DIFFERENTIAL CONTRIBUTION OF PROSODIC CUES IN NATIVE AND NON-NATIVE SPEECH SEGMENTATION

Carly L. Bahler, Caitlin E. Coughlin & Annie Tremblay

University of Illinois at Urbana-Champaign, USA
atrembla@illinois.edu

ABSTRACT

The present study investigates the contribution of fundamental frequency (F0) in native English and native French listeners' segmentation of French speech. The results of a word-monitoring task with resynthesized stimuli show that pitch accents modulated speech segmentation for both groups, but unlike native listeners, the English listeners, who were at mid and high proficiencies in French, were not able to use F0 rise as a cue to word-final boundaries in French. These findings are attributed to their native language, in which F0 rise is a cue to word-initial boundaries, and to the co-occurrence of F0 and duration cues in word-final syllables in French, thus preventing them from attending to F0 rise as a cue to word-final boundaries.

Keywords: speech segmentation, prosodic cues, French, second/foreign language, lexical access

1. INTRODUCTION

An increasingly large body of evidence shows that native adult listeners use both accentual cues [1, 7, 11, 13] and higher-level prosodic cues [3] to segment speech into words, suggesting that prosodic information constrains lexical access. These prosodic cues have been proposed to be encoded either at the lexical level [8, 9] or at a more abstract level [6, 12]. By comparison, fewer studies have examined non-native listeners' use of prosodic cues in word recognition and speech segmentation [15, 17]. Doing so becomes particularly interesting when the native and target languages differ in their prosodic structure and in the primary acoustic cues to prosodic prominence.

In English, accented syllables tend to be word-initial [4]. Native English listeners use this probabilistic information to identify word-initial boundaries in continuous speech [7, 11, 13]. Because prosodic information is correlated with segmental information (e.g., vowel reduction), English listeners' reliance on suprasegmental information increases when the segmental quality of the speech stream is degraded [13]. Conversely, in French, accented syllables occur at the right edge of the Accentual Phrase (AP), and thus are word-final [10, 18]. Native French listeners indeed parse accented syllables as word-final syllables and use pitch accents to rule out phonemically overlapping lexical competitors (e.g., *chagrin* 'heartache' in *chat grincheux* 'cranky cat') [1, 3].

It is not surprising, then, that the prosodic cues to word boundaries differ in English and French. Using an artificial language learning paradigm, Tyler and Cutler [13] found that both English and French listeners use increased duration as a cue to word-final boundaries, but English listeners use F0 rise as a cue to word-initial boundaries whereas French listeners use F0 rise as a cue to word-final boundaries. These findings were attributed to the universality of increased duration as a cue to word-final boundaries and to the language specificity of F0 rise as a cue to word boundaries. It should be mentioned, however, that F0 rise can also mark the beginning of the AP in French [10, 18], and native French listeners use this phrase accent to identify word-initial boundaries [14]. Hence, English listeners must learn to distinguish phrase-final from phrase-initial F0 rises and use phrase-final F0 rise as a cue to word-final boundaries in French.

The present study investigates the ways in which native English listeners at mid and high proficiencies in French learn to use prosodic cues to word-final boundaries in French. The results of a previous study we conducted [5], adapted from Christophe, et al. [3], indicated that both mid- and high-level English L2 learners of French can use accented syllables to identify word-final boundaries in French, but correlational analyses suggested that the L2 learners and native listeners relied on, respectively, the duration and F0 of the accented syllable. In this study, we isolated F0 cues by resynthesizing the speech stimuli such that unaccented syllables would contain the F0 rise of accented syllables (and vice versa). We will thus examine the relative contribution of F0 as a cue to word-final boundaries in French.

2. METHOD

2.1. Participants

Twelve native French speakers (mean age: 29) and 24 native-English-speaking L2 learners of French (mean age: 22) participated in this study. The L2 learners were placed into two proficiency groups based on cloze (i.e., fill-in-the-blank) test scores [16]. The L2 learners' mean cloze test scores (out of 45), age of first exposure to French, years of instruction in/on French, months of residence in a French-speaking environment, and percent weekly use of French (and the corresponding standard errors) are provided in Table 1.

Table 1: Language background of the L2 learners.

