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

Gradient Effects of Tonal Scaling in the Segmentation of Korean Speech: An Artificial-Language Segmentation Study

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

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

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

Abstract

French and Korean have similar intonational systems but differ in the alignment of the phrase-final High (H) tone and scaling of the following phrase-initial Low (L) tone. Tremblay et al. [1] found that Korean listeners have difficulty using tonal cues to segment French speech, raising the question of whether Korean listeners' segmentation of French was inhibited by the different alignments of the phrase-final H tone or scaling of the phrase-initial L tone in the two languages. This study investigates this issue, thereby shedding light on the importance of fine-grained language-specific tonal cues in speech segmentation.

Native Korean listeners completed three artificial-language (AL) segmentation tasks over three sessions. In Experiment 1, one AL contained no tonal cues to word-final boundaries (control), one contained French alignment cues, and one contained Korean alignment cues. Experiments 2 and 3 were identical to Experiment 1, except the phrase-initial L tone in the ALs containing prosodic cues was lowered by 20 Hz and by 40 Hz, respectively. The results showed that Korean listeners' segmentation of the ALs improved as the phrase-initial L tone was lowered, highlighting the gradient effects of tonal scaling in Korean listeners' speech segmentation, consistent with the intonational grammar of the language.

Index Terms: speech segmentation, Korean, tonal cues, artificial language

1. Introduction

A large body of research has shown that speech segmentation is a language-specific skill: To segment an unfamiliar language, listeners use those cues that are reliable for locating word boundaries in the native language [2–8]. The general finding from this research is that if a given cue signals the same word boundary in the native and unfamiliar languages, segmentation of the unfamiliar language is enhanced, whereas if a given cue signals different word boundaries in the two languages, segmentation of the unfamiliar language is inhibited [3, 6–7]. More recently, however, Tremblay et al. [1] have shown that for a non-native or unfamiliar language to be successfully segmented, it is not sufficient for a particular cue to signal the same boundary in the native and target languages; the two languages must also be similar in how this cue is realized at a fine-grained phonetic level.

Tremblay et al. [1] examined the use of fundamental frequency (F0) rise as a cue to word-final boundaries in French by native French listeners and Korean-speaking second-language (L2) learners of French (and English-speaking L2 learners of French). In French, prominence is realized at the level of the Accentual Phrase (AP), with the last non-reduced syllable of non-utterance-final APs ending with an F0 rise categorized as an H tone [9–11]. Korean is similar to French in that prominence is also phrasal, with the final syllable of the AP also ending with an F0 rise categorized as an H tone [12–13]. Importantly, French and Korean show subtle differences in the alignment of the AP-final H tone with the syllable and in the scaling of the following AP-initial L tone: In French, the AP-final H tone peaks at the very end of the AP-final syllable, and thus the pitch lowering begins in the following AP-initial syllable [10, e.g., Figure 8a]; by contrast, in Korean, the AP-final H tone peaks earlier in the AP-final syllable, and the pitch lowering begins in that syllable such that the pitch is low by the beginning of the following AP-initial syllable [13, e.g., (9)]. This alignment difference between French and Korean results in a relatively lower AP-initial L tone in Korean than in French.

The participants in Tremblay et al. [1] completed an eye-tracking experiment with stimuli where the monosyllabic target word and the first syllable of the following adjective (e.g., *chat lépreux* 'leprous cat') were temporarily ambiguous with a disyllabic competitor word (e.g., *chalet* 'cabin'). The monosyllabic target word was manipulated so that it would contain or not contain an AP-final H tone. The presence of the H tone was predicted to reduce the activation of the lexical competitor and thus result in higher proportions of target fixations and lower proportions of competitor fixations. Such results were confirmed for native French listeners (and English L2 learners of French) but not for Korean L2 learners of French. The authors argued that the intonational differences between French and Korean were subtle enough that Korean listeners likely assimilated the French intonational system to their native system, and thus were unable to use the AP-final H tone to locate word-final boundaries in French (i.e., the H tone came in too late for them to be able to use it).

