FIRST AND SECOND LANGUAGE SIMILARITY CAN HURT THE LEARNING OF SECOND-LANGUAGE SPEECH SEGMENTATION: THE CASE OF PROSODY

Caitlin E. Coughlin(1), Annie Tremblay(1), Jiyoun Choi(2), & Mirjam Broersma(3)

1University of Kansas, 2Hanyang University, 3Radboud University Nijmegen
atrembla@ku.edu

ABSTRACT

This study investigates whether learning to use prosodic cues to word boundaries in second-language speech segmentation is easier or more difficult if the native and second languages have similar (though non-identical) prosodies than if they have markedly different prosodies. It compares French, Korean, and English listeners' use of fundamental-frequency rise and lengthening as cues to word-final boundaries in French. Fundamental-frequency rise and lengthening signal word-final boundaries in French and Korean but can signal word-initial boundaries in English. Proficiency-matched Korean-speaking and English-speaking second-language learners of French and native French listeners completed a 'visual-world' eye-tracking task where fundamental-frequency rise and/or lengthening signaled the final boundary of target words. Results show that the French and English groups used both fundamental-frequency rise and lengthening to locate word-final boundaries in French, whereas the Korean group used only lengthening. We attribute Korean listeners' non-use of fundamental-frequency rise in French to first- and second-language perceptual assimilation.

Keywords: speech segmentation, second language, prosody, French, Korean

1. INTRODUCTION

Segmenting continuous speech into individual words in the native language (L1) is accomplished in a seemingly effortless fashion. Conversely, finding word boundaries in a second language (L2) is much more difficult. One reason for this difficulty is that the cues to word boundaries that are efficient to segment the L1 may be inefficient or misleading to segment the L2. Non-native listeners have been shown to learn new segmentation routines derived from phonotactic information (e.g., [1]), but they have difficulty in learning to use prosodic cues (e.g., fundamental-frequency (F0) rise, lengthening) that signal different word boundaries in the L1 and L2 (e.g., [2]).

Previous research on L2 learners' use prosodic cues to word boundaries has focused on L1-L2 pairs that have different prosodic systems (e.g., Hungarian-English [2,3]). Unclear, however, is how the degree of similarity between the L1 and L2 prosodic systems affects the learning and use of prosodic cues in L2 speech segmentation.

The present study tests two competing hypotheses. The first hypothesis is that the learning of a new segmentation cue will be facilitated if the L1 and L2 prosodic systems are similar. The second hypothesis posits is the learning of a new segmentation cue will instead be very difficult if the L1 and L2 prosodic systems are similar. This second hypothesis stems in part from L2 speech perception theories that attribute non-native listeners' difficulty with similar L1-L2 sounds to perceptual assimilation [4,5]. For both hypotheses, similarity is operationalized as a given prosodic cue (e.g., F0 rise, lengthening) signaling the *same word boundary* in both the L1 and the L2; if a given prosodic cue signals *different word boundaries* in the L1 and in the L2, the two prosodic systems are considered different. We test these two hypotheses by examining French, Korean, and English listeners' use of F0 rise and lengthening as cues to word-final boundaries in French.

In French, prominence is phrasal, with the last non-reduced syllable of the Accentual Phrase (AP) receiving a pitch accent [6,7,8]. In non-utterance-final position, this AP-final syllable has an F0 rise and is lengthened [6,7,8]. Research suggests that native French listeners use these two prosodic cues to locate word-final boundaries in speech [9,10].

In Korean, prominence is also phrasal. In the Seoul dialect, the AP-final syllable has an F0 rise [11,12,13], and if this syllable is at the end of an Intonational Phrase, it is also lengthened [13]. Korean is thus similar to French in that word-final boundaries can be signaled with both F0 rise and lengthening. Importantly, however, the F0 rise in Korean differs in its alignment from that of French: Whereas the F0 rise in French peaks at the end of the AP-final syllable and falls in the following (word-initial) syllable, in Korean the F0 both peaks and falls within the AP-final syllable, such that the following (word-initial) syllable is already low (cf. [7], p. 163 vs. [13], p. 21). Like French listeners, Korean listeners have been shown to use both F0 rise and lengthening as cues to word-final boundaries [14,15].

