The Role of Prosodic Information in L2 Speech Segmentation

Caitlin E. Coughlin
Annie Tremblay
University of Illinois at Urbana-Champaign

1. Introduction

Unlike written language, where word boundaries are often denoted by blank spaces (e.g., *le_chat* 'the_cat'), for spoken language, no single device allows for the reliable identification of word boundaries: words are typically uttered without a pause between them, and sound processes further blur word boundaries. A crucial challenge for second/foreign language (L2) learners is that the cues to word boundaries differ across languages. Thus, an English speaker's experience with her native language may prove misleading when attempting to segment speech in a new language such as French. L2 learners often can identify words in writing or when spoken in isolation, and yet they may fail to recognize them in continuous speech. To segment language into words, non-native listeners must know not only the word forms uttered by the speaker, but also the sound processes that apply at word boundaries and the factors regulating the application of these processes.

One type of sound process that has been shown to play an important role in speech segmentation is prosodic prominence. For example, from the age of 7.5 months, English-acquiring infants use accented syllables to identify word-initial boundaries (e.g., Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999). Young infants are also sensitive to higher-level prosodic information: in English, they can detect disruptions in intonational phrases at 4.5 months (e.g., Hirsh-Pasek et al., 1987; Jusczyk et al., 1992) and disruptions in phonological phrases at 9 months (e.g., Kemler-Nelson et al., 1989; Gerken, Jusczyk, & Mandel, 1994); and in French, they can detect phonological phrase boundaries at 13 months (Christophe et al., 2003). Adults have similarly been shown to use both accentual cues (e.g., in English: Cooper, Cutler, & Wales, 2002; Cutler & Butterfield, 1992; McQueen, Norris, & Cutler, 1994; Mattys, 2004; in French: Banel & Bacri, 1994) and higher-level prosodic cues (e.g., in French: Christophe et al., 2004; Welby, 2006; in Korean: Kim, 2004; Kim &

*We are very grateful to Stéphanie Gaillard for her tremendous contribution to this study. Many thanks also go to Sarah Brown-Schmidt, Zsuzsanna Fagyal, Scott Frauendorf, Angie Isaacs, Tuan Lam, Eun-Kyung Lee, Molly Lewis, Nathan Owens, Rachel Ryskin, Kristin Tooley, Alison Trude, and Duane Watson for their valuable comments on this work.

 Proceedings of the 35th Boston University Conference on Language Development (Supplement). Somerville, MA: Cascadilla Press.

Cho, 2009; in Japanese: Warner, Otake, & Arai, 2010) to segment speech into words.

By comparison, few studies have examined non-native listeners' use of prosodic information in word recognition and speech segmentation (e.g., Cooper et al., 2002; Kim, Broersma, & Cho, to appear; Sanders, Neville, & Woldorff, 2002; Tremblay, 2008; Tyler & Cutler, 2009; White, Melhorn, & Mattys, 2010). Doing so becomes particularly interesting when the native and target languages differ not only in their prosodic structure, but also in the primary acoustic cues associated with prosodic prominence. In such cases, L2 learners must establish the correct mapping between these cues and prosodic prominence before they can hear prominent syllables as such and use this information to recognize words in continuous speech. If non-target-like, this mapping may potentially impede L2 learners from making higher-level prosodic generalizations (e.g., Tremblay & Owens, 2010; for discussion, see Carroll, 2004).

The present study investigates how English speakers at mid and high proficiencies in French learn to use prosodic information to segment French speech into words. French and English differ both in their prosodic structure and in the acoustic cues associated with prosodic prominence. This constellation of native and target languages thus provides a crucial window into the prosodic cues that L2 learners extract from the speech signal (if any) and their use of these cues to recognize words in continuous speech. Before presenting our study, we give an overview of the prosodic structure of French and the cues it provides for locating word boundaries; we then review research on the use of prosodic information in the segmentation of French and English, and make predictions for the L2 learners in the present study.