	Mid	High
Cloze test scores	20.1 (1.3)	29.6 (1.1)
Age of first exposure to French	9.6 (1.3)	10.1 (0.7)
Years of instruction in/on French	7.6 (0.7)	9.4 (0.9)
Months of residence in French	1.9 (0.8)	10.8 (3.1)
Percent weekly use of French	7.0 (1.1)	15.8 (2.4)

2.2. Materials

The participants were asked to detect words that were not present in the sentence, but that were created phonemically between a monosyllabic noun and the syllable following it. In the first condition, the monosyllabic noun (e.g., *chat* 'cat') was accented, and thus the target word (e.g., *chalet* 'cottage') crossed an AP boundary (**natural across-AP condition**; e.g., *[Ce chat]₎ₐₚ [lépreux et légendaire]₎ₐₚ s'endort doucement* 'This leprous and legendary cat is slowly falling asleep'); in the second condition, the monosyllabic noun was not accented, and thus the target word was located within an AP (**natural within-AP condition**; e.g., *[Ce chat lépreux]₎ₐₚ s'endort doucement*). Two additional conditions were created by resynthesizing the speech stimuli using PSOLA [16] such that their first four syllables contained the F0 values of the opposite condition (**resynthesized across-AP condition**: across-AP sentence with the F0 of the within-AP condition; **resynthesized within-AP condition**: within-AP sentence with the F0 of the across-AP condition). The target words contained open syllables and/or syllables closed by /l/ or /ʁ/.

The task included 32 experimental items, with each participant hearing eight stimuli per condition. The participants were assigned to one of four lists and saw each experimental item in only one condition. The critical items were interspersed with

40 distracter items. The target words in the distracter items were located in different syntactic positions (e.g., subject, object). Half of all the test items contained the target word, and half did not.

The stimuli were recorded by a female native French speaker from Bordeaux (France). The speaker produced a pitch accent on the monosyllabic noun in the across-AP condition but on the last syllable of the post-nominal adjective in the within-AP conditions. The pitch accent on the monosyllabic noun in the across-AP condition was never followed by a pause, and no phrase accent was produced on the syllable following the pitch accent in the across-AP condition or on the monosyllabic noun in the within-AP condition.

The recordings were normalized for intensity, and acoustic analyses were performed in PRAAT [2]. To resynthesize the stimuli, the first four syllables of the experimental items were divided into 20 segments (each), and the average F0 value of each segment was extracted. The existing pitch points in each segment were then dragged vertically using PSOLA so that they would approximate the value of the extracted average in the corresponding segment of the opposite condition. After the resynthesis, a stop Hann-band filter from 500 to 1000 Hertz with a smoothing of 100 Hertz was applied to all the stimuli to mask the occasionally robotic sound that resulted from the F0 manipulation. Figure 1 shows an example of natural and resynthesized stimulus in the across-AP and within-AP conditions.

Figure 1: Example of stimulus (target: *chalet*).
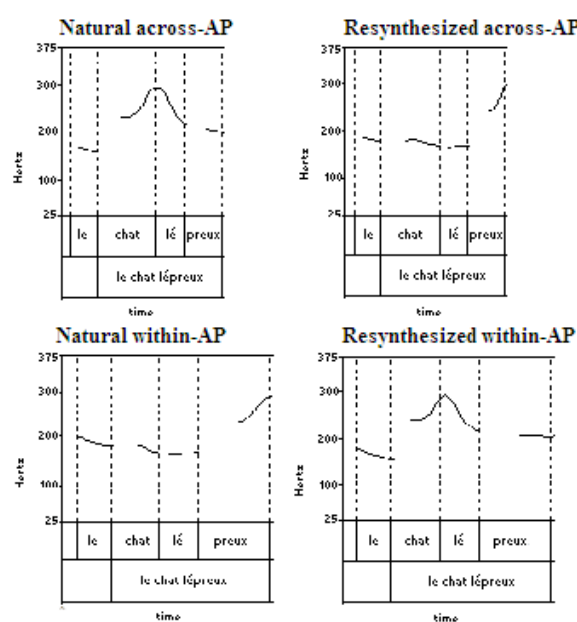
The mean F0 (in Hertz) and duration (in milliseconds) of the critical monosyllabic nouns in each condition (and the corresponding standard errors) are provided in Table 2.

Table 2: Acoustic analysis of the monosyllabic nouns.