These results highlight the importance of fine-grained language-specific tonal cues in speech segmentation: For a non-native or unfamiliar language to be segmented, a given cue must not only signal the same boundary in the native and target languages, but also be realized similarly at a fine-grained phonetic level. However, these results also raise the question of whether Korean listeners' segmentation of French in [1] was inhibited by the different alignments of the AP-final H tone or by the different scaling of the AP-initial L tone in the two languages. Korean speech segmentation has been shown to benefit from both the AP-final H and AP-initial L tones [14], but it is unclear whether differences in the alignment of the AP-final H tone and scaling of the AP-initial L tone affect segmentation. This study investigates this issue, thus clarifying the role of fine-grained tonal cues in speech segmentation.

10.21437/SpeechProsody.2018-13

2. Method

This study uses an AL segmentation paradigm. This paradigm consists of two phases: a familiarization phase during which listeners hear a continuous flow of syllables (i.e., the AL) for a given duration of time; and a test phase during which, in each trial, participants hear two short strings and decide which of the two strings formed a word in the AL.

2.1. Experimental design

In Experiment 1, three ALs containing CVCVCV words were heard in three tonal alignment conditions: one AL contained no tonal cues to word-final boundaries (control), one contained the AP-final H tone peaking at the very end of the word-final syllable (French alignment cues), and one contained the AP-final H tone peaking earlier in the word-final syllable (Korean alignment cues). Experiments 2 and 3 were identical to Experiment 1, except that the early pitch of the AP-initial L tone in the ALs containing prosodic cues was lowered by 20 Hz and by 40 Hz, respectively. The same listeners heard the three AP-final tonal alignment conditions (none, French cues, Korean cues) with three different ALs over three sessions (within-subject), and different listeners heard the three AP-initial tonal scaling (Experiments 1–3) conditions (between-subject).

2.2. Participants

A total of 108 adult native Korean listeners (mean age: 23.9, standard deviation: 2.9, 43 females) participated in this study, with 36 listeners completing each experiment. All listeners were tested at a university in Seoul, South Korea.

2.3. Materials

2.3.1. Artificial languages (familiarization phase)

Seven consonants (/p, t, k, s, n, m, l) and five vowels (/a, e, i, o, u/) were used to create 33 syllables; these syllables were then combined to create 18 trisyllabic words, six of which occurred in each of the three ALs: (i) [lapame], [nelaki], [liteno], [patute], [tunomu], [kilipo]; (ii) [setika], [nipuko], [sukolo], [monipu], [pemoma], [kamati]; and (iii) [pinuku], [soleta], [leketo], [nutake], [sakumi], [nasopi]. All the syllables were phonotactically possible in Korean, and none of the created words existed in Korean.

The syllables were recorded individually by a female native speaker of Castilian Spanish. (Since this study was going to be a cross-linguistic study, we selected a native speaker of a language that we did not intend to test in order to avoid possible segmental biases.) Following Kim et al. [3], the individual syllables were normalized to have a duration of 252 ms. They were then concatenated to create the trisyllabic words.

To create the French AP-final tonal alignment condition of Experiment 1, a female native speaker of French recorded a series of 14 consecutive CVCVCV French words three times, all of which contained only voiced consonants. The pitch contour of the second through the thirteenth word in each repetition was extracted, and the average pitch contour was calculated. This average pitch contour was then superimposed over the trisyllabic words of all three ALs using the PSOLA (Pitch Synchronous Overlap Add) function of Praat [15].

