9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Use of Tonal Information in Korean Lexical Access

Annie Tremblay[1], *Seulgi Shin*[1], *Sahyang Kim*[2], *Taehong Cho*[3]

[1]University of Kansas, USA
[2]Hongik University, South Korea
[3]Hanyang University, South Korea

atrembla@ku.edu, seulgi.shin@ku.edu, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Abstract

Prominence in Seoul Korean is realized at the level of the Accentual Phrase (AP), with the AP-final High (H) tone signaling prosodic word-final boundaries and the AP-initial Low (L) tone signaling word-initial boundaries [1–2]. Using word-spotting experiments, Kim and Cho [3] showed that Korean speech segmentation benefits from both the AP-final H and AP-initial L tones, but it is unclear whether (and if so, how) tonal information also constrains lexical access in Korean. The present study investigates this issue using a visual-world eye-tracking experiment.

Native Korean listeners heard sentences containing a temporary lexical ambiguity between a disyllabic target word in AP-initial position (e.g., *[saesinbu-ga]$_{AP}$ [masul-eul]$_{AP}$* 'the-new-bride-subj magic-obj') and a disyllabic competitor word spanning the AP boundary (e.g., *gama* 'palanquin'). The auditory stimuli were resynthesized to create four tonal boundary conditions: H#L, H#H, L#L, and L#H, where # represents an AP boundary. Listeners' eye movements to the printed target and competitor words were monitored as they heard the auditory stimuli. The results showed independent effects of the AP-initial and AP-final tones on lexical access, suggesting that the intonational system of Korean modulates lexical activation and highlighting the importance of language-specific tonal cues in lexical access.

Index Terms: speech segmentation, Korean, tonal cues, eye tracking

1. Introduction

An increasingly large body of research has shown that the intonational system of the native language helps listeners break the speech signal down into individual words [3–7] and modulates lexical access [4–7]. Seoul Korean (henceforth, Korean) is one example of language whose intonation helps listeners locate word boundaries in continuous speech.

One prosodic level that has been analyzed as being marked by intonation in Korean is the AP. APs have the basic underlying tonal pattern of L(HL)H or H(HL)H, with the first tone being H if the AP-initial segment is fortis or aspirated and L otherwise [1–2]. Across APs of different lengths, the initial L tone is aligned with the first syllable of the AP and the final H tone is aligned with the last syllable of the AP [1–2]. Thus, in Korean, the AP-initial L tone signals prosodic word-initial boundaries (for words not beginning with a fortis/aspirated segment) and the AP-final H tone signals prosodic word-final boundaries.

Using word-spotting experiments, Kim and Cho [3] found that Korean listeners' speech segmentation was less error prone if the AP-initial tone was L than if it was H (for target words not beginning with a fortis/aspirated segment), and if the AP-final tone was H than if it was L. This suggests that the tonal pattern of the AP in Korean guides listeners' speech segmentation. Crucially, the AP-initial L tone enhanced segmentation only in the presence of an AP-final H tone, and the AP-final H tone enhanced segmentation only in the presence of an AP-initial L tone. Kim and Cho [3] thus suggested that it is the contrast between the H and L tones that helps Korean listeners break the speech signal down into individual words.

Korean is thus a language where prosodic word boundaries are systematically marked by tonal information, making it possible for listeners to use that information in speech segmentation. One question that arises, however, is whether (and if so, how) tonal information constrains lexical activation in Korean. Finding that intonational cues to word boundaries modulate lexical access in Korean would suggest that tonal information is processed in parallel with segmental information [5, 8]. Such a finding would have important implications for existing computational models of spoken word recognition (e.g., [9–11]), which do not incorporate the use of suprasegmental information in word recognition (but for an attempt with lexical tones, see [12]).