Condition	Hz	Ms.
Natural across-AP	251 (6)	231 (10)
Resynthesized across-AP	189 (5)	231 (10)
Natural within-AP	189 (4)	187 (10)
Resynthesized within-AP	253 (7)	187 (10)

Repeated-measures ANOVAs, with prosody (across-AP, within-AP) and resynthesis (no, yes) as within-item variables, were performed on each acoustic measurement. They revealed: For F0, a significant prosody × resynthesis interaction, $F(1,31)=144.3$, $p<.001$ (other F's<1); and for duration, a significant effect of prosody, $F(1,31)=28.1$, $p<.001$ (other F's<1). Paired-samples t-tests were thus performed on the F0 values. The results, presented in Table 3, yielded effects of prosody and resynthesis on the F0 values of all pairs. These analyses indicate that both F0 and duration distinguished between the monosyllabic nouns in the natural across-AP and within-AP conditions, but only F0 distinguished between the natural and resynthesized nouns.

Table 3: Paired-samples t-tests on the F0 of the monosyllabic nouns (***$p<.001$).

|	Natural	Resynthesized	
	Across-AP	Across-AP	$t(31)=11.2$***
Hz	Within-AP	Within-AP	$t(31)=-11.3$***
	$t(31)=11.4$***	$t(31)=-9.5$***	

2.3. Procedure and predictions

The participants saw a target word, heard a sentence through headphones, and decided whether that sentence contained the target word ("o" for *oui* 'yes', no response for *non* 'no'). Their accuracy rates were measured. To increase the false alarm rates, the participants were put under time pressure: 14 distracter items containing the target word were followed by a *plus vite* 'faster' screen.

If the participants parse accented syllables as word-final, they should have fewer false alarms in the across-AP condition than in the within-AP condition. Furthermore, if they are sensitive to F0 rise as a cue to word-final boundaries, they should have more false alarms in the resynthesized across-AP condition than in the natural one, and fewer false alarms in the resynthesized within-AP condition than in the natural one.

3. RESULTS

Figure 2 presents the participants' proportion of false alarms in each condition.

Figure 2: Participants' false alarm rates.
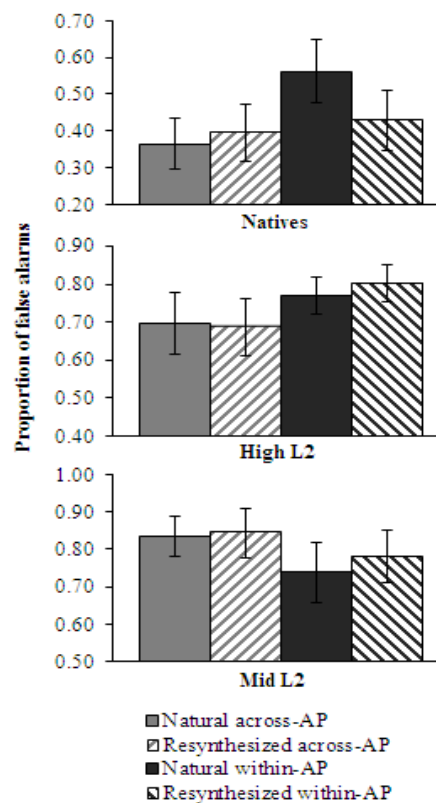
Linear mixed models on the native listeners' false alarm rates, with prosody (across-AP, within-AP) and resynthesis (no, yes) as fixed variables and with participant and item and random variables, revealed a significant effect of prosody, $F(1,369)=6.5$, $p<.011$, and a marginally significant interaction between prosody and resynthesis, $F(1,369)=3.4$, $p<.065$ (resynthesis: $F(1,369)=1.3$, $p<.248$). Separate linear mixed models performed on the across-AP and within-AP conditions revealed a significant effect of resynthesis only in the within-AP conditions, $F(1,179)=4.6$, $p<.034$. These results indicate that native listeners made fewer incorrect detections of the target word when it crossed an AP boundary then when it did not, and in the within-AP conditions, they made fewer incorrect detections of the target word when the stimuli had the F0 of the across-AP condition (where the monosyllabic word was accented) than when they had their natural (unaccented) prosody.

