CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study

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

[1]University of Kansas, [2]Hanyang University and [3]Hongik University
*Corresponding author. E-mails: atrembla@ku.edu; tcho@hanyang.ac.kr

Abstract

This study investigates how the fine-grained phonetic realization of tonal cues impacts speech segmentation when the cues signal the same word boundary in the native and unfamiliar languages but do so differently. Korean listeners use the phrase-final high (H) tone and the phrase-initial low (L) tone to segment speech into words (Kim, Broersma, & Cho, 2012; Kim & Cho, 2009), but it is unclear how the alignment of the phrase-final H tone and the scaling of the phrase-initial L tone modulate their speech segmentation. Korean listeners completed three artificial-language (AL) tasks (within-subject): (a) one AL without tonal cues; (b) one AL with later-aligned phrase-final H cues (non-Korean-like); and (c) one AL with earlier-aligned phrase-final H cues (Korean-like). Three groups of Korean listeners heard (b) and (c) in three phrase-initial L scaling conditions (between-subject): high (non-Korean-like), mid (non-Korean-like), or low (Korean-like). Korean listeners' segmentation improved as the L tone was lowered, and (b) enhanced segmentation more than (c) in the high- and mid-scaling conditions. We propose that Korean listeners tune in to low-level cues (the greater H-to-L slope in [b]) that conform to the Korean intonational grammar when the phrase-initial L tone is not canonical phonologically.

Keywords: artificial language; Korean; speech segmentation; tonal cues

A large body of research has shown that speech segmentation is a language-specific skill: to segment an unfamiliar language, listeners use those phonological cues that are reliable for locating word boundaries in the native language (e.g., Cunillera, Toro, Sebastian-Galles, & Rodriguez-Fornells, 2006; Finn & Hudson Kam, 2008; Kim et al., 2012; Ordin & Nespor, 2013, 2016; Shukla, Nespor, & Mehler, 2007; Toro, Pons, Bion, & Sebastian-Galles, 2011; Toro, Sebastián-Gallés, & Mattys, 2009; Tremblay et al., 2017; Tyler & Cutler, 2009; Vroomen & de Gelder, 1995). 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 (e.g., Kim et al., 2012; Tremblay et al., 2017; Tyler & Cutler, 2009). What is unclear from this research, however, is how the fine-grained phonetic realization of a given cue to word boundary impacts speech segmentation in cases where the cue signals the same word boundary in the native and unfamiliar languages but does so differently in the two languages. A great deal of research has shown that fine-grained phonetic details modulate the *process* by which speech segmentation takes place (i.e., the degree to which target and competitor words are activated and compete as the continuous speech signal unfolds; e.g., Cho, McQueen, & Cox, 2007; Salverda, Dahan, & McQueen, 2003; Salverda et al., 2007; Spinelli, McQueen, & Cutler, 2003; Tremblay & Spinelli, 2014). Much less attention has been paid to whether this fine-grained information also has an important impact on the *outcome* of speech segmentation (i.e., whether the intended words are ultimately extracted from the speech signal; e.g., see Spinelli, Grimault, Meunier, & Welby, 2010).

The present study sheds light on this question by examining Korean listeners' use of fine-grained tonal cues in speech segmentation. The intonational system of Korean has been analyzed as having an accentual phrase (AP), with L(HL)H (L = low tone, H = high tone) as the basic underlying tonal pattern in non-utterance-final APs that begin with a lenis segment (Jun, 1998, 2000).[1] The AP-final H tone and AP-initial L tone can thus signal word-final and word-initial boundaries in continuous speech. However, whereas the AP-final H tone is loosely anchored to the AP-final syllable, the following AP-initial L tone is tightly anchored to the AP-initial syllable; this means that the alignment of the AP-final H tone can vary, but it is sufficiently early to allow for the AP-initial L tone to reach its target (and have a low scaling) early on in the AP-initial syllable (Jun, 2000, Example 9). One question that arises from these phonetic details is whether Korean listeners' ability to segment speech into words would be affected by the alignment of the AP-final H tone (in relation to the word-final boundary) and by the scaling (i.e., pitch level) of the subsequent AP-initial L tone. Answering this question would not only clarify how AP-final and AP-initial tonal cues affect Korean listeners' speech segmentation in relation to the intonational phonology of Korean but also shed light on the role of the fine-grained phonetic details of tonal cues on the outcome of speech segmentation.

Korean listeners' speech segmentation has been shown to benefit from both the AP-final H tone and the AP-initial L tone when these tones are realized in their canonical form. For example, using word-spotting experiments where the AP-final tone of the nonsense carrier phrase and the tones of the AP-initial disyllabic target word were independently manipulated as H or L, Kim and Cho (2009, Experiment 1) showed that Korean listeners' ability to extract and recognize real Korean words from nonsense carrier phrases benefited from a preceding AP-final H tone only in the presence of an AP-initial L tone in the target word, and it benefited from an AP-initial L tone in the target word only in the presence of a preceding AP-final H tone. Similarly, using an artificial-language (AL) segmentation paradigm in which word boundaries were signaled by an H-L tone sequence (corresponding to an

AP-final H tone and an AP-initial L tone) or by an L-H tone sequence (corresponding to an AP-final L tone and an AP-initial H tone), Kim et al. (2012, Experiment 1) showed that Korean listeners' ability to extract words from the AL improved (relative to a control condition without tonal cues) only when word boundaries were signaled by a sequence of H-L tone. The results of these two studies indicate that Korean listeners' speech segmentation is enhanced only when word boundaries are signaled by both the AP-final H tone and the AP-initial L tone. However, these two studies leave open the question of whether Korean listeners' speech segmentation would also be affected by the fine-grained phonetic details of the AP-final H tone and AP-initial L tone such as their alignment and scaling, respectively.

A recent study on spoken word recognition with native Korean listeners suggests that the fine-grained phonetic details of the native intonational system have an important effect on speech segmentation in a second language when the native and second languages have similar intonational systems. Tremblay, Broersma, Coughlin, and Choi (2016) investigated the use of fundamental frequency (F0) rise as a cue to word-final boundaries in French by native French listeners and Korean-speaking learners of French (and English-speaking learners of French). Like Korean, the intonational system of French has been analyzed as having an AP, with L(HL)H also being the basic underlying tonal pattern of non-utterance-final APs (Jun & Fougeron, 2000, 2002). Importantly, French differs from Korean 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 tends to peak at or toward the end of the AP-final syllable, and the pitch lowering of the subsequent AP-initial L tone tends to begin in the following AP-initial syllable (e.g., Jun & Fougeron, 2002, Figure 8a; Welby, 2006). Thus, the AP-initial L tone has a relatively higher scaling in French than in Korean.[2] The participants in Tremblay et al. (2016) 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 H tone, realized as an F0 rise that peaked at the offset of the monosyllabic word. The presence of the H tone was predicted to reduce the activation of the lexical competitor (as it signals the end of the word aligned with the right edge of the AP) and thus result in a higher target-over-competitor fixation advantage. Such results were confirmed for native French listeners (and English L2 learners of French) but not for Korean L2 learners of French, who did not show an enhancing effect of F0 rise in their target-over-competitor fixation advantage. On the basis of these results, the authors argued that Korean listeners were unable to learn the fine-grained intonational differences between the two languages, namely, that the AP-final H tone peaks later and the AP-initial L tone has a higher scaling in French than in Korean. Consequently, Korean listeners, as suggested by the authors, were unable to use the French AP-final H tone to locate word-final boundaries in French.

