English Lexical Stress and Spoken Word Recognition in Korean Learners of English

Jeonghwa Shin and Shari R. Speer

Department of Linguistics, Ohio State University, USA
jshin@ling.osu.edu, speer@ling.osu.edu

Abstract

Two experiments explore how Korean-speaking L2 learners of English process English lexical stress during spoken word recognition. Korean doesn't employ lexical-level prosodic distinctions like English lexical stress, but it has phrase-level prosodic structure ((T) HLH), with the initial tone determined by the phonation type of phrase-initial sound. Results from eye-tracking and gating experiments showed slower processing times and lower accuracy for Korean learners' processing of English lexical stress information during word recognition compared to native English speakers. Processing difficulty was greater for iambic than for trochaic words, and time-course data indicate that processing difficulty increased between initial contact with the target word and the completion of word recognition. Post-hoc analyses of the eye tracking data also revealed that Korean L2 learners' word recognition was facilitated when the stress pattern matched the Korean tonal pattern induced by the word-initial phonation type contrast. Taken results together, the study implies that Korean L2 learners' processing of English lexical stress is modulated by both the processing strategy and constraints from perceptual experience with L1 phrase-level prosody.

Index Terms: lexical stress, phrase-level prosody, L1 transfer

1. Introduction

1.1 Background

Prosody at the word level is primarily characterized by three acoustic dimensions of the suprasegmental parameters: durational patterning, fundamental frequency (F0), and signal amplitude. These primary acoustic dimensions combine with other subtle variations, corresponding to the perception of timing, pitch, and loudness. Many previous studies have shown that word recognition is facilitated when words are produced with appropriate prosodic information [2, 4]. This was especially evident in languages that are known to employ word-level prosody such as English, Dutch, Spanish (lexical stress), Japanese (lexical pitch accent), and Mandarin Chinese (lexical tone). Listeners' perceptual sensitivity to word prosody in these languages is part of their native language endowment, developed over time due to L1 exposure since infancy.

While it is clear that prosody plays an important role in L1 spoken word recognition, relatively little is known about how L2 learners process L2 word prosody during spoken word recognition. With regard to L2 speech processing, it's more likely that the perceived similarities or dissimilarities between phonetic properties of the nonnative contrasts and those of native phonology constrain L2 learners' perception and processing of L2 speech. The scope of L1 influence can encompass the entire native language system, from the phonemes and their variants, stress and rhythmic patterns, to intonational patterns and their interaction with other phonemes.

In this light, the present study explores how language-specific variation of prosodic structure affects L2 learners' processing of L2 prosodic contrasts in the context of spoken word recognition. Particularly, we examine L2 learners' perception and processing of L2 word prosody when comparable distinctions are not present in the L1. Such is the case for Korean speaking learners of English whose L1 does not have word prosody such as English lexical stress, but has a phrase-level prosodic structure with some tonal variation.

1.2 Intonational structure in English and Korean

1.2.1 English lexical stress and intonation structure

English has meaningful differences in stress at the level of individual words ("lexical stress", e.g., the language allows for minimal pairs such as *INsight* vs. *inCITE;* capitals indicate stressed syllables). At the phrasal level, English also includes "pitch accents", (e.g., *I want CHOcolate pie* (not vanilla) vs. *I want chocolate PIE* (not cake); capitals indicate pitch accented stressed syllables) [1]. Manipulation of duration, F0, and amplitude produces perceived prominence differences across syllables in a word. These can be categorized either as trochaic (SW: Strong associated with the syllable carrying a primary stress, Weak associated with the syllable carrying a secondary or no stress) or iambic (WS) stress pattern at the lexical level. In addition to suprasegmental variation, English unstressed syllables are often accompanied by a reduced vowel such as a schwa (*OBject* vs. *obJECT*). In utterances of English, though every word contributes its own lexical stress pattern, the realization of a stressed syllable as the most prominent syllable in an utterance is determined by pitch accent location and type (e.g. H*, L*, L+H*, L*+H, H+!H*), assigned at a phrase level.