The present study thus re-examines Korean listeners' use of tonal information in speech segmentation. It does so with a visual-world eye-tracking experiment, thus elucidating the effect of tonal cues on Korean listeners' word activation over time, as indexed by their fixations to target and competitor words (for a discussion of the linking hypothesis between word activation and target and competitor fixations, see [13]). In a design similar to that of Tremblay and colleagues [6–7], participants heard sentences containing a temporary lexical ambiguity between a disyllabic target word in AP-initial position (e.g., *[saesinbu-ga]$_{AP}$ [masul-eul]$_{AP}$* 'the-new-bride-subj magic-obj') and a disyllabic competitor word spanning the AP boundary (e.g., *gama* 'palanquin'). The AP-final and AP-initial tones were manipulated to create four tonal boundary conditions: H#L, H#H, L#L, and L#H. The present study thus sheds direct light on how language-specific tonal information modulates target and competitor word activation in Korean.

2. Method

2.1. Participants

A total of 31 adult native Korean listeners (mean age: 25.5, standard deviation: 2.6, 10 females) participated in this study. All listeners were tested at a university in Seoul, South Korea.

10.21437/SpeechProsody.2018-166

2.2. Materials

Experimental sentences were created that contained a temporary lexical ambiguity between a disyllabic target word in post-boundary AP-initial position and a disyllabic competitor word spanning the AP boundary. The experimental sentences had the structure [Adverb]$_{AP}$ [Subject+case-marker]$_{AP}$ [Object+case-marker]$_{AP}$ [Verb]$_{AP}$. In these sentences, the subject always ended with the case marker *–ga, –i,* or *–to.* The target word was the following disyllabic object (e.g., *masul* 'magic'), and the competitor word was a disyllabic word that began with the same syllable as the case marker and ended with the first syllable of the target word (e.g., *gama* 'palanquin'). All sentential contexts preceding the target word were semantically compatible with both the target and competitor words (e.g., *banggeum saesinbu-ga masul-eul*... 'just-now the-new-bride-subj magic-obj …' vs. *banggeum saesinbu gama*... 'just now the new-bride's palanquin…'). The experiment contained 36 experimental sentences, with the above three case markers each appearing in 12 sentences.

The 36 experimental sentences were interspersed with 108 filler sentences, 8 of which were used in the practice session. Of these filler sentences, 36 had the same sentence structure and the target word in the same position as the experimental sentences, but the subject did not contain a case marker, and the target word instead began with the syllables *ga–, i–,* or *to–,* each in 12 sentences. In these filler sentences, the competitor word began with the second syllable of the target word (e.g., target: *gaji* 'eggplant'; competitor: *jijin* 'earthquake'). The remaining filler sentences had a similar sentence structure but the location of the adverb varied across sentences. Of these sentences, 40 had the target word in subject position and 32 sentences had the target word in the adverb position. For these sentences, the competitor word overlapped with the target word in its first syllable (e.g., target: *namu* 'tree'; competitor: *nabi* 'butterfly').

The visual display contained orthographic representations of the target and competitor words, and of two distracter words (for a validation of the use of orthography in visual-world eye-tracking experiments, see [14]). The distracter words were phonologically and semantically unrelated to the target and competitor words, and showed the same type of segmental overlap with each other as did the target and competitor words (e.g., for the experimental items, the first syllable of one distractor was the second syllable of the other distracter).

Three repetitions of the sentences were recorded by a female native speaker of Korean. The experimental sentences were then resynthesized in order to create four tonal boundary conditions: H#L, H#H, L#L, and L#H. In the natural productions, the subject ended with the AP-final H tone and the object began with the AP-initial L tone. For each sentence, the stimulus in the H#L condition was created by mimicking the H#L tones from a different recording of the same sentence; the stimulus in the H#H condition was created by extending the AP-final H tone of the resynthesized H#L stimulus to the following AP-initial syllable, with a slight decline over the AP-initial H tone so that the rest of the contour would sound natural; the stimulus in the L#L condition was created by extending the AP-initial L tone of the resynthesized H#L stimulus to the previous AP-final syllable; and the stimulus in the L#H condition was created by reversing the AP-final H tone and the AP-initial L tone of the resynthesized H#L stimulus, with a slight decline over the AP-initial H tone so that the rest of the contour would sound natural. The filler sentences were similarly resynthesized so that the experimental sentences would not stand out. The resynthesis was done manually using the PSOLA function in Praat [15]. Figure 1 shows the four pitch contours created for an example experimental sentence.

The experimental items were distributed in four lists, with participants hearing sentences in only one of the four tonal conditions. The four tonal conditions were counterbalanced across the 108 filler sentences.