Tremblay et al.'s (2016) results highlight the importance of fine-grained language-specific tonal cues in the speech segmentation *process*: for tonal information to modulate lexical activation and competition in a targetlike manner, 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, it is unclear from these results what kind of fine-grained phonetic detail of prosody led to the inhibitory effect, that is, whether Korean listeners' segmentation of French was inhibited by the different alignments of the AP-final H tone or by the different scaling (or the different phonetic targets) of the AP-initial L tone in the two languages (or by both). In other words, from the perspective of Korean speech segmentation, Tremblay et al.'s (2016) results raise the question of whether Korean listeners are more dependent on the alignment of the AP-final H tone or on the scaling of the AP-initial L tone (or on both). In addition, it is unclear whether the independent manipulation of such fine-grained tonal cues would impact Korean listeners' speech segmentation *outcome* (i.e., listeners' ability to successfully segment words from a continuous speech stream).

The present study seeks to fill these gaps by investigating the effects of AP-final H tone alignment and AP-initial L tone scaling on Korean listeners' speech segmentation. It does so using an AL segmentation paradigm similar to that of Kim et al. (2012). Typical AL segmentation studies with adult listeners consist of two phases: a familiarization phase and a test phase. In the familiarization phase, participants listen to a string of consecutive syllables for a given period of time (e.g., 20 mins); these strings of syllables are concatenated words, and listeners' task is to recognize words in the signal by figuring out where words begin and end. In each trial of the test phase, participants hear two short strings of syllables and must decide which of the two strings formed a word in the AL; one string is a word in the AL, and the other string is either a nonword foil (the syllables were never heard together in the AL) or a partword foil (two consecutive syllables came from a given word in the AL and the other syllable came from an adjacent word in the AL, such that the sequence of the three syllables was heard together at least once in the AL). Research has shown that transitional probabilities alone (i.e., the probability that a given syllable would follow another given syllable) can help listeners' segment an AL in which no other cues to word boundaries are present, with segmentation accuracy typically being significantly above chance (e.g., Monaghan, Chater, & Christiansen, 2005; Saffran, 2001; Saffran, Newport, & Aslin, 1996). The question in such a design is whether tonal cues to word boundaries would enhance or inhibit speech segmentation above and beyond transitional probabilities alone.

An AL segmentation paradigm is ideal for examining the role of fine-grained tonal information, because it makes it possible to isolate effects of subtle tonal cues from effects of lexical information, all the while revealing whether the effects of the different tonal cues are independent. If Korean listeners' speech segmentation benefits from an earlier-aligned AP-final H tone, word-identification performance should be higher following exposure to an AL that contains such a tone than following exposure to an AL that contains a later-aligned AP-final H tone. Similarly, if Korean listeners' speech segmentation benefits from an AP-initial L tone that is phonetically lower and thus more consistent with the phonetic realization of the intonational phonology of Korean, word-identification performance should be higher following exposure to an AL that contains such a tone than following exposure to an AL that contains an AP-initial L tone with a higher scaling. Note that an L tone may also occur in the middle of an AP (i.e., in LHLH, which is the canonical AP tonal distribution in Korean). However, the AP-medial L tone

(in LHLH) is typically scaled higher than the AP-initial L tone (Jun, 2000). The higher scaling of an L tone may therefore be interpreted as suggesting that the L tone occurs in the AP-medial position rather than in the AP-initial position. Given that an AP-medial L tone often occurs in the middle of a word (whereas an AP-initial L tone always co-occurs with a word-initial syllable), the higher L tone scaling consistent with an AP-medial L tone may therefore inhibit speech segmentation. Finally, if the two tonal cues work together rather than independently to signal the prosodic boundary to be aligned with the assumed lexical boundary, we might find that the AP-final H tone will enhance speech segmentation only in the presence of an AP-initial L tone with a low scaling. Relatedly, if one cue is more important than the other for signaling lexical boundaries, we might find a weaker effect of the less important cue when the more important cue provides strong support for the boundary marking. Given the importance of the AP-initial L tone in Korean listeners' segmentation (Kim & Cho, 2009), we anticipate that the effect of AP-initial L tone scaling will be greater than that of the AP-final H tone alignment, and that an effect of AP-final H tone alignment is less likely to emerge when the AP-initial L tone is canonical phonologically (i.e., when it is realized with a low scaling).

These predictions were tested by examining Korean listeners' segmentation of an AL whose prosody mimics that of French, thus providing an ecologically valid way to test how tonal cues impact speech segmentation when the cues signal the same word boundary in the native language and unfamiliar AL but differ in their fine-grained phonetic detail.

Method

Experimental design

This study uses a 3 (AP-final H tone manipulations) × 3 (AP-initial L tone manipulations) design. All listeners heard three ALs containing CVCVCV (consonant–vowel–consonant–vowel–consonant–vowel) words (within-subject design). Two of these ALs contained tonal cues and thus served as experimental conditions. The two experimental conditions differed in their AP-final H tone alignment: one AL contained the AP-final H tone peaking at the very end of the word-final syllable (later-aligned AP-final H tone, consistent with the French F0 rising pattern and deviating from the Korean F0 rising pattern); and one AL contained the AP-final H tone peaking earlier in the word-final syllable (earlier-aligned AP-final H tone, consistent with the Korean F0 rising pattern and deviating from the French F0 rising pattern). The third AL did not contain any tonal cues to word boundaries and thus served as a control condition. The AP-final H tone alignment and control manipulations were treated as a within-subject variable because we anticipated its effect to be subtle and more likely to emerge in the same group of individuals. Consequently, participants were exposed to three different ALs (containing different words) over three testing sessions, with the order of conditions and of ALs being counterbalanced across participants, with the exception that the control condition (without tonal cues) was always heard in the second session. Crucially, three different groups of listeners heard these three ALs, with the

Table 1. Experimental design

	Session 1	Session 2	Session 3
Group 1 (n = 36)	High AP-initial L tone scaling + earlier/later-aligned AP-final H tone	Control (no tonal cues)	High AP-initial L tone scaling + later/earlier-aligned AP-final H tone
Group 2 (n = 36)	Mid AP-initial L tone scaling + earlier/later-aligned AP-final H tone	Control (no tonal cues)	Mid AP-initial L tone scaling + later/earlier-aligned AP-final H tone
Group 3 (n = 36)	Low AP-initial L tone scaling + earlier/later-aligned AP-final H tone	Control (no tonal cues)	Low AP-initial L tone scaling + later/earlier-aligned AP-final H tone

experimental ALs being heard in three AP-initial L tone scaling conditions (between subject): high scaling (most deviating from the canonical realization of the AP-initial L tone in Korean), mid scaling, and low scaling (matched with the canonical F0 realization of the AP-initial L tone in Korean). The experimental design is summarized in Table 1.