1.2.2 Korean intonation structure

Unlike English, the Korean language doesn't have word level prosody. Korean intonational phonology posits two hierarchical levels of prosodic phrases above phonological words, that is, accentual phrase (AP) and Intonation phrase (IP) [3]. The AP is defined by a fixed prosodic pattern, (T) HLH, which contains one or more lexical items, but most frequently one, and demarcates the beginning and the end of a phrase. The initial tone is realized as a high tone (H) with a relatively high F0 if the AP-initial obstruent is aspirated or tense and as a low tone (L) otherwise.

1.3 The current study

The study specifically explores 1) whether Korean learners of English use English lexical stress information for spoken word recognition in L2 and 2) whether Korean phrase-level prosodic structure exerts any influence on Korean L2 learners' processing of English lexical stress during word recognition. We explore the issues in an eye tracking experiment and a gating experiment.

If Korean learners of English are not sensitive to the relevant acoustic phonetic cues to English lexical stress due to the lack of experience with word prosody in L1, their word recognition is less likely to be facilitated by lexical stress information as compared to native English speakers. Meanwhile, if the influence of L1 Korean phrase-level prosody manifests in the perception and processing of English lexical stress, the phonation type of the word initial segment may influence the processing of English lexical stress. That is, word that begins with an aspirated consonant, which would have a H initial tone in Korean, serves as a better match for a H* pitch accented trochaic English word, as compared to a word that begins with a lax-initial consonant, which would have a L initial tone in Korean.

2. Eye tracking experiment

An eye tracking experiment examined how native English speakers and Korean L2 learners of English used lexical stress information over the time course of spoken word recognition.

2.1. Materials

Sixteen trisyllabic nonword near stress minimal pairs (e.g., *ZAkunaɪ - zaKUner*) were constructed for the testing stimulus - the first two syllables differed only in stress, either trochaic (SW) or iambic (WS), and the last syllable was segmentally disambiguating. The test nonwords did not include reduced vowels in unstressed syllables. A trained female native English speaker produced the nonwords in citation form and in a carrier sentence, "*Click on the (target word) now*", with the same pitch accent on the lexically stressed syllable. Visual stimulus materials were line-drawings of space aliens.

2.2. Participants and procedure

Twenty-three native English speakers and nineteen Seoul Korean speaking learners of English participated in a three-day training session, an eye tracking word recognition task, and an orthographic transcription task.

The training session included look-and-listen tasks, picture naming tasks, and picture-choice tasks. In the training session, participants learned the associations of nonword names and alien picture referents.

In the following word recognition test, participants were asked to mouse-click on one of the three alien pictures on the computer screen upon hearing the English instruction, "*Click on the (target word) now*". The three pictures were the target with either trochaic or iambic stress (e.g., *ZAkunaɪ*), a stress competitor name (e.g., *zaKUner*), and a distractor (e.g., *FUgiser*). The *ASL E6000* head-mounted eye tracker recorded gaze position at a sampling rate of 60 Hz.

After completing the eye tracking word recognition test, participants transcribed the target words in their L1 orthography.

2.3. Analysis on the eye gaze data

The data came from the forty two participants. Their accuracy rates were above 90% in the picture-choice task of the last training session and above 75% in the eye tracking word recognition task.

Analyses were conducted on looks to the target in correct responses using weighted empirical-logit regression in a linear mixed effect model with random effects for subjects and items. The three categorical predictors, *L1* (English vs. Korean), *Stress pattern* (Trochaic vs. Iambic), and *Phonation type* (Aspirated vs. Lax) were sum coded, and *Time* was treated as a continuous variable. The span of the analysis window was from -153 to 867ms, with data aligned at the boundary between the first and second syllable of the target word.