2. Prosodic Structure of French

In their autosegmental-metrical account of French prosody, Jun and Fougeron (2002) propose that the domain of prominence in French is the Accentual Phrase (AP), which corresponds roughly to Verluyten's (1982) Accentual Group and Nespor and Vogel's (1986) Phonological Phrase. Each AP contains one accented syllable at its right edge, as shown in (1), where the asterisk (*) represents a pitch accent; this accented syllable is realized as a low-high (LH) rising tone, except in sentence-final position of declarative sentences, where only the L tone surfaces before a falling Intonational Phrase (IP) boundary (L%). In neutral (i.e., non-contrastive) prosodies, the accented syllable (H*) is predictably aligned with the last non-schwa syllable of the last word in the AP; it thus coincides with a word-final boundary, but functions as an edge tone (see also Welby, 2006). The acoustic correlates of the LH pitch accent (i.e., in non-sentence-final position) include both an increase in fundamental frequency (F0) and an increase in duration (Welby, 2006).

APs can also have a rising phrase accent (LHi) at their left edge, also illustrated in (1). The rise, when present, typically occurs on content words rather than on function words. This phrase accent is structurally different from

the pitch accent, in that it is optional (it is not realized in short APs), and the location of its rise (Hi), if realized, varies as a function of the length of the AP and of the content words inside it; when present, either of its tone delineates the left edge of the AP, and thus tends to be aligned with word-initial syllables. Unlike the pitch accent, which is associated with the word, the phrase accent, as its name suggests, is a property of the phrase. Since it is not anchored to a particular syllable, its primary acoustic correlate is an increase in F0 (Welby, 2006).

(1)
L (Hi) L H* L Hi L L%
[[Nous aimons tous]$_{AP}$ [le chocolat noir]$_{AP}$]$_{IP}$.
'We all like dark chocolate.'

Given their alignment with (respectively) the right and left edges of APs, both pitch and phrase accents provide cues to word boundaries in French. Let us now turn to studies showing that French listeners indeed exploit these cues when recognizing words in continuous speech.

3. Use of Prosodic Cues in the Segmentation of French and English Speech

Because pitch accents predictably fall on the last syllable of APs in French, they are reliable cues to word-final boundaries. A number of studies have indeed shown that native French listeners use pitch accents to segment speech into words. For example, Banel and Bacri (1994) found that French speakers who listened to phonemically ambiguous sequences (e.g., /bagaʒ/) were more likely to hear a single disyllabic word (e.g., *bagage* 'luggage') if the duration of the second syllable had been increased, and they were more likely to hear two monosyllabic words (e.g., *bas gage* 'low pledge') if the duration of the first syllable had been increased, as it would be if a pitch accent was located on, respectively, the second or first syllable. Bagou, Fougeron, and Frauenfelder (2002) similarly showed that French listeners used both increased duration and F0 rise to segment an artificial language into words, with the later yielding slightly more accurate segmentation than the former (see also Bagou & Fraudenfelder, 2006).

In a series of online experiments, Christophe et al. (2004) provided further evidence that phrase-final prosodic boundaries (and pitch accents) mediate lexical access in French. They found that monosyllabic words such as *chat* 'cat' were recognized more slowly when they were temporarily ambiguous with a competitor word created phonemically between the monosyllabic word and the first syllable of the word following it (e.g., *chagrin* /ʃagʁɛ̃/ 'heartache' in ... *[d'un chat grincheux]$_{AP}$* ... '... of a cranky cat ...') than when they were not temporarily ambiguous with such a competitor (e.g., ... *[d'un chat drogué]$_{AP}$* ... '... of a drugged cat ...'); however, if the monosyllabic word was at a prosodic boundary and thus received a pitch accent (e.g., ... *[le gros chat]$_{AP}$ [grimpait aux arbres]$_{AP}$* '... the big cat was climbing trees'), the target word was no longer

recognized more slowly when it was temporarily ambiguous with a phonemic competitor than when it was not (e.g., ... *[le gros chat]$_{AP}$ [dressait l'oreille]$_{AP}$* 'the big cat was sticking up his ears'). These findings suggest that phrase-final boundaries, marked with a pitch accent, acted as filter by constraining lexical access.