Participants

A total of 108 adult native Korean listeners (mean age = 23.9 years, $SD = 2.9$, 43 women) participated in this study, with 36 listeners completing only one of the three AP-initial L tone scaling conditions. All listeners were tested at Hanyang University in Seoul, South Korea.

Materials

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, 6 of which occurred in each of the three ALs: AL1 contained the words [lapame], [nelaki], [liteno], [patute], [tunomu], [kilipo]; AL2 contained the words [setika], [nipuko], [sukolo], [monipu], [pemoma], [kamati]; and AL3 contained the words [pinuku], [soleta], [leketo], [nutake], [sakumi], [nasopi].[3] 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. Because Korean speakers produce segmental cues to AP-initial (and thus word-initial) boundaries (e.g., Cho & Keating, 2001), it was deemed preferable to elicit productions from someone who does not speak Korean so that the AL would not contain Korean-specific segmental cues to word boundaries. Following Kim et al. (2012), the individual syllables were normalized to have a duration of 252 ms, and they were then concatenated to create the trisyllabic words.

To create a later-aligned AP-final H tone condition that would mirror the AP-final H tone alignment of a French speaker, a female native speaker of French recorded a series of 14 consecutive CVCVCV French words three times, all of which contained
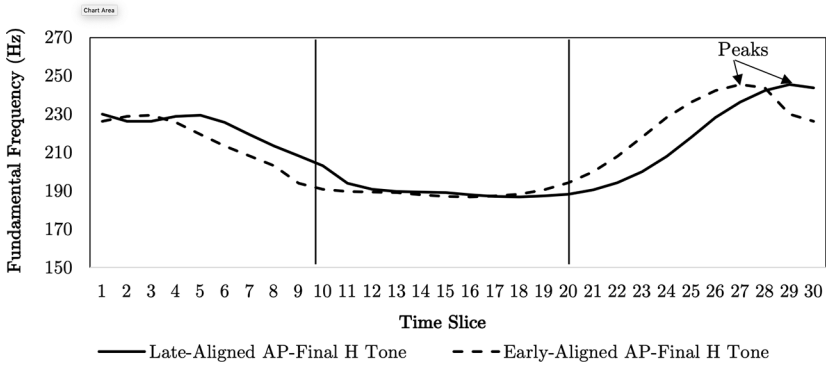
Figure 1. Mean fundamental frequency (F0, in Hertz) for each time slice in the later-aligned and earlier-aligned AP-final H tone conditions with a high AP-initial L tone scaling.

only voiced consonants. The pitch contour of the second through the thirteenth word in each repetition (total: 36 tokens) was extracted in 30 time slices (10 time slices for every syllable), and the average pitch contour was calculated. The average F0 was 220 Hz for the first syllable, 188 Hz for the second syllable, and 220 Hz for the third syllable; the maximum F0 was 230 Hz for the first syllable (time slice 1), 194 Hz for the second syllable (time slice 11), and 245 Hz for the third syllable (time slice 29). The mean F0 (in Hz) for each of the time slices of this later-aligned AP-final H tone condition with a high AP-initial L tone scaling is provided in Figure 1 (solid line). The average pitch contour extracted over the 30 time slices was then superimposed over the trisyllabic words of the ALs using the PSOLA (Pitch Synchronous Overlap Add) function of Praat (Boersma & Weenink, 2017).

To create an earlier-aligned AP-final H tone condition that would mirror the AP-final H tone alignment of a Korean speaker, 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. To identify where the AP-final H tone typically peaks during the AP-final syllable in Korean, the pitch contour of the 2nd through the 13th word in each repetition (total: 36 tokens) was similarly extracted, and the average pitch contour was calculated. The Korean speaker's average pitch contour is presented in Figure A.1 of Appendix A. The peak of the AP-final H tone produced by the Korean speaker was then identified, and the later-aligned AP-final H 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 that of the AP-final H tone in Korean. In other words, in the earlier-aligned AP-final H tone condition, listeners heard the French contour realigned such that its AP-final H tone would peak earlier. (Note that the difference in the peak alignment of the AP-final H tone was clearly observed in our French and Korean recordings: whereas the AP-final H tone in the French recordings peaked 85% into the AP-final syllable [time slice 29], the AP-final H tone in the Korean recordings peaked 65% into the AP-final syllable [time slice 27]; see Figure 1 and Figure A.1 for a comparison.) Because the entire contour was shifted, some of the AP-initial L tone from the
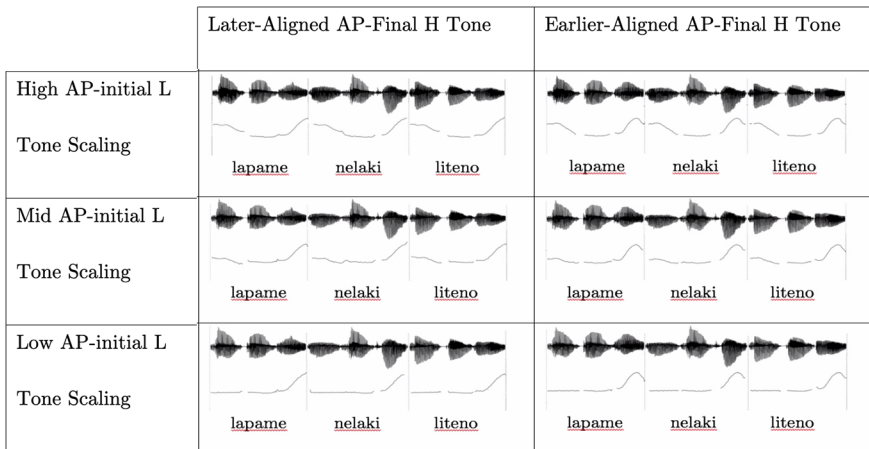
	Later-Aligned AP-Final H Tone	Earlier-Aligned AP-Final H Tone
High AP-initial L Tone Scaling	lapame nelaki liteno	lapame nelaki liteno
Mid AP-initial L Tone Scaling	lapame nelaki liteno	lapame nelaki liteno
Low AP-initial L Tone Scaling	lapame nelaki liteno	lapame nelaki liteno

Figure 2. Example words from the artificial language; the later-aligned and earlier-aligned AP-final H tones are represented in the left and right panels, respectively; the high, mid, and low AP-initial L tone scaling are represented in the top, middle, and bottom panels, respectively.