2.4. Results and discussion

The mean empirical log ratios of gaze observations to the target object for the two stress conditions are presented in Figure 1. The solid vertical line indicates the first syllable offset. Dotted vertical lines indicate mean durations of the first two syllables - 238 ms and 164 ms for trochaic words and 180 ms and 230 ms for iambic words.

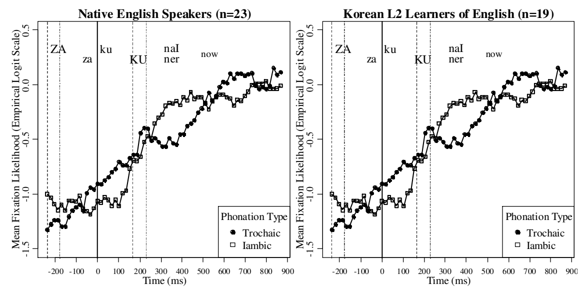

Figure 1: *Mean fixations to the target object, empirical log odds scale.*

Figure 1 shows that the slopes for the two stress conditions were steeper for English than Korean speakers, with more fluctuations in the slope change for the L2 learners. English speakers showed more looks to trochaic words than iambic words immediately after the first syllable. Korean speakers showed steeper slope in the trochaic condition than in the iambic condition up to 450 ms, the average end of the last syllable in the nonword.

Table 1 summarizes the statistical analyses. For the purpose of the experiment, we focus on the effects of the categorical predictors on the slope.

Table 1: *Summary of L1, stress pattern, phonation type, and time effects on gaze (N = 11040; logLik = -17634).*

Predictor	Est.	*SE*	*t*	*p* <
Intercept	-0.90	0.13	-7.13	.001
Lg1 = *Korean*	0.05	0.20	0.25	n.s.
Stress1 = *Iambic*	-0.21	0.10	-2.16	0.05
Phonation1 = *Lax Initial*	0.05	0.19	0.28	n.s.
Lg1:Stress1	-0.06	0.20	-0.29	n.s.
Lg1:Phonation1	-0.65	0.20	-3.24	.001
Stress1:Phonation1	-0.70	0.20	-3.52	.001
Lg1:Stress1:Phonation	0.64	0.40	1.60	n.s.
Time	0.95	0.09	10.66	.001
Lg1:Time	-0.44	0.17	-2.45	.05
Stress1:Time	0.19	0.18	1.04	n.s.
Phonation1:Time	-0.25	0.18	-1.38	n.s.
Lg1:Stress1:Time	-0.34	0.33	-0.95	n.s.
Lg1:Phonation1:Time	0.60	0.36	1.67	n.s.
Stress1:Phonation1:Time	1.37	0.36	3.84	.001
Lg1:Stress1:Phonation1:Time	-1.85	0.71	-2.59	.05

There was a significant effect of *L1* on the slope (*ps* < .05). The negative parameter estimate for *L1* effect on the slope (β = - 0.44) indicates that the increase of fixation likelihood over time was significantly higher for English speakers than for Korean speakers, supporting the prediction. The effect of *Stress type* on the slope was not significant (*p* > .05), suggesting that activation of the trochaic and iambic words was overall similar over the time course of word recognition.

However, in a post-hoc analysis conducted on successive 200 ms windows after the first syllable offset using the same mixed effect model, the interaction of *L1* and *Lexical stress* was significant in windows from 200 - 400 ms and 400 - 600 ms (*ps* < .05). This indicates that trochaic word recognition was temporarily more facilitated than iambic word recognition even after the second syllable in Korean L2 learners than it was in English speakers.

To examine whether the lexical stress processing was affected by what Koreans heard as aspirated vs. lax initial contrast, we refer to interaction of *L1, Stress pattern,* and *Phonation type* with *Time*. There was a significant interaction of *Stress pattern* and *Phonation type* on the slope with a positive parameter estimate (β = 1.37, $p < 0.001$), suggesting that the slope change for the trochaic word was significantly higher for the aspirated initial word than for the lax initial word, and it was the opposite for the iambic word. However, the significant interaction of *L1, Stress pattern,* and *Phonation type* on the slope (β = -1.85, $p < 0.05$) indicates the interaction of *Phonation type* and *Stress pattern* over time was stronger for native English speakers than for L2 learners of English.

