18	0.04	4.43	< 0.01**
Duration	0.11	0.04	2.98	< 0.01**
Korean learners of English: ip	0.11	0.21	0.52	0.60
Korean learners of English: IP	1.06	0.24	4.48	< 0.01**
Korean learners of English: F0	0.17	0.07	2.66	< 0.01**
Korean learners of English: duration	0.04	0.06	0.66	0.51

Figures 3-4 are visualizations of the GLMM results in Table 1. Figure 3 shows the estimated effects of boundary types on perceived boundaries by native English speakers (left panel) and Korean learners of English (right panel). The x-axis indicates boundary types. The y-axis shows estimated effects of the boundary types on perceived boundaries. The log-odds of the GLMM outcome were converted into the probability scale ranging from 0 to 1. The overall pattern is similar between the two groups: they were more likely to mark boundaries on IP than ip. This tendency, however, was significantly greater for Korean learners of English than for native English speakers.

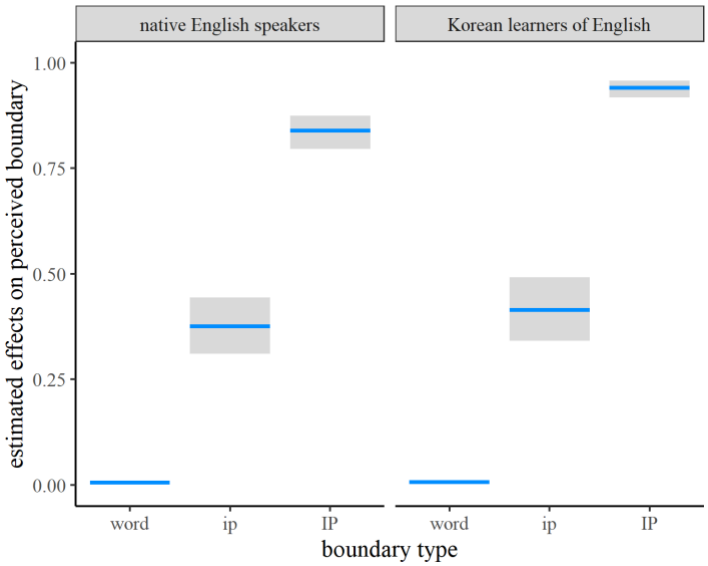

Figure 3. Estimated effects of boundary types on perceived boundaries by native English speakers (left panel) and Korean learners of English (right panel).

Figure 4 shows the estimated influence of max F0 on perceived boundaries by native English speakers and Korean learners of English. The x-axis indicates z-normalized values of max F0. The y-axis shows estimated effects of max F0 on perceived boundaries in the probability scale. For both groups of speakers, higher max F0 was associated with greater likelihood of perceived boundaries. Nevertheless, this tendency was found to be significantly greater for Korean learners of English than for native English speakers.

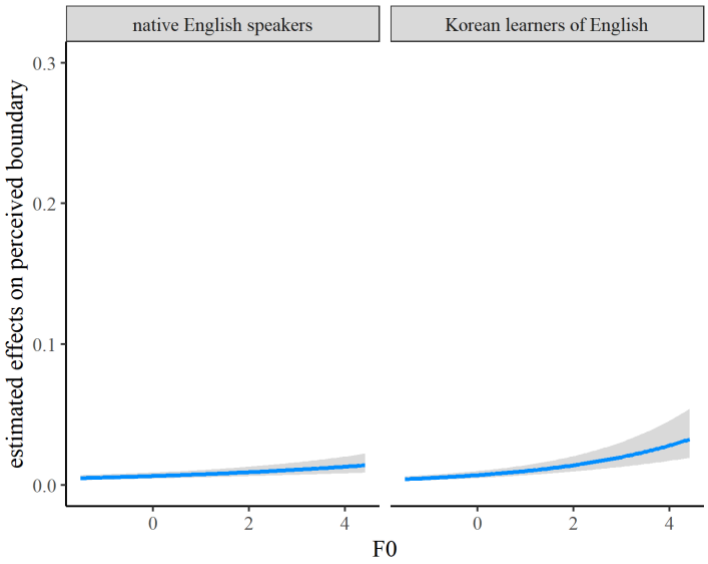

Figure 4. Estimated influence of F0 on perceived boundaries by native English speakers and Korean learners of English.

It was suspected that the two groups of speakers might differ in relying on max F0 in their perception of boundaries because the speaker of this TED talk utilized max F0 differently to encode boundary types. Therefore, we further examined the relationship between phonetic cues and boundary types in Figures 5-6. The x-axis indicates boundary types in Figures 5-6. The y-axis shows z-normalized max F0 in Figure 5 and z-normalized mean phone duration in Figure 6. In Figure 5, the range of max F0 was greater for IP than for ip. The speaker in this speech was more likely to lower his pitch for IPs than for ips, although he tended to raise his pitch similarly for both boundaries. In Figure 6, the range of mean phone duration was similar for both boundaries. Overall, the results suggest the phonetic cues are weakly associated with boundary types in this TED talk.