French contour was heard as part of the AP-final syllable (following the AP-final H tone). This resulted in a slight pitch fall after the AP-final H tone in the earlier-aligned condition, one that closely resembles Korean listeners' tone production (see Jun, 2000, Example 9).[4] The mean F0 (in Hz) for each of the time slices of this earlier-aligned AP-final H tone condition with high AP-initial L scaling is provided in Figure 1 (dashed line). The average pitch contour extracted over the 30 time slices was again superimposed over the trisyllabic words of the ALs using the PSOLA function of Praat (Boersma & Weenink, 2017).

The AP-initial L tone produced by the French speaker had a high scaling and thus was used for the high AP-initial L tone scaling conditions. The low-scaling condition was created by lowering the onset of the AP-initial L tone (i.e., 230 Hz) such that it would be as low as the lowest point of the AP-initial L tone produced by the French speaker. This corresponded to a decrease of 40 Hz in the onset of the AP-initial L tone relative to the high-scaling condition (i.e., 3.31 semitones lower than the high-scaling condition). The mid-scaling condition was created by lowering the onset of the AP-initial L tone by the median distance (i.e., 20 Hz) between the AP-initial L tone onset in the high- and low-scaling conditions (i.e., 1.58 semitones lower than the high-scaling condition and 1.73 semitones higher than the low-scaling condition). The ensuing pitch contours were created by interpolating between the new onset and the natural offset of the AP-initial L tone. This resulted in an F0 with a negative slope for the mid AP-initial L tone scaling condition and in a flat F0 for the low AP-initial L tone scaling condition. Example words from the later-aligned and earlier-aligned AP-final H tone conditions with each of the AP-initial L tone scalings are presented in Figure 2. Finally, to create the control condition (which did not contain tonal cues to word boundaries), the F0 of the trisyllabic words was flattened to 210 Hz (average pitch of the French contour).

All manipulated 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 word randomization was identical across tonal alignment and tonal scaling conditions. 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. The total duration of the AL was approximately 10 minutes, and participants listened to it twice.

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 two syllables of a word together with the last syllable of a preceding (adjacent) word in the AL or the last two syllables of a word together with the first syllable of a following (adjacent) word in the AL (i.e., respectively, the underlined syllables in CVCVCV#CVCVCV and CVCVCV#CVCVCV, where # is a word boundary). It is important to note that all the part-word foils had been heard in the AL. The trisyllabic sequences were heard with a flat F0 of 210 Hz, like that of the control condition.

Procedures

Each group of participants completed three AL segmentation tasks over three sessions, with at least 2 days between each session. Within each group, the different ALs were counterbalanced across the different within-subject conditions (later-aligned AP-final H tone, earlier-aligned AP-final H tone, and control), and the order of conditions and of ALs was counterbalanced across participants, with the exception that the control condition (without tonal cues) was always heard in the second session (see Table 1).

The experiments were administered using Paradigm software (Paradigm Stimulus Presentation, 2007). 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 by pressing one of two designated keys on the keyboard. 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.

Data analyses

We ran logit mixed-effects models on participants' word-identification accuracy using the glmer() function of the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). A first model compared listeners' accuracy in each condition against

chance, with participant and word as crossed random effects. Chance or below-chance performance on any of the experimental conditions would suggest that tonal information interfered with listeners' extraction of transitional probabilities from the input and thus inhibited speech segmentation. A second model examined the effects of AP-final H tone alignment (control [no tone], later-aligned H tone, or earlier-aligned H tone) and AP-initial L tone scaling (high scaling, mid scaling, or low scaling) and their interaction when the baseline was listeners' performance in the control condition whose corresponding experimental conditions contained the AP-initial L tone with a mid scaling. This model also had participant and word as crossed random effects. Using listeners' performance in the control condition as baseline allowed us to examine whether the AP-final H tone and the AP-initial L tone enhanced speech segmentation above and beyond transitional probabilities alone. We compared this model to simpler models in a pairwise fashion using log-likelihood ratio tests, and found that the most complex model (including the interaction between the two tone manipulations) was significantly better than the simpler model with the two simple effects of tone manipulations but no inter-action between them. We therefore report the results of the most complex model, with p values being calculated using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2016). In order to directly compare listeners' perfor-mance in the two alignment and two scaling conditions, the second model was releveled, with the condition containing the later-aligned AP-final H tone and the AP-initial L tone with a mid scaling as baseline. This releveled model therefore shed direct light on whether the AP-final H tone alignment and AP-initial L tone scaling independently modulated speech segmentation. Whenever significant inter-actions were found between the AP-final H tone alignment and the AP-initial L tone scaling, we proceeded to further relevel the models with the high and/or low AP-initial L tone scaling as baseline(s). This releveling allowed us to interpret the simple effects of AP-final H tone alignment for each level of the AP-initial L tone scaling manipulation.

Results

Participants' proportions of correct responses in the word-identification tasks fol-lowing exposure to the nine different ALs are presented in Figure 3. The results of the model that compared listeners' accuracy in all conditions against chance are presented in Table 2.

As can be seen from this model, listeners' performance did not differ significantly from chance when the AL contained the later-aligned AP-final H tone and the AP-initial L tone with a high scaling, and it was significantly below chance when the AL contained the earlier-aligned AP-final H tone and the AP-initial L tone with a high scaling. In all other conditions, performance was significantly above chance. These results confirm that, in the control conditions, listeners were able to use transitional probabilities alone to segment the speech stream into words. Furthermore, these results suggest that speech segmentation was *inhibited* when the AL contained the AP-initial L tone with a high scaling.

Table 2. Results of logit mixed-effects model comparing listeners' accuracy against chance

Condition		Estimate	SE	z	p
High AP-initial L tone scaling	Control (no tonal information)	0.551	0.162	3.406	.001
	Later-aligned AP-final H tone	–0.172	0.161	–1.065	>.1
	Earlier-aligned AP-final H tone	–0.394	0.161	–2.441	.015
Mid AP-initial L tone scaling	Control (no tonal information)	0.764	0.164	4.670	<.001
	Later-aligned AP-final H tone	0.821	0.164	5.013	<.001
	Earlier-aligned AP-final H tone	0.570	0.163	3.499	<.001
Low AP-initial L tone scaling	Control (no tonal information)	0.687	0.163	4.210	<.001
	Later-aligned AP-final H tone	1.549	0.168	9.200	<.001
	Earlier-aligned AP-final H tone	1.530	0.168	9.118	<.001

Note: 11,664 observations, 108 participants, 18 words.