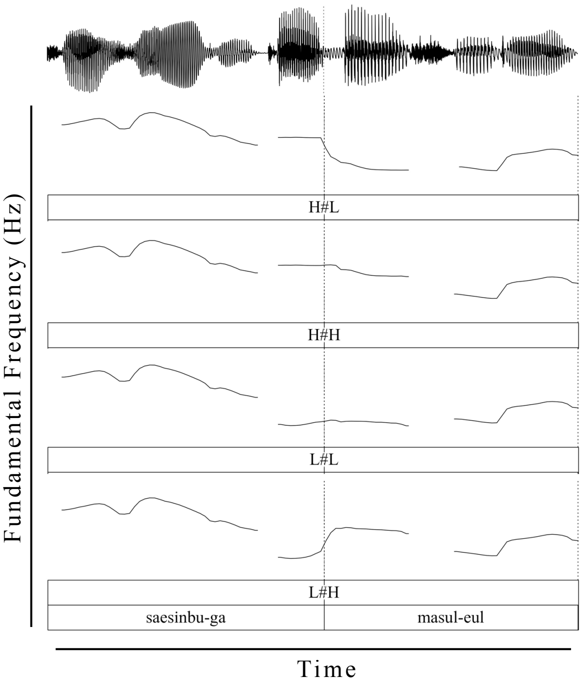

Figure 1: *Subject-object phrases from example sentence in all four tonal boundary conditions.*

2.3. Procedures

Experiment Builder software (SR Research) was used to create and administer the eye-tracking experiment, and Eyelink software (SR Research) was used to monitor participants' eye movements. Eye movements were recorded at a sampling rate of 250 Hz using the head-mounted EyeLink II eye-tracker. The stimuli were heard with an ASIO-compatible sound card, ensuring accurate audio timing in relation to the recording of eye movements.

The eye-tracking experiment began with the practice session (8 trials) followed by the main experiment (136 trials). In each trial, participants saw four orthographic words in a (non-displayed) 2 x 2 grid for 3,000 milliseconds before a fixation cross appeared in the middle of the screen for 500 milliseconds; as the fixation cross disappeared, the four words reappeared on the screen in their original position and the auditory stimulus was heard (synchronously) over headphones. Participants were instructed to click on the target word with the mouse as soon as they heard the target word in the stimulus. Their eye movements were recorded from the onset of the target word (e.g., *masul*). The trial ended with the participants' response, with an inter-trial interval of 1,000 milliseconds.

The 36 experimental trials and the 100 filler trials from the main experiment were presented in four blocks (34 trials per block), with each block containing 9 experimental trials. The order of the experimental and filler trials within a block and the order of blocks were randomized across participants. The participants were offered to take a break after the second block. The experiment lasted approximately 20 minutes.

2.4. Data analyses

Experimental trials that received distracter responses or no response, or for which eye movements could not reliably be tracked, were excluded from the analyses. This resulted in the exclusion of 1.25% of the trials.

For the remaining trials, we analyzed participants' eye movements in the four regions of interest, corresponding to the four orthographic words on the screen. Proportions of fixations to the target, competitor, and distracter words were extracted in 8-ms time windows from the onset of the target word to 1,400 ms post-target-word onset for the purpose of data visualization. The proportions of target and competitor fixations were then averaged from 500 ms to 1,000 ms (when lexical competition effects begin and end) for the statistical analysis.

Linear mixed-effects models were conducted on the *difference* between participants' averaged proportions of target fixations and their averaged proportions of competitor fixations (in the 500-1,000-ms time window). (Analyses conducted separately on the proportions of target fixations and on the proportions of competitor fixations yield the same pattern of results.) We ran two types of analysis. In the first analysis, the largest model included the AP-final tone (L, H), the AP-initial tone (H, L), and their interaction as fixed effects, with participants' performance in H#L condition as the baseline. In the second analysis, the largest model included the tonal boundary condition (H#L, H#H, L#L, L#H) as fixed effect, with participants' performance in the H#H and L#L conditions as baselines. Both types of analyses included participant and item as crossed random effects. We then backward fit these models using log-likelihood ratio tests. We report the results of the simplest models that accounted for significantly more of the variance than simpler models.