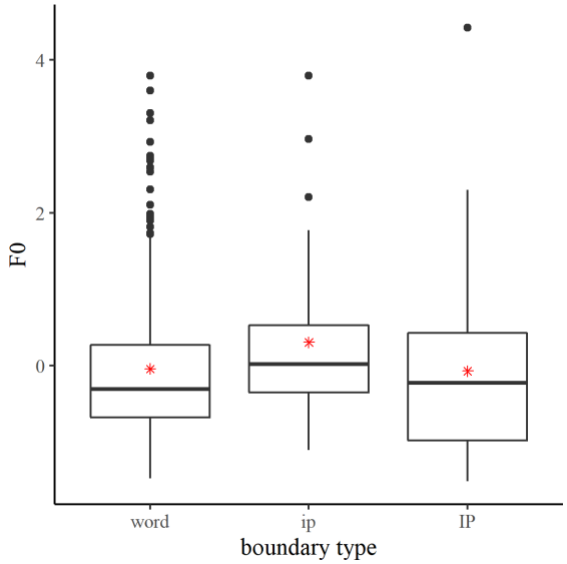

Figure 5. Relation between boundary types and z-normalized max F0.

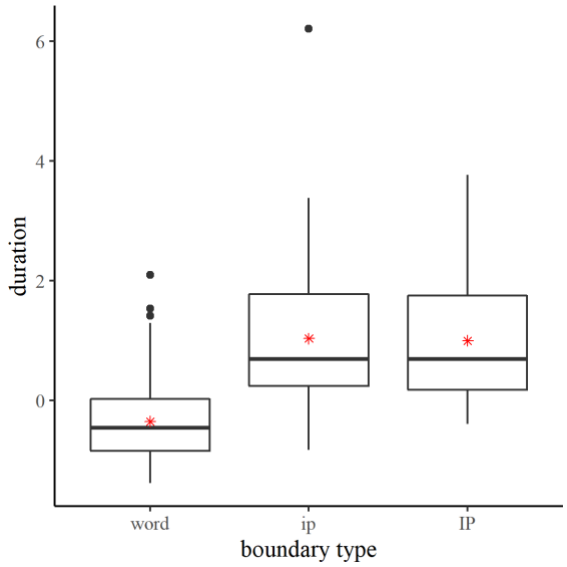

Figure 6. Relation between boundary types and z-normalized mean phone duration.

DISCUSSION AND CONCLUSION

In the current study, Korean learners of English tended to perceive prosodic boundaries similarly with native English speakers, but the two groups of speakers differed in that Korean learners of English were more likely to mark IP boundaries than were native English speakers. The overall similarity of boundary ratings between the two groups of speakers can be considered as supporting

evidence of the Markedness Theory. As discussed, Korean has a more marked prosodic system—a greater number of boundary types and tones—than English. Therefore, Korean learners of English were expected to have few difficulties in perceiving prosodic boundaries in English. Indeed, Korean learners of English showed high agreement (high b-scores) on perceived boundaries, especially IP boundaries, and their judgments were comparable with those of native English speakers as well as the linguistic experts who labeled the boundary types in the current study. Considering such high agreement, Korean learners of English can be regarded as super-annotators (Cole et al., 2017), even if they are non-native speakers of English as well as linguistically non-experts. The differences in IP boundary markings between the two groups of speakers may arise from the more complex distinctions of IP boundary tones in Korean than in English. Due to their L1 Korean influence, Korean learners of English might be more attuned to English IP boundaries than are native English speakers.

Also, both Korean learners of English and native English speakers were found to employ max F0 and mean phone duration in their boundary ratings, but they differed in that Korean learners of English were more likely to rely on max F0 than were English native speakers. This can be rather surprising considering the results from our qualitative examination in Figure 5 that the speaker in this TED talk was likely to use a similarly high pitch for both ip and IP boundaries. He utilized a wider F0 range for IP boundaries than for ip boundaries, but this was achieved by lowering F0 for IP boundaries more than for ip boundaries. In other words, phonetic cues were weakly differentiated between boundary types in this TED talk. Despite the weak association between the acoustic cues and phonological features, Korean learners of English still favored max F0 to a greater extent than native English speakers. More specifically, Korean learners of English were more likely to rate prosodic boundaries than native English speakers if the max F0 of a word increased. This again might lead us to consider the rich IP tonal inventories in Korean, which are mostly signaled by F0. It is possible that Korean learners of English relied on max F0 in their perception of English prosodic boundaries because F0 is an efficient cue in Korean prosody. This is in line with the previous findings on word boundaries and acoustic cues in cue-weighting theory (Tremblay et al., 2016; Tremblay et al., 2018).