To create the Korean AP-final tonal alignment condition of Experiment 1, a female native speaker of Korean recorded a series of 14 consecutive CVCVCV Korean words three times, all of which contained only phonetically voiced consonants in word-internal position. The pitch contour of the second through the thirteenth word in each repetition was similarly extracted, and the average pitch contour was calculated. The peak of the word-final (also AP-final) H tone was then identified, and the French contour (recorded from the French speaker) was superimposed over the trisyllabic words of the three ALs such that the peak of its AP-final H tone would occur with the same alignment as the AP-final H tone in Korean: In other words, in the Korean alignment condition, listeners heard the French contour realigned such that its AP-final H tone would peak earlier. (Recall that the goal of the study was to test whether Korean listeners had difficulty segmenting French because of the late alignment of the AP-final H tone or because of the higher scaling of the AP-initial L tone; it was therefore necessary to use the shape of the French contour in both alignment conditions.) Whereas the AP-final H tone in French peaked 95% into the AP-final syllable, the AP-final H tone in Korean peaked 75% into the AP-final syllable.

In the control condition, which did not contain prosodic cues to word boundaries, the trisyllabic words all had a flat F0 of 210 Hz (average of the French contour).

All words were randomly concatenated such that each word would be heard a total of 126 times throughout the AL. No word occurred twice in a row, and there was no pause between any of the words. The total duration of the AL was approximately 10 minutes, and the participants listened to it twice.

As indicated above, Experiments 2–3 were identical to Experiment 1, except that the early pitch of the AP-initial L tone in the ALs containing prosodic cues was lowered by 20 Hz and by 40 Hz, respectively. The word randomization was identical across tonal alignment and tonal scaling conditions.

Example words with the French and Korean AP-final tonal alignment cues and with the different AP-initial tonal scaling of Experiments 1–3 are presented in Figure 1.

Twenty-second fade-in and fade-out periods were added to the beginning and end of all ALs so that listeners could not use the onset of the initial word and the offset of the final word to locate word boundaries.

2.3.2. Word identification (test phase)

Each AL had its own word identification task. Each word identification task contained 36 pairs of trisyllabic sequences, which were created by combining the six words heard in the AL with six part-word foils. These part-word foils contained the first or last two syllables of a word together with a syllable from another (adjacent) word. Importantly, the part-word foils had been heard in the AL (i.e., the underlined syllables in CV<u>CVCV#CV</u>CVCV and CVCV<u>CV#CVCV</u>CV, where # is a word boundary, represent part-word foils). The trisyllabic sequences were heard with a flat F0 of 210 Hz.

2.4. Procedures

Each group of participants completed three AL segmentation tasks over three sessions. Within each group, the AP-final tonal alignment conditions were counterbalanced across the different ALs, and the order of conditions and of ALs was counterbalanced across participants, with the control condition (without prosodic cues) being heard in the second session.

The experiments were administered using Paradigm software [16]. For each segmentation task, participants heard the AL twice (familiarization phase, approx. 20 mins) and then completed the word identification task (test phase, approx. 5 mins). In each trial of the word identification task, participants

heard a word and a part-word foil with an interstimulus interval of 800 ms. Participants were asked to select which of the two sequences they heard in the AL. The correct answer (first sequence or second sequence) was counterbalanced across trials. Participants' accuracy rates were recorded, and the next trial began as participants entered their responses.