3. Results

Figure 2 shows participants' proportions of target, competitor, and distracter fixations in the four tonal boundary conditions over time, and Figure 3 shows the averaged difference between participants' proportions of target and competitor fixations in the 500-1,000-ms time window for all four conditions.

3.1. Effects of AP-final and AP-initial tones

In the first analysis, the model with the best fit included the effects of AP-final and AP-initial tones, but no interaction between them. The results of this model are shown in Table 1.

The model in Table 1 yielded significant effects of AP-final and AP-initial tones: Compared with the baseline H#L condition, both the AP-final L tone and AP-initial H tone *decreased* the difference between participants' proportions of target and competitor fixations. This suggests that there is more lexical competition between the post-boundary target and its competitor straddling the boundary when the AP-final tone is L and when the AP-initial tone is H than when the boundary tone is H#L. The model was not improved by adding the interaction between the AP-final and AP-initial tones.

These results indicate that the AP-final and AP-initial tones independently modulated Korean listeners' target-over-competitor word activation, as reflected by the difference between their fixations to the target and competitor words.

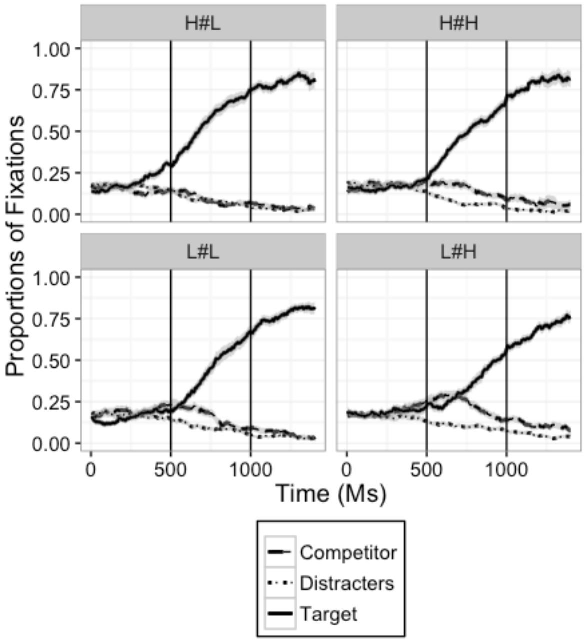

Figure 2: *Proportions of target, competitor, and distracter fixations in the four tonal conditions over time (the shaded area represents 1 standard error above/below the mean; the vertical lines represent the time window of analysis).*

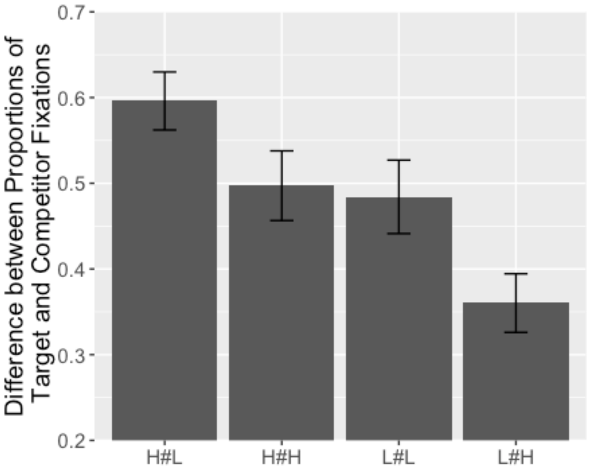

Figure 3: *Difference between participants' averaged proportions of target and competitor fixations in the 500-1,000-ms time window of analysis (the error bars represent 1 standard error above/below the mean).*

Table 1: *Results of linear mixed-effects model with best fit on the difference between participants' proportions of target and competitor fixations in the 500-1,000-ms time window with AP-final and AP-initial tones as fixed effects (est. = estimate; SE = standard error).*

Effect	Est.	SE	*t*	*p*
Intercept	0.631	0.027	23.439	<.001
AP-final tone (L)	–0.133	0.022	–6.144	<.001
AP-initial tone (H)	–0.106	0.022	–4.840	<.001

3.2. Effect of tonal boundary condition

In the second analysis, the model with the best fit included the effect of tonal boundary condition. The results of this model are presented in Table 2 for when the baseline was the H#H condition and in Table 3 for when the baseline was the L#L condition.