In the post-hoc segmented analysis, the interaction of *Stress pattern* and *Phonation type* turned out to be significant in the last window, 800 - 867 ms ($p < .05$), with no interaction with *L1* ($p > .05$). This implies that the aspirated and lax initial segments respectively facilitated the processing of English trochaic and iambic words toward the end of word recognition (when disambiguating final syllable information was also available) in both English and Korean speakers. The results support the prediction that the experience with acoustic-phonetic cue to L1 phrase-level tonal variation (i.e., F0) affects the processing of L2 lexical stress at some point during spoken word recognition.

3. Cross-modal gating experiment

A cross-modal gating experiment further investigated the effect of phonation type contrast on the perception of stress pattern for word identification. Gating methodology was chosen as it measures successive responses as the spoken word unfolds, allowing examination of the processing of the initial syllable and subsequent syllables separately.

3.1. Materials

Sixteen near minimal stress pairs were constructed and recorded as for the eye tracking experiment. Word-initial phonation type was counterbalanced so that 8 pairs began with obstruent Seoul Korean speakers heard as aspirated (/p, t, k, s/) and the other 8 pairs began with those they heard as lax (/b, d, g, m/) in a pre-listening test.

3.2. Participants and procedure

Thirty three native English speakers and thirty two Korean speakers participated in training tasks modeled after those used in the eye tracking experiments, a gating experiment, and an orthographic transcription task.

In the cross-modal gating experiment, the participants chose a target picture from three picture referents upon hearing a nonword fragment in the carrier instruction (e.g., '*Click on the* (*target fragment*)'). Each task consisted of three gating blocks - the first block (short fragment) consisted of 16 word-initial syllable fragments (e.g., '*Click on the za-*') and the second block (long fragment) gave 16 two-syllable fragments (e.g., '*Click on the zaKU-*'). The third block (whole fragment) presented nonwords in a full form.

3.3. Analysis

The analyses included only data from the sixty five participants who performed with at least 90% accuracy in the picture-choice task of the last training session and above 75% accuracy rate in the whole fragment block in the gating word identification task.

A linear mixed effect model was separately performed on accuracy and RTs for the correct responses. The three categorical predictors, *L1* (English vs. Korean), *Stress pattern* (Trochaic vs. Iambic), and *Phonation type* (Aspirated vs. Lax) were sum coded. The *Gating block* was treated as a continuous variable with the shortest fragment block as the referent point.

3.4. Results and discussion

Table 2 shows the mean percent accuracy for each gating block.

Table 2: *Mean accuracy for L1, stress pattern, phonation type for each gating block (standard errors are presented in parentheses).*

Block	Stress Pattern	English speaker		Korean speaker	
		Asp	Lax	Asp	Lax
Short Frag.	Trochaic	75.76 (3.74)	68.94 (4.04)	66.41 (4.19)	56.25 (4.40)
	Iambic	57.58 (4.32)	52.27 (4.36)	53.91 (4.42)	56.25 (4.40)
Long Frag.	Trochaic	76.52 (3.70)	78.03 (3.62)	74.22 (3.88)	67.19 (4.17)
	Iambic	78.79 (3.57)	71.97 (3.92)	72.66 (3.96)	73.44 (3.92)
Whole Frag.	Trochaic	84.85 (3.13)	85.61 (3.07)	83.59 (3.29)	84.38 (3.22)
	Iambic	87.88 (2.85)	83.33 (3.26)	89.84 (2.68)	87.50 (2.93)

Table 3 summarizes the analysis result on accuracy. As we are primarily concerned with processing of the initial stress information given the word-initial phonation type, main effect(s) and interactions of the predictors at the intercept are discussed here.