In the current study, two signal-driven factors, phonological features and acoustic cues, were examined in relation to the perception of prosody. We learned from our GLMM analysis that both phonological features and phonetic cues were important factors in the perception of prosodic boundaries. However, considering higher estimated effects in the GLMM, phonological features were considered as stronger predictors of perceived boundaries than were phonetic cues in this speech. We suspect that the weak contribution of acoustic cues to perceived boundaries can be due to the speech style of the speaker in this TED talk. As discussed, the speaker in this speech employed pitch and duration in a similar manner for all the phonological features (boundary types), although his low pitch was differentiated between ip and IP boundaries. As F0 and duration were not such informative cues in this speech, listeners might not rely on the acoustic cues as much as the phonological features in their boundary judgments.

Overall, we examined how native English speakers and Korean learners of English perceived English prosodic boundaries as a function of phonological features and acoustic cues. To the best of our knowledge, this is the first study to address the perception of prosodic boundaries beyond word boundaries (ip and IP) by Korean learners of English compared with native English speakers.

Also, this study considers not only acoustic cues but also phonological features in the perception of boundaries. Our findings suggest that language learners with more marked prosodic systems in their L1 than L2 are able to perceive L2 prosodic boundaries similarly with native L2 speakers. And, they favor the acoustic cues that are considered efficient cues in their L1 in the perception of L2 prosodic boundaries. In sum, the current study has implication for the effects of (1) phonological features and phonetic cues and (2) transfer from L1 prosody to L2 prosody on the perception of prosodic boundaries.

ABOUT THE AUTHORS

Suyeon Im is a postdoctoral research coordinator in Hanyang Institute for Phonetics and Cognitive Sciences of Language at Hanyang University in Seoul in Korea. She obtained her PhD in Linguistics from the University of Illinois at Urbana-Champaign. Her research uses experimental and computational methods to understand sound structure in L1 and L2.

Hanyang Institute for Phonetics and Cognitive Sciences of Language
Hanyang University
222 Wangsimni-ro, Seongdong-gu
Seoul 04763
Republic of Korea
E-mail: suyeon.im@hanyang.ac.kr

Hyunkee Ahn is full professor in the Department of English Language Education at Seoul National University in Seoul, Republic of Korea. He obtained a PhD in Linguistics from the University of Texas at Austin. His research interests include perception and production of English as a foreign language, L2 phonemic awareness, and L2 pronunciation instruction and learning.

Department of English Language Education
Seoul National University
1 Gwanak-ro Gwanak-gu
Seoul 08826
Republic of Korea
E-mail: ahnhk@snu.ac.kr

REFERENCES

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3*, 255-309.

Cole, J., Mahrt, T., & Hualde, J. I. (2014, January). Listening for sound, listening for meaning: Task effects on prosodic transcription. In *Proceedings of Speech Prosody* (vol. 7, pp. 859-863).

Cole, J., Mahrt, T., & Roy, J. (2017). Crowd-sourcing prosodic annotation. *Computer Speech and Language, 45*, 300-325.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate/r–l/on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*(2), 255-274.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-B57.

Jun, S. (2000). K-ToBI (Korean ToBI) labelling conventions. Retrieved from http://www.linguistics.ucla.edu/people/jun/ktobi/k-tobi.html.

Jun, S. (2006). Intonational phonology of Seoul Korean revisited. *Japanese-Korean Linguistics, 14*, 14-25.

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition, 34*(3), 415-444.

Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *The Journal of the Acoustical Society of America, 125*(5), 3373-3386.

Liberman, M. Y. (1975). *The intonational system of English* (Doctoral dissertation). Massachusetts Institute of Technology.

Mahrt, T. (2013). *Language markup and experimental design software* [Computer software]. Retrieved from http://www.timmahrt.com/lmeds.html.

Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation* (Unpublished doctoral dissertation). Boston, MA: Massachusetts Institute of Technology.

R Core Team (2019). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org.

Shattuck-Hufnagel, S., & Turk, A. (1998, June). The domain of final-lengthening in English. In *Proceedings of 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America* (vol. 2, pp. 1235–1236).

Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the native language on the learning of fundamental frequency in second-language speech segmentation. *Frontiers in Psychology, 7*, 985.

Tremblay, A., Broersma, M., & Coughlin, C. E. (2018). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition, 21*(3), 640-652.

Veilleux, N., Shattuck-Hufnagel, S., & Brugos, A. (2006). *Transcribing prosodic structure of spoken utterances with ToBI* [PowerPoint slides]. Retrieved from https://ocw.mit.edu.

Wang, M. Q., & Hirschberg, J. (1992). Automatic classification of intonational phrase boundaries. *Computer Speech & Language, 6*(2), 175-196.

White, L. (1987). Markedness and second language acquisition: The question of transfer. *Studies in Second Language Acquisition, 9*(3), 261-285.

Xu, Y. (2013, August). ProsodyPro—A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (pp. 7-10).