Table 2: *Results of linear mixed-effects model with best fit on the difference between participants' proportions of target and competitor fixations in the 500-1,000-ms time window, with tonal boundary condition as fixed effect and with H#H condition as baseline (est. = estimate; SE = standard error).*

Effect	**Est.**	**SE**	***t***	***p***
Intercept	0.534	0.029	18.418	<.001
Condition (H#L)	0.089	0.031	2.892	<.004
Condition (L#L)	–0.028	0.031	<|1|	>.1
Condition (L#H)	–0.150	0.031	–4.884	<.001

Table 3: *Results of linear mixed-effects model with best fit on the difference between participants' proportions of target and competitor fixations in the 500-1,000-ms time window, with tonal boundary condition as fixed effect and with L#L condition as baseline (est. = estimate; SE = standard error).*

Effect	**Est.**	**SE**	***t***	***p***
Intercept	0.613	0.022	27.935	<.001
Condition (H#L)	0.074	0.022	3.327	<.001
Condition (H#H)	0.027	0.022	1.226	>.1
Condition (L#H)	–0.069	0.022	–3.107	<.001

The model in Tables 2 and 3 revealed a significant effect of tonal boundary condition: Compared with the baseline H#H and L#L conditions, the difference between participants' proportions of target and competitor fixations was *larger* in the H#L condition and *smaller* in the L#H condition, and did not differ between the H#H and L#L conditions.

These results suggest that the AP-final H tone and the AP-initial L tone cumulatively enhanced Korean listeners' target-over-competitor word activation, and the AP-final L tone and the AP-initial H tone cumulatively inhibited Korean listeners' target-over-competitor word activation. In other words, the two tones had an additive effect on listeners' word activation.

4. Discussion and Conclusion

The results of the eye-tracking experiment showed that Korean listeners' lexical access was independently modulated by the AP-final and AP-initial tones: Target-over-competitor word activation, as reflected by the difference between listeners' proportions target and competitor fixations, was greater when the AP-final tone was H than when it was L, and it was greater when the AP-initial tone was L than when it was H. Importantly, the effect of the AP-final and AP-initial tones was cumulative, with participants showing an intermediate level of target-over-competitor word activation when one tone was expected but the other was not (i.e., H#H, L#L).

The finding that the canonical H#L tones enhance Korean listeners' speech segmentation is in line with that of Kim and Cho [3], and additionally suggests that tonal information modulates lexical access. However, the current results differ from those of Kim and Cho [3] in that the H#H and L#L conditions yielded intermediate levels of target-over-competitor word activation: The conditions where only one tone was unexpected (i.e., H#H, L#L) generated higher target-over-competitor word activation than the condition where both the AP-final and AP-initial tones were unexpected (i.e., L#H). This suggests that a contrast between the AP-final and AP-initial tones may not be necessary for listeners to use tonal information in speech segmentation and lexical access. The discrepancy between the present findings and those of Kim and Cho [3] may be due to the different methodologies used in the two studies, with eye-tracking being extremely sensitive to the effects of fine-grained acoustic information [16–18] and better capturing how this information modulates word recognition as the speech signal unfolds over time.

These results confirm that intonational cues are processed in parallel with segmental information and constrain lexical access [5, 8]. This finding has important implications for existing computational models of spoken word recognition (e.g., [9–11]). In order to explain and simulate the effects of tonal information on word recognition, these models would have to be able to simultaneously take as input a variety of acoustic cues, both segmental and suprasegmental, and make lexical inferences based on these cues. For example, Cho and colleagues [8] proposed that the word recognition system could include a "Prosody Analyzer," which would extract the prosodic structure of the utterance being heard based on both segmental (e.g., some instantiations of domain-strengthening) and suprasegmental (e.g., duration, pitch) cues, and use this prosodic structure to make inferences about the location of word boundaries and thus to modulate the lexical activation process. Existing computational models of spoken word recognition do not currently incorporate the use of suprasegmental information in word recognition (though see [12]). To simulate the use of this information, computational models should be developed that parse input in a sufficiently fine-grained manner and incorporate the use of multiple cues (both segmental and suprasegmental) in the word recognition process.