Table 3. Results of logit mixed-effects model with AP-final H tone alignment, AP-initial L tone scaling, and their interaction as fixed effects (baseline: control condition whose corresponding experimental conditions contained the mid AP-initial L tone scaling)

Fixed effects	Estimate	SE	z	p
(Intercept)	0.764	0.164	4.670	<.001
AP-final H tone (later-aligned)	0.058	0.088	<|1|	>.1
AP-final H tone (earlier-aligned)	–0.194	0.087	–2.235	.025
AP-initial L tone (high scaling)	–0.213	0.209	–1.018	>.1
AP-initial L tone (low scaling)	–0.077	0.210	<|1|	>.1
AP-final H tone (later-aligned) × AP-initial L tone (high scaling)	–0.780	0.121	–6.447	<.001
AP-final H tone (earlier-aligned) × AP-final L tone (high scaling)	–0.751	0.121	–6.227	<.001
AP-final H tone (later-aligned) × AP-initial L tone (low scaling)	0.804	0.130	6.169	<.001
AP-final H tone (earlier-aligned) × AP-initial L tone (low scaling)	1.038	0.129	8.040	<.001

Note: 11,664 observations, 108 participants, 18 words.

Table 3 reports the results of the model with the best fit that examined the effects of AP-final H tone alignment, AP-initial L tone scaling, and their interaction, with the baseline being listeners' performance in the control condition whose corresponding experimental conditions contained the AP-initial L tone with a mid scaling. The model with the best fit included all simple effects and their interaction.

As can be seen in Table 3, the simple effect of AP-final H tone alignment (later-aligned) was not significant, suggesting that listeners performed similarly in the later-aligned AP-final H tone condition compared to the control condition when the AP-initial L tone of the corresponding experimental AL had a mid scaling. By contrast, the simple effect of AP-final H tone alignment (earlier-aligned) was significant, with listeners' accuracy being significantly *lower* in the earlier-aligned

Figure 3. Korean listeners' accuracy on the word-identification task following exposure to the nine artificial languages; the x-axis represents the AP-final H tone manipulations; the y-axis represents the proportion of correct responses on the word-identification task; the top, middle, and bottom panels represent the AP-initial L tone scaling manipulations; the error bars represent 1 SE above/below the mean; the horizontal line represents chance performance (0.5).
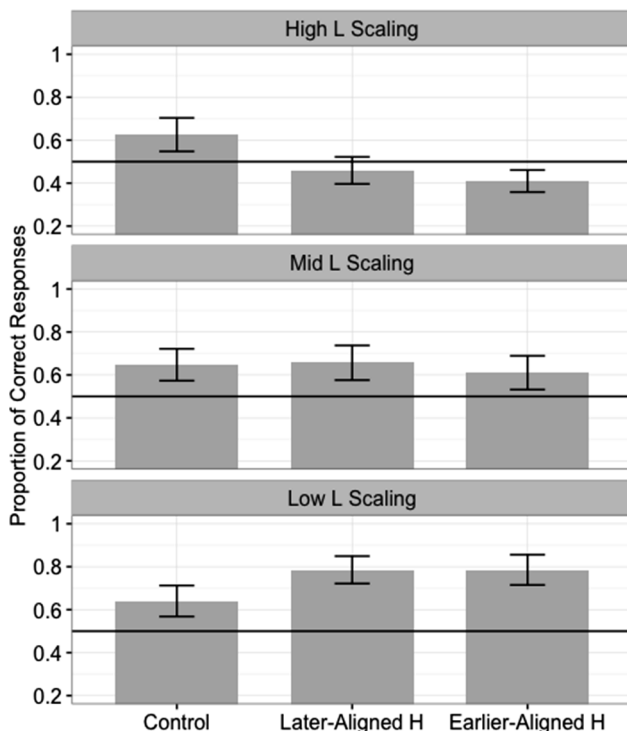
AP-final H tone condition than in the control condition when the corresponding experimental ALs had an AP-initial L tone with a mid scaling. The simple effects of AP-initial L tone scaling (high scaling, low scaling) were not significant, as should be expected as the baseline in this comparison was the control condition where no tonal information was present. The model revealed significant interactions between the AP-final H tone alignment (later-aligned, earlier-aligned) and the AP-initial L tone scaling (high scaling, low scaling). The negative directionality of the interactions for the high-scaling conditions and the positive directionality of the interactions for the low-scaling conditions indicate that the effect of AP-final H tone *hindered* performance in the high-scaling conditions but *enhanced* it in the low-scaling conditions. These results suggest that the AP-initial L tone scaling had an important effect on listeners' speech segmentation in the presence of a co-occurring AP-final H tone, inhibiting performance when it was high and enhancing it when it was low.

In order to better understand the effects of the AP-final H tone alignment, we releveled the model in Table 3 twice, once with the high AP-initial L tone scaling condition as baseline and once with the low AP-initial L tone scaling as baseline. The releveled model with the high-scaling condition as baseline yielded significant simple effects of AP-final H tone alignment (later-aligned: $\beta = -0.722$, $SE = 0.083$, $z = -8.702$, $p < .001$; earlier-aligned: $\beta = -0.945$, $SE = 0.084$, $z = -11.270$, $p < .001$), with accuracy being *lower* in these conditions than in the control condition where tonal information was not present. These results confirm that the high scaling of the

Table 4. Results of logit mixed-effects model with AP-final H tone alignment, AP-initial L tone scaling, and their interaction as fixed effects (baseline: condition containing the later-aligned AP-final H tone with the mid AP-initial L tone scaling)

Fixed effects	Estimate	SE	z	p
(Intercept)	0.821	0.164	5.013	<.001
AP-final H tone (control)	–0.058	0.088	<|1|	>.1
AP-final H tone (earlier-aligned)	–0.252	0.087	–2.886	.004
AP-initial L tone (high scaling)	–0.993	0.209	–4.758	<.001
AP-initial L tone (low scaling)	0.727	0.214	3.396	.001
AP-final H tone (control) × AP-initial L tone (high scaling)	–0.804	0.130	–6.169	<.001
AP-final H tone (earlier-aligned) × AP-final L tone (high scaling)	0.234	0.134	1.741	.082
AP-final H tone (control) × AP-initial L tone (low scaling)	0.780	0.121	6.446	<.001
AP-final H tone (earlier-aligned) × AP-initial L tone (low scaling)	0.029	0.120	<|1|	>.1

Note: 11,664 observations, 108 participants, 18 words.