Linear mixed models on L2 learners' false alarm rates, with prosody, resynthesis, and proficiency (mid, high) as fixed variables and with

participant and item and random variables, revealed a significant interaction between prosody and proficiency, $F(1,738)=9.7$, $p<.002$ (all other F's<1). Separate linear mixed models performed on each group revealed significant effects of prosody for the mid-level L2 learners, $F(1,369)=4.8$, $p<.029$, and for the high-level L2 learners, $F(1,369)=4.9$, $p<.027$ (other F's<1), with the only the high-level L2 learners showing the effect in the predicted direction. These results indicate that unlike the native listeners and high-level L2 learners, the mid-level L2 learners made more incorrect detections of the target word when it crossed an AP boundary than when it did not, and neither L2 group was influenced by the resynthesis manipulation.

4. DISCUSSION AND CONCLUSION

The present results confirm that native French listeners identify word-final boundaries at the offset of accented syllables, and they can use (resynthesized) F0 rise as a cue to word-final boundaries. The flattening of the F0 rise in the resynthesized across-AP stimuli did not decrease the false alarm rates, probably because the duration of the accented syllable in these stimuli had not been altered, and thus provided a cue to word-final boundaries that native listeners were also able to exploit. These results are in line with those of Christophe, et al. [3], who showed that accented syllables mediate lexical access in French, and with those of Tyler and Cutler [17], who found that French listeners treat both duration and F0 as good cues to word-final boundaries.

The results also showed that as English listeners learn French, they go from parsing accented syllables as word-initial (mid-level) to parsing accented syllables as word-final (high-level), unlike the results of our previous study [5] and despite the comparability of the proficiency groups in the two studies. This development is not surprising, however, given that accented syllables provide a reliable cue to word-initial boundaries in English [7, 11, 13]. Yet, neither group was able to use F0 rise alone as a cue to word-final boundaries in French. Since the high-level L2 learners segmented words at the offset of accented syllables, our results suggest that they were able to use other cues to word-final boundaries (e.g., duration). The mid-level L2 learners' failure to do so suggests that English listeners at lower proficiencies may favor F0 as a cue to word-initial boundaries over

duration as a cue to word-final boundaries. Given the reliability of duration as a cue to word-final boundaries in French, we propose that English listeners may not ever attend to F0 as a cue to word-final boundaries, thus parsing the speech signal in a non-native-like manner.

5. REFERENCES

[1] Banel, M.H., Bacri, N. 1994. On metrical patterns and lexical parsing in French. *Speech Comm.* 15, 115-126.

[2] Boersma, P., Weenink, D. 2004. Praat: Doing phonetics by computer [Computer program]. Available at *http://www.praat.org*.

[3] Christophe, A., Peperkamp, S., Pallier, C., Block, E., Mehler, J. 2004. Phonological phrase boundaries constrain lexical access: I. Adult data. *J. Mem. Lg.* 51, 523-547.

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

[5] Coughlin, C.E., Tremblay, A. In press. The role of prosodic information in L2 speech segmentation. *Proceedings of the 35th BUCLD* (Supplement).

[6] Cutler, A. 2010. Abstraction-based efficiency in the lexicon. *J. Lab. Phon.* 1, 301-318.

[7] Cutler, A., Butterfield, S. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *J. Mem. Lg.* 31, 218-236.

[8] Goldinger, S.D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Review* 105, 251-279.

[9] Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In Johnson, K., Mullennix, J.W. (eds.), *Talker Variability in Speech Processing*. San Diego: Academic, 145-165.

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

[11] Mattys, S.L. 2004. Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *J. Exp. Psych.: Hum. Perc. Perf.* 30, 397-408.

[12] McQueen, J.M., Cutler, A., Norris, D. 2006. Phonological abstraction in the mental lexicon. *Cog. Science* 30, 1113-1126.

[13] McQueen, J.M., Norris, D., Cutler, A. 1994. Competition in spoken word recognition: Spotting words in other words. *J. Exp. Psych.: Learn. Mem. Cog.* 20, 621-638.

[14] Spinelli, E., Grimaud, N., Meunier, F., Welby, P. 2010. An intonational cue to segmentation in phonemically identical sequences. *Att. Perc. Psych.* 72, 775-787.

[15] Tremblay, A. 2008. Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Ap. Psych.* 29, 553-584.

[16] Tremblay, A. In press. Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition*.

[17] Tyler, M.D., Cutler, A. 2009. Cross-language differences in cue use for speech segmentation. *J. Acoust. Soc. Am.* 126, 367-376.

[18] Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *J. Phon.* 34, 343-371.