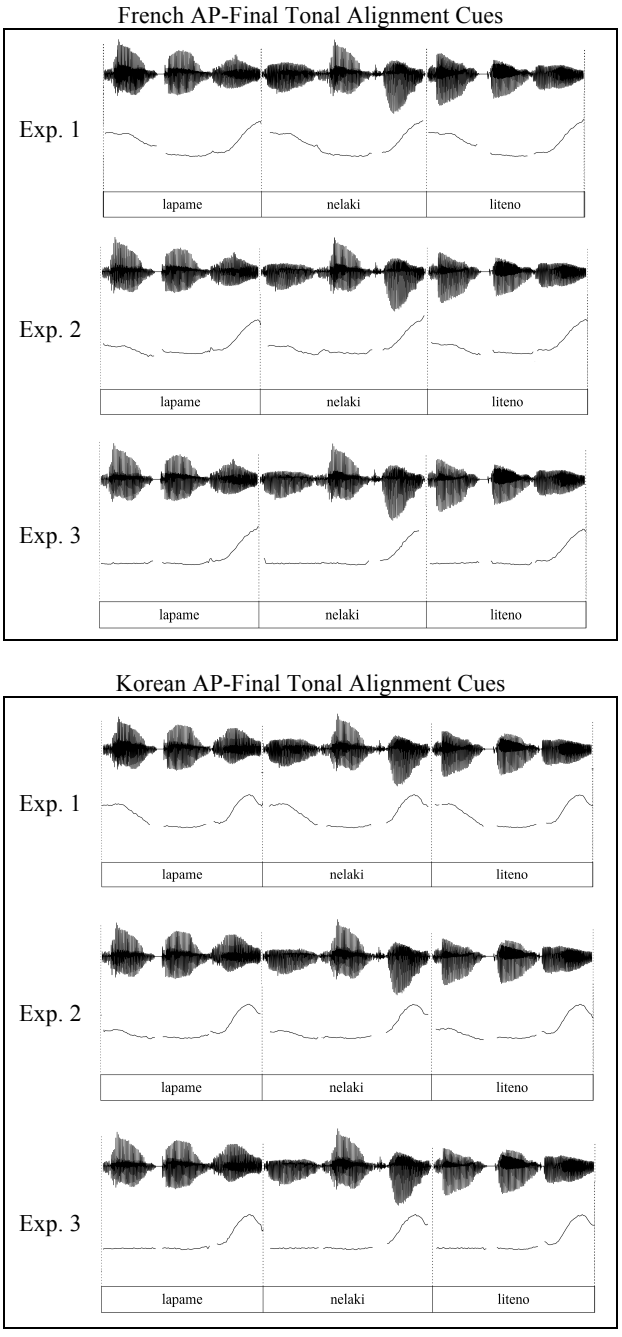

Figure 1: *Example words from the artificial language (French AP-final tonal alignment cues represented in top panel, Korean AP-final tonal alignment cues represented in bottom panel).*

2.5. Data analyses

For each experiment, we ran logit mixed-effects models on participants' word identification accuracy. Each model included the AP-final tonal condition as fixed effect, and participant and item as crossed random effects. We then compared these models to models without any fixed effect. We found that the models with the AP-final tonal condition consistently accounted for significantly more of the variance than the models without this effect, as determined by log-likelihood ratio tests. We therefore report the results of these more complex models. In these models, the baseline was listeners' performance on the control condition.

3. Results

3.1. Experiment 1

Beginning with Experiment 1, participants' proportion of correct responses on each AP-final tonal condition is presented in Figure 2, and the results of the logit mixed-effects model with the best fit on participants' word identification accuracy are presented in Table 1.

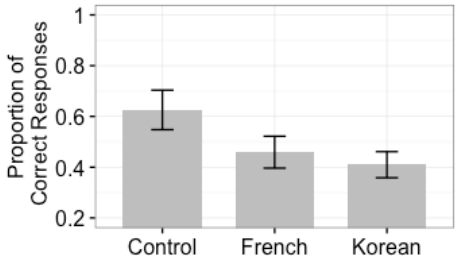

Figure 2: *Proportions of correct responses in Experiment 1 (the error bars represent 1 standard error above/below the mean).*

Table 1: *Results of logit mixed-effects model with best fit on participants' accuracy in Experiment 1 (est. = estimate; SE = standard error).*

Effect	Est.	SE	*z*	*p*
Intercept	0.541	0.110	4.918	<.001
Tonal condition (French alignment)	–0.712	0.082	–8.644	<.001
Tonal condition (Korean alignment)	–0.927	0.083	–11.154	<.001

The model in Table 1 yielded significant effects of AP-final tone for both the French and Korean alignment conditions, with listeners' performance on these conditions being *lower* than their performance on the control condition.