AP-initial L tone *inhibited* speech segmentation, and it did so both with the later-aligned and with the earlier-aligned AP-final H tone. The releveled model with the low-scaling condition as baseline revealed significant simple effects of AP-final H tone alignment (later-aligned: $\beta = 0.862$, $SE = 0.096$, $z = 8.959$, $p < .001$; earlier-aligned: $\beta = 0.844$, $SE = 0.095$, $z = 8.838$, $p < .001$), with accuracy being *higher* in these conditions than in the control condition. These results confirm that the low-scaling of the AP-initial L tone *enhanced* speech segmentation in the presence of a co-occurring AP-final H tone, and it did so for both AP-final H tone alignment conditions.

Table 4 presents the results of the model reported in Table 3 but with the condition containing the later-aligned AP-final H tone and the AP-initial L tone with a mid scaling as baseline. We focus on the fixed effects that compare later-aligned and earlier-aligned AP-final H tones, as the effects that compare the later-aligned AP-final H tone to the control condition were previously discussed.

As can be seen in Table 4, the simple effect of AP-final H tone alignment (earlier-aligned) was significant, with listeners' accuracy being significantly *lower* in the earlier-aligned AP-final H tone condition than in the later-aligned one when the AP-initial L tone had a mid scaling. The directionality of this effect is thus the opposite from that predicted at the onset of the study. The model also yielded significant effects of AP-initial L tone scaling (high scaling, low scaling), with listeners' accuracy in the condition with the later-aligned AP-final H tone being significantly lower when the AP-initial L tone had a high scaling and significantly higher when the AP-initial L tone had a low scaling. These results further confirm that in the presence of a later-aligned AP-final H tone, the high AP-initial L tone scaling hindered listeners' speech segmentation, whereas the low AP-initial L tone scaling enhanced it. The interaction between the AP-final H tone alignment (earlier-aligned) and the AP-initial L tone scaling (high scaling) was not significant. This lack of interaction means that the effect of AP-initial L tone scaling (mid > high)

generalizes both tone alignment conditions, and the effect of tone alignment (later-aligned > earlier aligned) generalizes both the mid and high tone scaling conditions. The interaction between the AP-final H tone alignment (earlier-aligned) and the AP-initial L tone scaling (low scaling) showed a trend toward approaching significance. The positive directionality of this trend suggests that the (negative) effect of AP-final H tone alignment (earlier-aligned) became more positive (i.e., decreased) in the low AP-initial L tone scaling condition compared to the mid-scaling condition. To better understand this trend, we releveled the model one more time, with the low AP-initial L scaling condition as baseline. This releveled modeled did not yield a significant effect of AP-final H tone alignment (earlier-aligned, $\beta = 0.018$, $SE = 0.102$, $z < |1|$, $p > .1$), indicating that the later-aligned and earlier-aligned AP-final H tone conditions did not differ when the AP-initial L tone had a low scaling. These results thus suggest that the earlier-aligned AP-final H tone inhibited listeners' speech segmentation performance compared to the later-aligned one except when the AP-initial L tone had a low scaling, in which case it had no effect on speech segmentation.

Discussion

The present study used an AL segmentation paradigm to investigate whether differences in the alignment of the AP-final H tone and in the scaling of the AP-initial L tone would modulate Korean listeners' speech segmentation, and whether they would do so independently. By answering these questions, this study would not only clarify how AP-final and AP-initial tonal cues affect Korean listeners' speech segmentation in relation to the intonational phonology of Korean but also shed light on the role of the fine-grained phonetic details of tonal cues on the outcome of speech segmentation.

The results showed that in the presence of an AP-final H tone, the lower the AP-initial L tone, the better Korean listeners' speech segmentation. 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 (2009). Using word-spotting experiments, Kim and Cho (2009) 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, the authors 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 that the AP-final and AP-initial tones interact and do not independently modulate Korean listeners' speech segmentation. The current results are in line with those of Kim and Cho (2009), with Korean listeners using the AP-final H tone as a cue to word boundaries only if the scaling of the AP-initial L tone is low enough to be consistent with the canonical F0 lowering in Korean. The results are also consistent with those of Kim et al. (2012), whose AL included both an AP-final H tone and an AP-initial L tone with a low scaling as cues to, respectively, the end of the preceding word and the beginning of the following word. Together, these findings suggest that it is the contrast between the H and L tones that signals the prosodic juncture, which in turn enhances speech segmentation for Korean listeners.

The results of the present study also indicate that a difference as small as 20 Hz in the onset of the word-initial tone (i.e., 1.58 semitones from the mid- to the high-scaling condition and 1.73 semitones from the mid- to the low-scaling conditions) has enough of an impact on listeners' speech segmentation to, respectively, inhibit and enhance speech segmentation. This suggests that fine-grained tonal cues to word boundaries affect listeners' speech segmentation in a gradient fashion.

The results of the present study also showed that when the AP-initial L tone is undershot (i.e., in the mid- and high-scaling conditions), an AP-final H tone that is aligned earlier in the syllable *inhibits* speech segmentation compared to an AP-final H tone that is aligned later in the syllable. This effect is contrary to what was predicted at the onset of the study: it was predicted that Korean listeners' speech segmentation might benefit from an earlier-aligned AP-final H tone compared to a later-aligned H tone because the earlier-aligned AP-final H tone is more consistent with the intonational system of Korean, as suggested by independent acoustic analyses of Korean (Jun, 2000) as well as our own recordings. Instead, the earlier-aligned AP-final H tone made it more difficult for listeners to locate word boundaries in the speech signal when the AP-initial L tone was not sufficiently low.

This alignment effect may be taken to be a low-level psychoacoustic effect that interacts with the phonological effect driven by the intonational grammar of Korean: when the AP-initial L tone was not canonical phonologically (i.e., when it was not realized as expected), Korean listeners appear to have resorted to low-level acoustic cues that can signal word boundaries across languages. Compared to the earlier-aligned AP-final H tone, the later-aligned H tone provided stronger low-level acoustic cues to word boundaries, in that it resulted in a steeper H-to-L pitch slope from the word-final syllable to the word-initial syllable. The degree of steepness in F0 rises and falls has been known to serve as a salient tonal feature to mark prominence across languages (e.g., Baumann & Winter, 2018; DiCanio, Benn, & García, 2018; Knight, 2008; Ots, 2017; Rietveld & Gussenhoven, 1985). It is therefore reasonable to assume that the steeper lowering created by the later-aligned AP-final H tone is perceptually salient and enhances the percept of the possible boundary. When the AP-initial L tone deviated from that stipulated by the intonational grammar of Korean, listeners tuned in to this low-level cue to locate word boundaries in the speech signal. This would explain not only why the effect of AP-final H tone alignment was contrary to predictions but also why it emerged only when the AP-initial L tone was undershot. The AP-initial L tone in the mid- and high-scaling conditions is likely to have been interpreted by Korean listeners as the interpolation between an AP-final H tone and an AP-medial L tone rather than as the AP-initial L tone itself. This both prevented listeners from being able to use the AP-final H tone to locate word-final boundaries and resulted in their use of low-level acoustic cues (e.g., steeper H-to-L pitch slope) to locate word-final boundaries. This is also reminiscent of what Kim and Cho (2009) observed: Korean listeners' lexical segmentation performance did not benefit from the universally driven phrase-final lengthening cue when the tonal contrast across the boundary was robust enough (with the H#L sequence), and they resorted to the durational cue only in the absence of the canonical tonal patterns. Most importantly, here, the perceptual salience created by the F0 steepness comes into effect only when it is consistent with the intonational phonology of Korean: it must be a falling tone in a direction to signal