Table 3: *Summary of L1, stress pattern, phonation type, and gating block effects on accuracy (N = 3120; logLik = -1667).*

Predictor	**Est.**	***SE***	***z value***	***p <***
(Intercept)	0.47	0.11	0.43	.001
Lg1 = *Korean*	-0.3	0.16	-1.8	n.s.
Stress1 = *Iambic*	-0.5	0.12	-4.1	.001
Phonation1 = *Lax initial*	-0.23	0.19	-1.21	n.s.
Lg1:Stress1	0.47	0.25	1.91	n.s.
Lg1:Phonation1	0.08	0.25	0.32	n.s.
Stress1:Phonation1	0.34	0.25	1.37	n.s.
Lg1:Stress1:Phonation	0.61	0.49	1.24	n.s.
Gate	0.7	0.06	12.83	.001
Lg1:Gate	0.15	0.11	1.40	n.s.
Stress1:Gate	0.4	0.11	3.64	.001
Phonation1:Gate	0.06	0.11	0.51	n.s.
Lg1:Stress1: Gate	-0.11	0.22	-0.50	n.s.
Lg1:Phonation1: Gate	-0.02	0.22	-0.10	n.s.
Stress1:Phonation1:Gate	-0.35	0.22	-1.61	n.s.
Lg1:Stress1:Phonation1: Gate	-0.08	0.44	-0.19	n.s.

At the intercept, the effect of *Stress pattern* was significant ($p < .001$), with no other effects and interactions approaching significance ($p > .05$). The negative estimate for the effect of

Stress pattern (β = - 0.5) indicates that there was overall higher accuracy for trochaic word identification than for iambic word identification regardless of the phonation type of the word initial consonant (72.35 % vs. 54.92 % for English speakers, 61.33 % vs. 55.08 % for Korean speakers). Though Table 2 shows that Korean L2 learners were numerically more accurate for aspirated initial trochaic word identification than for lax initial trochaic word identification and it was the opposite for iambic word identification, the difference between lax and aspirated initial iambic words was too small to produce any significant interaction of *L1, Stress pattern,* and *Phonation type*.

Table 4 shows the mean RTs of correct responses in each gating block (4.9 % outlier RTs were removed).

Table 4: *Mean RTs for L1, stress pattern, phonation type for each gating block (standard errors are presented in parentheses).*

Block	Stress Pattern	English speaker		Korean speaker	
		Asp	Lax	Asp	Lax
Short Frag.	Trochaic	2001.10 (155.33)	2279.67 (160.22)	3263.06 (371.59)	3564.45 (440.56)
	Iambic	2286.01 (173.17)	2383.78 (179.08)	4146.09 (478.40)	4819.84 (504.38)
Long Frag.	Trochaic	1805.79 (110.15)	1352.23 (95.11)	2042.44 (182.62)	1640.08 (217.20)
	Iambic	1468.44 (98.72)	2390.64 (107.72)	1834.86 (146.95)	2417.72 (210.44)
Whole Frag.	Trochaic	1187.59 (87.44)	1210.21 (88.75)	1390.29 (104.84)	1562.10 (123.36)
	Iambic	1281.93 (77.58)	1312.24 (84.48)	1753.70 (118.56)	1697.99 (133.38)

Table 5 summarizes the analysis result on the RTs.

Table 5: *Summary of L1, stress pattern, phonation type, and gating block effects on RTs (N = 2186; logLik = -19339).*