By contrast, in English, prominence is lexical, and it is aligned with stressed syllables. Stress has a statistical tendency to be word-initial [16,17], and (accented) stressed syllables tend to be higher in F0 and longer than unstressed syllables [18,19]. Native English listeners parse stressed syllables [20,21] and syllables with higher F0 [22] as word-initial. Although lengthening provides a good cue to stress (and thus to word-initial boundaries) in English [18], it can also signal phrase-final lengthening [19]; native English listeners have indeed been found to use duration as a cue to word-final boundaries [22].

French and Korean are thus much more similar (though not identical) in their prosody than French and English are. If L1-L2 prosodic similarity enhances the learning and use of L2 prosodic cues, Korean L2 learners of French should outperform proficiency-matched English L2 learners of French in their use of F0 cues to word-final boundaries in French. By contrast, if L1-L2 prosodic similarity interferes with the learning and use of L2 prosodic cues, English L2 learners of French should outperform proficiency-matched Korean L2 learners of French in the use of F0 cues to word-final boundaries. Since lengthening can cue both word-initial and word-final boundaries in English, predictions are unclear for this cue.

We test these predictions using a 'visual-world' eye-tracking experiment.

2. METHOD

2.1. Participants

Sixteen native French listeners (mean age: 26.9, SD: 5.1), 16 Korean L2 learners of French (mean age: 23.3, SD: 8.2), and 16 (American) English L2 learners of French (mean age: 23.9, SD: 0.9) participated in this study. The Korean and English listeners were selected so that they would be matched in their French proficiency (assessed with a cloze, i.e., fill-in-the-blank, test) and French experience. Table 1 provides the mean proficiency score, age of first exposure to French, number of years of French instruction, number of months of immersion in a French-speaking environment, and percent weekly use of French reported by the L2 learners (standard deviations in parentheses).

Table 1: L2 learners' proficiency and language background information

	Korean	English
Proficiency (/45)	21.0 (5.7)	23.3 (8.2)
Age first exposure	18.8 (2.1)	16.8 (4.3)
Years instruction	5.7 (2.2)	6.2 (3.2)
Months immersion	14.3 (15.8)	14.0 (23.5)
Weekly Fr use (%)	12.5 (10.6)	13.7 (10.2)

One-way ANOVAs with L1 as a between-group variable did not reveal a significant difference between the Korean and English groups on any of the five measurements (p>.1).

2.2. Materials

The stimuli were the same as those used in [2]. Participants heard sentences that contained a monosyllabic target noun and a disyllabic adjective (e.g., *chat lépreux* 'leprous cat'), the first two syllables of which were segmentally ambiguous with a disyllabic lexical competitor (e.g., *chalet* 'cottage'). All noun-adjective sequences in the experimental items were in subject position. They were produced such that an (AP-final) pitch accent occurred either on the monosyllabic noun (e.g., *chat*) or on the second syllable of the disyllabic adjective (e.g., *lépreux*). As a result, the lexical competitor (e.g., *chalet*) either crossed an AP boundary (Across-AP condition) or was located within an AP (Within-AP condition).

In our naturally produced stimuli, the target word (e.g., *chat*) had an F0 rise and was lengthened in the Across-AP condition but not in the Within-AP condition. In order to investigate whether F0 rise and lengthening can independently be used to locate word-final boundaries, we resynthesized our stimuli such that the target word would have a flat F0 in the Across-AP condition but an F0 rise in the within-AP condition. We thus had four conditions: (i) Natural Across-AP condition (F0 rise, lengthening); (ii) Resynthesized Across-AP condition (no F0 rise, lengthening); (iii) Resynthesized Within-AP condition (F0 rise, no lengthening); and (iv) Natural Within-AP condition (no F0 rise, no lengthening).