Recent research has shown that native French listeners can also use phrase accents to identify word-initial boundaries in continuous speech. Spinelli, Welby, and Shaegis (2007) found that French listeners could discriminate between and identify words such as *l'affiche* 'the poster' and *la fiche* 'the sheet': because the rise (Hi) in the phrase accent is typically aligned with the first syllable of content words, it occurs on the syllable /la/ in *l'affiche* but on the syllable /fiʃ/ in *la fiche*. French listeners could use this acoustic information (among others) not only in offline tasks, but also in online ones, with target words being activated more when they matched the intended segmentation than when they did not. In a follow-up study, Spinelli et al. (2010) demonstrated that raising the F0 of /la/ in *la fiche* resulted in the greater selection and easier recognition of vowel-initial words (e.g., *affiche*) than if the F0 had not been manipulated. This suggests that phrase-initial prosodic information has an immediate effect on French listeners' word recognition.

Not all studies have shown such an effect, however. Bagou and Frauenfelder (2006) report that French listeners exposed to an artificial language benefit from phrase-initial prominence only when phrase-final prominence is also present. This discrepancy with the previous results may suggest that pitch accents are more reliable cues to word boundaries than phrase accents in French.

English speakers also use prosodic information to recognize words in continuous speech. Pitch accents in English are generally aligned with stressed syllables (e.g., Beckman & Elam, 1997), which, statistically, tend to be word-initial rather than word-final, especially in disyllabic and trisyllabic nouns (e.g., Clopper, 2002). Hence, accented syllables in English can provide a reliable cue to word-initial boundaries. Yet, because prosodic information is highly redundant with segmental information (in particular, vowel reduction) in the language, it may play a more important role in word recognition when lexical information is less available.

In a juncture perception task where stimuli were barely audible, Cutler and Butterfield (1992) found that English listeners tend to hear word-initial boundaries at the onset of stressed syllables (e.g., *achieve her way instead* was perceived as *a cheaper way to stay*). Similar findings were reported in online studies: McQueen et al. (1994) showed that English words are more easily detected when they are the second syllable of nonsense iambic sequences (e.g., *mess* in /nəmés/) than when they are the first syllable of nonsense trochaic sequences (e.g., *mess* in /méstəm/), because a word-initial boundary can be detected at the onset of the stressed syllable in the former, but no word-final boundary is detected at the offset of the stressed syllable in the latter. Mattys (2004) also found that when stimuli are presented with background noise, stress-initial disyllabic primes embedded in quadrisyllabic nonsense sequences

(e.g., /díplənəútə/) facilitated the activation of matching trisyllabic targets (e.g., *notable*) more than did stress-final disyllabic embedded primes (e.g., /dípləmɪkǽ/) for matching trisyllabic targets (e.g., *mechanic*). These findings suggest that prosodic information may be particularly useful for word segmentation in English when lexical information is degraded or absent (see also Mattys, White, & Melhorn, 2005).

The primary acoustic correlates of prosodic prominence in English are duration, F0, and amplitude (e.g., Liberman, 1960; Beckman, 1986), and the importance of each correlate depends in part on the location of the stressed syllable in the word, with F0 being a strong cue to word-initial stress but with duration being a stronger cue to stress in non-word-initial positions (Tremblay & Owens, 2010). The close relationship between prosodic information and vowel reduction in English further confirms the importance of duration as a cue to prominence. In French, on the other hand, both F0 and duration are strong acoustic correlates of pitch accents in word-final position.

Using artificial language-learning experiments, Tyler and Cutler (2009) found that both English and French listeners used duration as a cue to word-final boundaries. They attributed these findings to the universality of duration as a cue to word-final boundaries across languages rather than to the relationship between prosodic information and vowel reduction in English (see also Hayes, 1995; Saffran, Newport, & Aslin, 1996; Vaissière, 1983). Tyler and Cutler also found, however, that only French listeners used F0 rises as a cue to word-final boundaries; English listeners instead used F0 rises as a cue to word-initial boundaries. Since the alignment of F0 rises with word edges varies cross-linguistically, L2 learners must learn the particular prosodic configuration of the target language in order to use this information successfully in speech segmentation. For the present study, this means that English L2 learners of French must not only learn to parse accented syllables as word-final (rather than word-initial), but also to use F0 as a cue to word-final boundaries. This can potentially be difficult given that the F0 rise in phrase accents, when present, can also provide a cue to word-initial boundaries in French. Hence, English L2 learners of French must also distinguish phrase accents from pitch accents and use them in a target-like fashion to segment the speech stream into words.