All in all, the present study provides additional evidence that the intonational system of the language helps listeners break the speech signal down into words and modulates lexical access when this system predictably marks word boundaries, as in Korean. The degree to which intonational cues constrain lexical access, however, may depend on the variability of the tonal realization allowed by the intonational grammar of the language. This warrants further cross-linguistic research on the role of intonation in lexical processing.

5. Acknowledgements

This research is based upon work supported by the National Science Foundation under grant no. BCS-1423905 awarded to the first author (AT).

6. References

[1] S.-A. Jun, "The accentual phrase in the Korean prosodic hierarchy," *Phonology,* vol. 15, no. 2, pp. 189–226, 1998.

[2] S.-A. Jun, "K-ToBI (Korean ToBI) labeling conventions," *UCLA Working Papers in Phonetics,* vol. 99, pp. 149–173, 2000.

[3] S. Kim and T. Cho, "The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean," *Journal of the Acoustical Society of America,* vol. 125, no. 5, pp. 3373–3386, 2009.

[4] A. Christophe, S. Peperkamp, C. Pallier, E. Block, and J. Mehler, J., "Phonological phrase boundaries constrain lexical access I: Adult data," *Journal of Memory and Language,* vol. 51, no. 4, pp. 523–547, 2004.

[5] A. P. Salverda, D. Dahan, and J. M. McQueen, "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," *Cognition,* vol. 90, no. 1, pp. 51–89, 2003.

[6] A. Tremblay, M. Broersma, C. E. Coughlin, and J. Choi, "Effects of the native language on the learning of fundamental frequency in second-language speech segmentation," *Frontiers in Psychology,* vol. 7, article 985, 2016.

[7] A. Tremblay, M. Broersma, and C. E. Coughlin, "The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in the second language," *Bilingualism: Language and Cognition,* in press.

[8] T. Cho, J. M. McQueen, and E. A. Cox, "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," *Journal of Phonetics,* vol. 35, no. 2, pp. 210–243, 2007.

[9] J. L. McClelland and J. L Elman, "The TRACE model of speech perception," *Cognitive Psychology,* vol. 18, no. 1, pp. 1–86, 1986.

[10] D. Norris, "Shortlist: A connectionist model of continuous speech recognition," *Cognition,* vol. 52, no. 3, pp. 189–234, 1994.

[11] D. Norris and J. M. McQueen, "Shortlist B: A Bayesian model of continuous speech recognition," *Psychological Review,* vol. 115, no. no. 2, pp. 357–395, 2008.

[12] L. Shuai, and J. G. Malins, "Encoding lexical tones in jTRACE: A simulation of monosyllabic spoken word recognition in Mandarin Chinese," *Behavior Research Methods,* vol. 49, no. 1, pp. 230–241, 2017.

[13] A. P. Salverda, M. Brown, and M. K. Tanenhaus, "A goal-based perspective on eye movements in visual world studies," *Acta Psycholica,* vol. 137, no. 2, pp. 172–180, 2011.

[14] J. M. McQueen and M. C. Viebahn, "Tracking recognition of spoken words by tracking looks to printed words," *Quarterly Journal of Experimental Psychology,* vol. 60, no. 5, pp. 661-671, 2007.

[15] P. Broersma and D. Weenink, "Praat: Doing phonetics by computer" [computer program], version 6.0.36, retrieved from from http://www.praat.org, 11 November 2017.

[16] D. Dahan, J. S. Magnuson, M. K. Tanenhaus, & E. M. Hogan, "Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition," *Language and Cognitive Processes,* vol. 16, no. 5–6, pp. 507–534, 2001.

[17] B. McMurray, M. K. Tanenhaus, and R. N. Aslin, "Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition," *Journal of Memory and Language,* vol. 60, no. 1, pp. 65–91, 2009.

[18] K. B. Shatzman and J. M. McQueen, "Segment duration as a cue to word boundaries in spoken-word recognition," *Perception & Psychophysics,* vol. 68, no. 1, pp. 1–16, 2006.