The stimuli were recorded by a phonetically trained native French speaker from France. The recordings were normalized for intensity, and acoustic analyses were performed in Praat [23]. To resynthesize the stimuli, the first four syllables of the experimental items were each divided into 20 segments, and the average F0 value of each segment was extracted. The existing pitch points in each segment were then dragged vertically using the PSOLA function of Praat [23] so that they would approximate the value of the extracted average in the corresponding segment of the opposite prosodic condition. A stop Hann-band filter from 500 to 1000 Hertz with a smoothing of 100 Hertz was then applied to all the stimuli in the experiment. This filter did not adversely affect the quality of the segmental or prosodic information, and it was successful in masking occasional differences in acoustic quality between the natural and resynthesized conditions.

A total of 32 experimental items were used, each appearing in all four conditions and counterbalanced in four lists so that no participant heard a single item in more than one condition. These experimental items were intermixed with 69 filler items.

Target, competitor, and distracter words were presented orthographically in a visual display [24]. Experimental trials always included the target word (e.g., *chat*), the competitor word (e.g., *chagrin*), and two unrelated distracter words that overlapped in form (e.g., *prince* ‘prince’ and *principe* ‘principle’) to avoid anticipatory fixations to the target and competitor words. For the same reason, filler trials included disyllabic target words.

2.3. Procedures

Participants’ eye-movements were recorded with an EyeLink eye-tracker (SR Research) at a sampling rate of either 250 Hz or 1,000 Hz (depending on the testing location). An ASIO-compatible sound card was used on the presentation computer to ensure that the audio timing of stimuli would be synchronized with the recording of eye movements. For each trial, participants saw four orthographic words in a (non-displayed) 2 × 2 design on the screen for 4,000 ms. The four orthographic words then disappeared and a fixation cross appeared at the center of the screen for 500 ms. When the fixation cross disappeared, the words reappeared and the auditory stimulus was simultaneously heard over headphones. The trial ended when the participant clicked the word they heard. The test items were presented over four blocks to allow for breaks and recalibration.

3. ANALYSES AND RESULTS

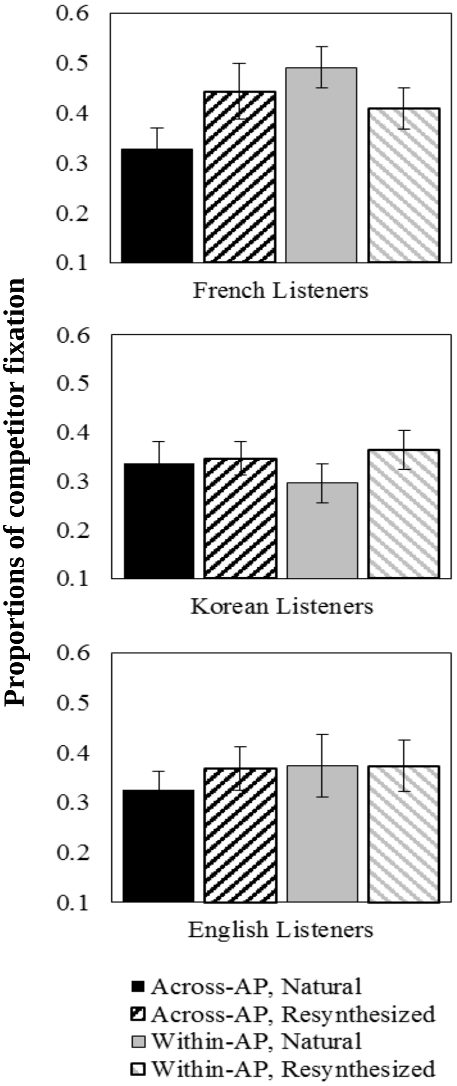
Only trials where participants clicked the target or competitor word were included in the analyses. This resulted in the loss of 5.8% of the data (French: 1.4%; Korean: 1.8%; English: 2.6%).

Because it takes approximately 200 ms to launch an eye-movement [25], proportions of eye fixation were analyzed starting at 200 ms after the onset of the target word. They were averaged for a time window that corresponded to the target noun (e.g., *chat*) and first syllable of the following adjective (e.g., *lé-*) (with the corresponding 200-ms delay in the offset of the time window), as these two syllables were segmentally ambiguous with the lexical competitor (e.g., *chalet*). At that point in time, the signal had not yet disambiguated (e.g., *-preux*, which is not a French word, had not yet been heard). Hence, this time window can determine whether listeners use prosodic cues to word boundaries prior to disambiguation in the signal. Since the disambiguation point differed from item to item, this averaging was done separately for each test item (and each participant).