the AP-initial L tone target. Thus, the observed effect is not a simple low-level au-ditory-phonetic effect, but it reflects the phonetic–prosody interface in the language, with the phonetic effect being fine-tuned by the higher order intonational phonol-ogy, as stipulated by the so-called phonetic grammar of the language (for a review, see Cho, 2015; Cho, 2016). This is again consistent with the gradient effects observed between the mid-scaling and the high-scaling conditions.

On the one hand, these results indicate that the effects of the AP-final H tone alignment and AP-initial L tone scaling are not independent: the AP-final H tone enhanced speech segmentation only when the AP-initial L tone was canoni-cal phonologically (see also Kim & Cho, 2009), and an effect of AP-final H tone alignment was more likely to emerge when the AP-initial L tone deviated from listeners' phonological expectations. On the other hand, the results suggest that the different alignments of the AP-final H tone in Korean and French is unlikely to have caused Korean listeners' speech segmentation difficulties in Tremblay et al. 2016)). Instead, Korean listeners' difficulty is much more likely to have stemmed from the higher scaling of the AP-initial L tone in French compared to Korean. These results highlight the gradient effects of incremental tonal changes in Korean listeners' speech segmentation, building on theories of lexical access, which postulate that lexical segmentation is modulated by the fine-grained phonetic details that arise due to phonological processes (e.g., Gow, 2002; Gow & Im, 2004) or to the phonetics–prosody interface (e.g., Cho et al., 2007; Salverda et al., 2003) or to both (e.g., Kim, Mitterer, & Cho, 2018). Cho et al. (2007), for example, suggested that listeners use the subphonemic (gradient) acoustic–phonetic information available in the speech signal to compute the prosodic structure of the current utterance, which is done by the so-called prosody analyzer, and in parallel use the same acoustic–phonetic information in the segmental analysis to create the lexical hypotheses. The locations of prosodic boundaries, detected by the prosody analyzer, then inform the lexical competition process, which is otherwise driven by the segmental information in the signal. The segmental analysis determines the content in the current input to be matched with a possible word, while the prosody analyzer indicates where words are likely to begin and end. The results of the present study enrich this theoretical consideration by providing a concrete case of how low-level acoustic details in the signal interact with the higher order intonational grammar of the language (Ladd, 2012), which may be processed by the prosody analyzer in lexical segmentation.

Finally, quite a few studies have explored how tonal information is aligned or "anchored" with the segmental string under the rubric of the segmental anchoring hypothesis (e.g., Arvaniti, Ladd, & Mennen, 2000; Atterer & Ladd, 2004; for a review, see Chapter 5 of Ladd, 2012) or segmental anchorage (Welby & Loevenbruck, 2006).[5] The tenet of these notions, setting aside the exact theoretical underpinnings, is that tonal targets such as the lowest F0 point and the F0 peak for L and H tones are aligned in principled ways with specific landmarks or regions in the segmental structure (such as the midpoint or the end of the vowel or the syllable) in a given language. Theories indicate that the same phonological tone–segment association may entail differential tone–segment alignment patterns in a language-specific way. For example, the F0 rising pattern for phonologically similar pitch accents in Germanic languages may differ across languages (e.g., English,

German, and Dutch): the F0 peak associated with nuclear pitch accents tends to be aligned earlier in (British) English than in Dutch (Ladd, Schepman, White, Quarmby, & Stackhouse, 2009); and the F0 peak for prenuclear pitch accents appears earlier in English and Dutch than in German (Atterer & Ladd, 2004). Such cross-linguistic differences in tonal alignment raise the question of whether and how language-specific fine-grained phonetic details in the realization of tonal features are exploited in lexical segmentation of both native and non-native languages. The results of the present study inform this question to some extent. Subtle phonetic differences in tonal alignment may not influence lexical segmentation in a phonologically viable context (i.e., when followed by a canonical L tone), showing some tolerance to variability. This may be consistent with segmental anchorage (e.g., Welby & Loevenbruck, 2006), which allows for variability within a specified region in the segmental structure. However, the results of the present study also indicate that listeners exploit subtle phonetic differences in tonal alignment when phonology does not come into play. Given that the cross-linguistic variation found across Germanic languages occurs for a similar phonological specification of pitch accent (e.g., L+H*), it will be worth exploring the degree to which cross-linguistic phonetic variation in tonal alignment is tolerated and the degree to which it is exploited in a language-specific way. The present study therefore calls for further cross-linguistic studies on how language-specific fine-grained phonetic details in the realization of tonal features may be exploited in the lexical segmentation of both native and the non-native languages.

Conclusion

The present study investigated whether differences in the alignment of the AP-final H tone and in the scaling of the AP-initial L tone would modulate Korean listeners' speech segmentation, thus shedding further light on the role of fine-grained tonal cues in speech segmentation. The results of an AL segmentation task suggested that Korean listeners' speech segmentation was modulated by both higher level phonological expectations (i.e., the expectation that the AP-initial L tone in Korean should have a low scaling) and low-level acoustic information (i.e., the steeper H-to-L pitch slope from the word-final syllable to the word-initial syllable in the later-aligned AP-final H tone compared to the earlier-aligned H tone), with listeners resorting to using low-level acoustic information only when higher level phonological expectations were not met. Further research should provide a refined investigation of the circumstances under which listeners treat tonal targets as no longer representative of those associated with the intonational grammar of the native language and begin relying on low-level acoustic information in relation to the high-order prosodic structure of the language.

Author ORCIDs. Annie Tremblay, 0000-0002-0748-172X

with the creation of stimuli, and to the three anonymous reviewers who, through their insightful comments, helped us strengthen the manuscript.