These results indicate that the French contour with a phonetically higher AP-initial L tone *inhibited* Korean listeners' segmentation of the AL. This was true independently of whether the AP-final H tone was aligned late or earlier in the syllable.

3.2. Experiment 2

Continuing with Experiment 2, participants' proportion of correct responses on each AP-final tonal condition is presented in Figure 3, and the results of the logit mixed-effects model with the best fit on participants' word identification accuracy are presented in Table 2.

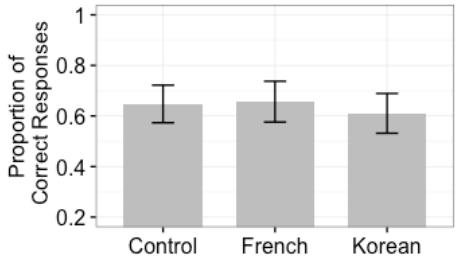

Figure 3: *Proportions of correct responses in Experiment 2 (the error bars represent 1 standard error above/below the mean).*

Table 2: *Results of logit mixed-effects model with best fit on participants' accuracy in Experiment 2 (est. = estimate; SE = standard error).*

Effect	Est.	SE	*z*	*p*
Intercept	0.786	0.196	4.011	<.001
Tonal condition (French alignment)	0.058	0.089	<|1|	>.1
Tonal condition (Korean alignment)	–0.197	0.087	–2.249	.025

The model in Table 2 revealed no effect of AP-final tone for the French alignment condition, and a significant effect of AP-final tone for the Korean alignment condition, with listeners' performance on the Korean alignment condition being *lower* than their performance on the control condition.

These results indicate that when the AP-final H tone was aligned late in the syllable (French alignment condition), the French contour with a 20-Hz lower AP-initial L tone *neither enhanced nor inhibited* Korean listeners' segmentation of the AL. However, when the AP-final H tone was aligned earlier in the syllable (Korean alignment condition), the French contour with a 20-Hz lower AP-initial L tone *inhibited* Korean listeners' segmentation of the AL, suggesting that the earlier alignment of the AP-final H tone is more hurtful than the late alignment of that tone in the presence of a 20-hz lower AP-initial L tone.

3.3. Experiment 3

Moving on to Experiment 3, participants' proportion of correct responses on each AP-final tonal condition is presented in Figure 4 and the results of the logit mixed-effects model with the best fit on participants' word identification accuracy are presented in Table 3.

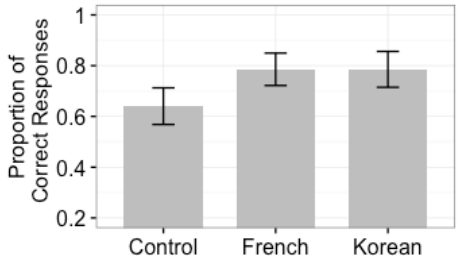

Figure 4: *Proportions of correct responses in Experiment 3 (the error bars represent 1 standard error above/below the mean).*

Table 3: *Results of logit mixed-effects model with best fit on participants' accuracy in Experiment 3 (est. = estimate; SE = standard error).*

Effect	Est.	SE	*z*	*p*
Intercept	0.703	0.197	3.560	<.001
Tonal alignment condition (French)	0.892	0.098	9.099	<.001
Tonal alignment condition (Korean)	0.866	0.097	8.948	<.001

The model in Table 3 yielded significant effects of AP-final tone for both the French and the Korean alignment conditions, with listeners' performance on those conditions being *higher* than their performance on the control condition.

These results indicate that the French contour with a 40-Hz lower AP-initial L tone *enhanced* Korean listeners' segmentation of the AL. This was true independently of whether the AP-final H tone was aligned late or earlier in the syllable.