Linear mixed-effects models were conducted on the log-odd-transformed proportions of competitor fixation using the lme4 package in R [26]. We first ran a big model on all three groups’ competitor fixations, with prosodic boundary (Within-AP, Across-AP), resynthesis (natural, resynthesized), and L1 as fixed variables. The baseline was French listeners’ fixations in the natural Across-AP condition. We also ran subsequent models separately for each group. For L2 learners, these subsequent models also had proficiency (centered cloze-test scores) and its interaction with prosodic boundary and resynthesis as fixed variables. In all models, participant and items were crossed random variables.

The participants’ proportions of fixation to the lexical competitor for the segmentally ambiguous time window are presented in Figure 1.

Figure 1: Proportions of competitor fixation in segmentally ambiguous time window



The first model revealed the following significant effects: prosodic boundary (t(1322)=2.98, p<.01); L1 for the Korean group (*t*(1322)=–2.45, *p*<.015) and the English group (*t*(1322)=–2.46, *p*<.015); prosodic boundary × resynthesis (*t*(1322)=–2.39, *p*<.017); prosodic boundary × L1 for the Korean group (*t*(1322)=–2.14, *p*<.04); and prosodic boundary × resynthesis × L1 for the Korean group (*t*(1322)=2.49, *p*<.02). The (positive) effect of prosodic boundary indicates higher proportions of competitor fixation for items without lengthening (Within-AP) than for items with lengthening (Across-AP). The (negative) effect of L1 indicates lower proportions of competitor fixation for the L2 groups than for the French group (likely due in part to L2 learners' slower processing).

A separate model on French listeners' proportions of competitor fixation revealed a significant effect of prosodic boundary (*t*(467)=3.06, *p*<.01) and a significant prosodic boundary × resynthesis interaction (*t*(467)=–2.49, *p*<.02). The (positive) effect of prosodic boundary indicates higher proportions of competitor fixation for items without lengthening (Within-AP) than for items with lengthening (Across-AP). The (negative) interaction between prosodic boundary and resynthesis indicates that adding an F0 rise reduced the proportion of competitor fixation for items without lengthening (Within-AP) but flattening F0 increased it for items with lengthening (Across-AP). This suggests that native French listeners are able to make use of both F0 rise and lengthening to segment speech into words.

A separate model on Korean listeners' proportions of competitor fixation revealed only a significant prosodic boundary × proficiency interaction (*t*(446)= 2.66, *p*<.01). This (positive) interaction indicates that Korean listeners show a greater effect of prosodic boundary (and thus lengthening) with increasing French proficiency. The lack of interactions with resynthesis indicates that Korean listeners did not make use of F0 rise to segment French speech. Thus, Korean listeners were able to use only lengthening to locate word-final boundaries in French.

Although none of the effects were modulated by L1 for English listeners, we ran a subsequent model on their proportions of competitor fixation to examine the effect of proficiency in relation to prosodic boundary and resynthesis. This model revealed a significant effect of prosodic boundary (*t*(403)=2.0, *p*<.05) and a marginally significant prosodic boundary × resynthesis × proficiency interaction (*t*(403)=–1.74, *p*<.08). The (positive) effect of prosodic boundary indicates higher proportions of competitor fixation for items without lengthening (Within-AP) than for items with lengthening (Across-AP). The (negative) three-way interaction indicates that English listeners show an increasingly strong interaction between prosodic boundary and resynthesis as their proficiency increases. This suggests that they can eventually use both F0 rise and lengthening to segment French speech.

4. DISCUSSION AND CONCLUSION

The results of our 'visual-world' eye-tracking experiment revealed that native French listeners made independent use of both F0 rise and lengthening to locate words in the segmentally ambiguous speech signal. Similarly, English L2 learners of French who were sufficiently advanced in French were also able to make independent use of both F0 rise and lengthening to locate word-final boundaries in French. Conversely, Korean L2 learners of French at sufficiently advanced proficiency could use only lengthening. These results are in line with the second hypothesis we tested, suggesting that greater similarity between the L1 and the L2 can in fact hurt the learning (and use) of speech segmentation cues.