Predictor	**Est.**	***SE***	***t***	***p <***
(Intercept)	2947.69	126.58	23.29	.001
Lg1 = *Korean*	1544.40	200.23	7.71	.001
Stress1 = *Iambic*	471.49	126.74	3.72	.001
Phonation1 = *Lax initial*	331.47	199.53	1.66	. n.s.
Lg1:Stress1	604.05	252.42	2.40	.05
Lg1:Phonation1	503.10	251.57	2.00	.05
Stress1:Phonation1	219.31	253.18	0.87	n.s.
Lg1:Stress1:Phonation	218.63	504.25	0.43	n.s.
Gate	-810.50	45.66	17.75	.001
Lg1:Gate	-636.92	91.32	-6.98	.001
Stress1:Gate	-211.36	91.52	-2.31	.05
Phonation1:Gate	-147.61	91.25	-1.62	n.s.
Lg1:Stress1: Gate	-296.70	182.78	-1.62	n.s.
Lg1:Phonation1: Gate	-175.94	182.50	-0.96	n.s.
Stress1:Phonation1:Gate	-142.28	182.82	-0.78	n.s.
Lg1:Stress1:Phonation1: Gate	-332.76	365.13	-0.91	n.s.

At the intercept, effects of *L1* and *Stress pattern* were significant (*ps* < .05), suggesting that Korean speakers took longer to respond than did English speakers (β = 1544.40), and that it took more time for iambic word identification than for trochaic word identification when only the initial syllable of the word was used for identification (β = 471.49).

There was a significant interaction of *L1* and *Stress pattern* (β = 604.05, p < .05) with no interaction with *Phonation type* (*p* > .05). This suggests that Korean L2 learners had more processing difficulty for iambic word identification as compared to trochaic word recognition than did English speakers, regardless of with what phonation type the nonword began.

4. Conclusions

The current study investigated how L1 experience with phrase-level prosodic structure influences the process of L2 lexical stress during spoken word recognition. The results from the eye tracking experiment and the gating experiment converge to show that Korean-speaking learners of English show difficulty integrating lexical stress information for word recognition as compared to native English speakers. In addition, Korean speakers showed a temporary processing advantage for trochaic over iambic word recognition at the end of critical word presentation, as compared to English speakers. This is attributable to Korean listeners' experience with L1that provides less incentive to pay attention to the prominence difference in syllables for spoken word recognition. In addition, as Korean phrase-level prosodic structure does not involve the manipulation of duration and amplitude cues for the initial tonal contrast, the information in the English unstressed initial syllable, which entailed reduction of all acoustic cues, would be psychoacoustically difficult to detect and thus yielded more processing difficulty to Korean L2 learners.

However, the results from the post-hoc eye tracking data analyses provided some evidence that word-initial phonation type can potentially modulate the processing of English lexical stress during word recognition in Korean L2 learners - recognition of aspirated-initial trochaic and lax-initial iambic words was facilitated toward the end of word recognition. As the interaction of word-initial phonation type and lexical stress was not found in the gating experiment, it would be impetuous to make any firm conclusion about the Korean L1 phrase-level prosodic effect on the processing of L2 lexical stress.

However, the different results are more likely due to the different experimental paradigms employed in the two experiments - unlike the eye tracking word recognition task, the gating word identification task allowed listeners extended time to make their decisions, possibly obscuring the immediate reflection of the processing of phonation and lexical stress in the response. Therefore, we leave the finding as the potential evidence that L1 prosody employed at a different level from L2 prosody can affect L2 prosodic processing at some point during word recognition, if they share the same acoustic features (e.g., F0) to make relevant prosodic manipulations.

5. Acknowledgements

This research is supported by research fund from Linguistics department at Ohio State University and by the Alumni Grants for Graduate Research and Scholarship (AGGRS) from Graduate College of Arts and Humanities at Ohio State University.

6. References

[1] Beckman, M. E., "The parsing of prosody", Language and Cognitive Processes, 11(1-2), 17-68, 1996.

[2] Cooper, N., Cutler, A. and Wales, R., "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners", Language and speech, 45(3), 207-228, 2002.

[3] Jun, S., "The phonetics and phonology of Korean prosody", Unpublished doctoral dissertation, The Ohio State University, 1993.

[4] Soto-Faraco, S., Sebastián-Gallés, N. and Cutler, A., "Segmental and suprasegmental mismatch in lexical access", Journal of Memory and Language, 45(3), 412-432, 2001.