4. Discussion and Conclusion

Three main findings stem from the above results: First, in the presence of an AP-final H tone, the lower the AP-initial L tone, the better Korean listeners' speech segmentation; second, if the AP-initial L tone is not sufficiently low, Korean listeners cannot use the AP-final H tone as a cue to word-final boundaries; third, even in the presence of a phonetically low AP-initial L tone, Korean listeners' speech segmentation does not benefit more from an AP-final H tone that is aligned earlier in the syllable than from an AP-final H tone that is aligned late in the syllable.

The finding that Korean listeners' successful speech segmentation is dependent on the low scaling of the AP-initial L tone is consistent with the results of Kim and Cho [14]. Using word-spotting experiments, Kim and Cho [14] found that Korean listeners' word detection was less error prone if the AP-final tone was H and the AP-initial tone was L; crucially, they found that the AP-final H tone enhanced segmentation only in the presence of an AP-initial L tone, and the AP-initial L tone enhanced segmentation only in the presence of an AP-final H tone, suggesting an interaction between the AP-final and AP-initial tones in Korean listeners' speech segmentation. The current results are similar, indicating that it is the contrast between the H and L tones that enhance segmentation. Although the earlier alignment of the AP-final H tone in Korean makes it possible for the AP-initial L tone to be phonetically low, Korean listeners' successful speech segmentation does not appear to depend on the alignment of the AP-final H tone itself. Given these results, it is likely that the Korean listeners in Tremblay et al. [1] were unable to use tonal cues to segment French speech because the AP-initial L tone is phonetically higher in French than in Korean.

These results highlight the gradient effects of incremental tonal changes in Korean listeners' speech segmentation, and indicate that the fine-grained phonetic details that arise with tonal alignment, as specified by the intonational grammar of the language [17], play a crucial role in speech segmentation.

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] 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.

[2] A. S. Finn and C. L. Hudson Kam, "The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation," *Cognition,* vol. 108, pp. 477–499, 2008.

[3] S. Kim, M. Broersma, and T. Cho, "The use of prosodic cues in learning new words in an unfamiliar language," *Studies in Second Language Acquisition,* vol. 34, no. 3, pp. 415–444, 2012.

[4] M. Shukla, M. Nespor, and J. Mehler, "An interaction between prosody and statistics in the segmentation of fluent speech," Cognitive Psychology, vol. 54, no. 1, pp. 1–32, 2007.

[5] J. M. Toro, F. Pons, R. A. H. Bion, and N. Sebastián-Gallés, "The contribution of language-specific knowledge in the selection of statistically coherent word candidates," *Journal of Memory and Language,* vol. 64, no. 2, pp. 171–180, 2011.

[6] A. Tremblay, J. Namjoshi, E. Spinelli, M. Broersma, T. Cho, S. Kim, M. T. Martínez-García, and K. Connell, "Experience with a second language affects the use of fundamental frequency in speech segmentation," *PLoS One,* vol. 12, article e0181709, 2017.

[7] M. D. Tyler and A. Cutler, "Cross-language differences in cue use for speech segmentation," *Journal of the Acoustical Society of America,* vol. 126, no. 1, pp. 367–376, 2009.

[8] J. Vroomen and B. de Gelder, "Metrical segmentation and lexical inhibition in spoken word recognition," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 21, no. 1, pp. 98–108, 1995.

[9] S.-A. Jun and C. Fougeron, "A phonological model of French intonation," in A. Botinis (Ed.), *Intonation: Analysis, Modeling and Technology,* 2000, pp. 209–242.

[10] S.-A. Jun and C. Fougeron, "Realization of accentual phrase in French intonation," *Probus,* vol. 14, no. 1, pp. 147–172, 2002.

[11] P. Welby, "French intonational structure: Evidence from tonal alignment," *Journal of Phonetics,* vol. 34, no. 3, pp. 343–371, 2006.

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

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

[14] 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.

[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] Perception Research Systems, "Paradigm stimulus presentation", retrieved from http://www.paradigmexperiments.com, 2007.

[17] D. R. Ladd, *Intonational Phonology.* Cambridge: Cambridge University Press, 2012.