We hypothesize that the F0 rise in French comes in too late for Korean listeners to be able to use it when segmenting French speech. Although F0 rise and lengthening signal AP-final (and thus word-final) boundaries in both French and Korean, in French the F0 peak aligns with the end of a syllable and the F0 falls on the following (word-initial) syllable, whereas in Korean the F0 peaks and falls within the AP-final syllable such that it is already low by the time the next (word-initial) syllable begins. This difference in F0-rise alignment between French and Korean may result in Korean listeners not using the F0 rise in time to segment words in French speech.

Critically, Korean L2 learners of French appear to have greater difficulty than proficiency-matched English L2 learners of French when adjusting the timing with which they expect the F0 rise to occur in French words. We hypothesize that this difficulty stems from perceptual assimilation processes similar to those proposed by existing L2 speech perception theories [4,5]. Specifically, Korean listeners may not hear the alignment of the F0 rise and word-final boundaries in French as different from that of Korean; consequently, they may fail to restructure their segmentation routines when learning to segment French speech. No such interference would occur with English listeners, because they would learn to associate the F0 rise with a new word boundary in French. Note, however, that given the non-categorical nature of prosody, the precise nature of the perceptual assimilation processes responsible for this difficulty would likely differ from those proposed in existing L2 speech perception theories [4,5]. Further research is needed to determine if Korean listeners indeed perceive the F0 rises in French and Korean as identical, and if so, how this perceptual assimilation differs from that with segments.

5. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant no. BCS-1423905 awarded to Dr. Annie Tremblay. Support for this research also comes from a *Language Learning* small research grant awarded to Dr. Annie Tremblay.

6. REFERENCES

[1] Weber, A., Cutler, A. (2006). First-language phonotac-tics in second-language listening. *Journal of the Acoustical Society of America, 119,* 597–607.

[2] Tremblay, A., Coughlin, C. E., Bahler, C., Gaillard, S. (2012). Differential contributions of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology, 3,* 385-423.

[3] White, L., Melhorn, J. F., Mattys, S. (2010). Segmentation by lexical subtraction in Hungarian speakers of second language English. *Quarterly Journal of Experimental Psychology, 63,* 544–554.

[4] Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Timonium, MD: York Press, 233–272.

[5] Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Timonium, MD: York Press, 167–200.

[6] Jun, S.-A., Fougeron, C. (2000). A phonological model of French intonation. In Antonis Botinis (Ed.), *Intonation: Analysis, modelling, and technology*. Dordrecht: Kluwer, 209–242.

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

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

[9] Christophe, A., Peperkamp, S., Pallier, C., Block, E., Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51,* 523–547.

[10] Michelas, A., D'Imperio, M. (2010). Accentual phrase boundaries and lexical access in French. In *Speech Prosody 2010.* http://speechprosody2010.illinois.edu/papers/100882.pdf.

[11] Jun, S.-A. (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean. In B. Connell, A. Arvaniti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV*. Cambridge, U.K.: Cambridge University Press, 235–253.

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

[13] Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions (version 3.1). http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html.

[14] 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.

[15] 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.

[16] Clopper, C. (2002). Frequency of stress patterns in English: A computational analysis. IULC Working Papers Online, 2. https://www.indiana.edu/~iulcwp/pdfs/02-clopper02.pdf.

[17] Cutler, A., Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133–142.

[18] Beckman, M. E., Jun, S.-A. (1996). K-ToBI (Korean ToBI) labeling conventions. Unpublished manuscript. Ohio State University & UCLA.

[19] Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America, 32,* 451–454.

[20] Cutler, A., Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

[21] McQueen, J. M., Norris, D., Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 621–638.

[22] 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.

[23] Boersma, P., Weenink, D. (2007). Praat: Doing phonetics by computer (Version 5.2.11) [computer program]. http://www.praat.org.

[24] McQueen, J. M., Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60,* 661–671.

[25] Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107.

[26] Baayen, H. D. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.