Notes

1. When the AP-initial segment is fortis or aspirated, the basic underlying tonal pattern is H(HL)H (Jun, 1998, 2000).

2. Another interesting difference between the two languages is that, unlike French, Korean does not have articles. Thus, APs more often begin with a noun in Korean than in French. This may result in Korean listeners being more reliant on the AP-initial L tone than French listeners (given that determiners are very helpful to segment speech into words; e.g., Yurovsky, Yu, & Smith, 2012). Note also that an F0 rise at the beginning of a content word (known as an optional "early rise") can serve as a cue to word beginnings in French (Welby, 2007).

3. The consonant /s/ was included as a syllable onset because otherwise it would not have been possible to create a sufficiently large number of Korean nonwords that could be used in the experiments. Of the three ALs created, only AL2 and AL3 had words that contained /s/. We ran a logit mixed-effects model on participants' accuracy in the conditions that contained tonal cues to word boundaries, with AL (AL1, AL2, AL3) as a fixed effect and with participant and word as crossed random effects (baseline = AL1). The model revealed no significant difference between listeners' accuracy on AL1, which did not contain /s/, and their accuracy on AL2 ($\beta = -0.187$, $SE = 0.148$, $z = -1.262$, $p > .1$) or AL3 ($\beta = 0.165$, $SE = 0.150$, $z = 1.103$, $p > .1$), both of which contained /s/. We can therefore conclude that the inclusion of /s/ in AL2 and AL3 did not adversely affect listeners' use of tonal cues.

4. The physical F0 falling at the end of the AP in the earlier-aligned condition is the by-product of the interpolation between the AP-final H tone target and the AP-initial L tone target in the following word. This interpolation is representative of Korean, as can be seen in the first (LHLH) AP of Example 9 in Jun (2000). As Ladd (2012) explains: "One key innovation of the AM [Autosegmental Metrical] theory is that it draws an explicit distinction between *events* and *transitions* [emphasis in original]. It recognises that certain localised pitch features are linguistically important, while much of the rest of the pitch contour is merely what happens between the important features" (p. 47).

5. As a reviewer pointed out, the term "segmental anchoring" often refers to specific segmental landmark points with which the beginning and the end of F0 movement are aligned, whereas the term "segmental anchorage" refers to a region, rather than a point, in the segmental string.

References

Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. In M. Broe, and J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 119–131). Cambridge: Cambridge University Press.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics*, **32**, 177–197.

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

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, **70**, 20–38.

Boersma, P., & Weenink, D. (2017). Doing phonetics by computer (Version 6.0.36). Retrieved from http://www.praat.org

Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 505–529). Hoboken, NJ: Wiley-Blackwell.

Cho, T. (2016). Prosodic boundary strengthening in the phonetics-prosody interface. *Language and Linguistics Compass*, **10**, 120–141.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, **29**, 155–190.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, **35**, 210–243.

Cunillera, T., Toro, J. M., Sebastian-Galles, N., & Rodriguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, **1123**, 168–178.

DiCanio, D., Benn, J., & García, C. (2018). The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics*, **68**.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, **108**, 477–499.

Gow, D. W., Jr. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, **28**, 163–179.

Gow, D. W., Jr., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, **51**, 279–296.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, **15**, 189–226.

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

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modeling and technology* (pp. 209–242). Dordrecht: Kluwer Academic Publishers.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, **14**, 147–172.

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**, 415–444.

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

Kim, S., Mitterer, H., & Cho, T. (2018). A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLOS ONE*, **13**, e0202912.

Knight, R.-A. (2008). The shape of nuclear falls and their effect on the perception of pitch and prominence: Peaks vs. plateaus. *Language and Speech*, **51**, 223–244.

Kuznetsova, A., Brockhoff, B., & Christensen, H. (2016). Tests in linear mixed effects models. Version 2.0.32. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Ladd, D. R. (2012). *Intonational phonology* Cambridge: Cambridge University Press.

Ladd, D. R., Schepman, A., White, L., Quarmby, L. M., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, **37**, 145–161.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, **96**, 143–182.

Ordin, M., & Nespor, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning*, **63**, 800–834.

Ordin, M., & Nespor, M. (2016). Native language influence in the segmentation of a novel language. *Language Learning and Development*, **12**, 461–481.

Ots, N. (2017). On the phrase-level function of F0 in Estonian. *Journal of Phonetics*, **65**, 77–93.

Paradigm Stimulus Presentation. (2007). Perception Research Systems. Retrieved from http://www.paradigmexperiments.com

Rietveld, A. M. C., & Gussenhoven, C. (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics*, **13**, 299–308.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, **44**, 493–515.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, **35**, 606–621.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90**, 51–89.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, **105**, 466–476.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, **54**, 1–32.

Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention Perception and Psychophysics*, **72**, 775–787.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, **48**, 233–254.

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastian-Galles, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language*, **64**, 171–180.

Toro, J. M., Sebastián-Gallés, N., & Mattys, S. L. (2009). The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology*, **21**, 786–800.

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., Namjoshi, J., Spinelli, E., Broersma, M., Cho, T., Kim, S., … Connell, K. (2017). Experience with a second language affects the use of fundamental frequency in speech segmentation. *PLOS ONE*, **12**, e0181709.

Tremblay, A., & Spinelli, E. (2014). Utilisation d'indices acoustico-phonétique dans la reconnaissance des mots en contexte de liaison. In C. Soum-Favaro, A. Coquillon, and J. P. Chevrot (Eds.), *La liaison: Approches contemporaines* (pp. 111–134). Berne: Lang.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*, **126**, 367–376.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 98–108.

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, **34**, 343–371.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, **49**, 28–48.

Welby, P., & Loevenbruck, H. (2006). Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics*, **18**, 74–124.

Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology*, **3**, 374.
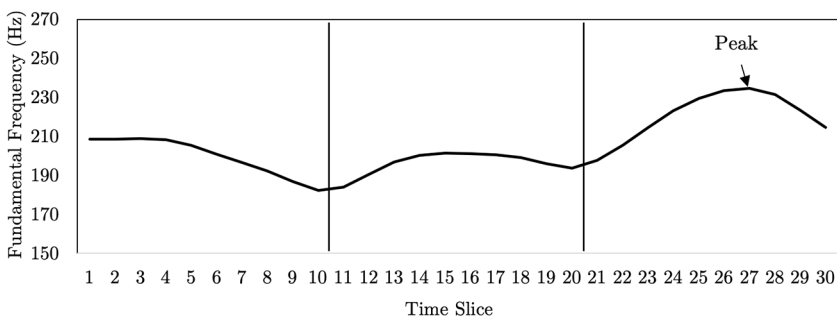
Appendix A

Figure A.1. Visual representation of the mean fundamental frequency (in Hertz) for each time slice in the prosodic contour recorded from a Korean speaker's production of Korean words.