Few studies have examined non-native listeners' use of prosodic information in word recognition and speech segmentation. Cooper et al. (2002) showed that when hearing phonemically identical stressed and unstressed fragments (e.g., /kǽm/ vs. /kæm/), Dutch L2 learners of English were in fact better than native English listeners at identifying the word to which the fragment belonged (e.g., respectively, *campus* and *campaign*). These findings were attributed to the fact that prosodic information is less correlated with segmental information in Dutch than in English. Using a similar task, Tremblay (2008) found that French L2 learners of English had much more difficulty than native English listeners in identifying the word to which the fragment belonged. While prosodic information can signal word boundaries in French, for words that begin (or end) with phonemically identical syllables, it does not distinguish between

different lexical competitors, thus making it difficult for French L2 learners of English to use this information in lexical access.

Non-native listeners can use prosodic information to segment speech into words, but like native listeners, their reliance on this information depends on the extent to which lexical information is available. Sanders et al. (2002) showed that Spanish and Japanese L2 learners of English had less difficulty detecting phonemes when they occurred in the onset of word-initial stressed syllables than when they occurred in the onset of word-medial stressed syllables. Importantly, this effect was larger when lexical information was missing from the stimuli then when it was present in the stimuli. On the other hand, White et al. (2010) found that Hungarian L2 learners of English recognized trisyllabic words (e.g., *corridor*) more rapidly when the disyllabic prime they heard was preceded by a lexical word (e.g., *anythingcorri*) than when it was preceded by a non-word (e.g., *imoshingcorri*), irrespective of whether the target word and prime were stressed on the first or second syllable (e.g., *confusing* for *anythingconfu* and *imoshingconfu*). These L2 learners' non-reliance on prosodic information, which did not vary as a function of proficiency in English, is somewhat surprising given that stress is word-initial in Hungarian and thus provides an excellent cue to word-initial boundaries.

The present study investigates English listeners' use of prosodic information in the recognition of French words. It uses an experimental paradigm adapted from Christophe et al. (2004), in which the participants were asked to detect words that were not present in the sentence, but that were created phonemically between a monosyllabic word and the syllable following it (e.g., *chagrin* 'heartache' in *chat grincheux* 'cranky cat').[1] In one condition, the monosyllabic word received a pitch accent, and thus the target word crossed an AP boundary (across-AP condition); in the other condition, the monosyllabic word did not receive a pitch accent, and thus the target word was located within an AP (within-AP condition). Unlike some of the previous studies, lexical information was not degraded, but the participants were put under time pressure so as to see whether prosodic information would further help them segment the speech stream into words. Since pitch accents fall on word-final syllables in French and phrase-final boundaries are aligned with word-final boundaries, if the participants use prosodic information, they should make fewer incorrect detections of the target word in the across-AP condition than in the within-AP condition. Given that the experiment used natural stimuli, the acoustic cues to pitch accents (and AP-final boundaries) were not manipulated explicitly, but the relationship between these cues in the stimuli and the participants' proportion of false alarms were examined.

[1] The present design was favored over one in which the participants would be asked to detect the monosyllabic word in the sentence (e.g., *chat*), because it is very difficult to assess whether non-native listeners have particular competitor words (e.g., *chagrin*) in their lexicon.

4. Method

4.1 Participants

Twenty-eight native English speakers (age 18–31, mean (M): 23.4, standard deviation (SD): 4.1) at mid and high proficiencies in French and 11 native French listeners (age 23–33, M: 27.2, SD: 3.8) participated in this study. They were undergraduate and graduate students at a Midwestern University. The participants had normal or corrected-to-normal vision and did not report hearing impairment. They received financial compensation in return for their participation.

The L2 learners had completed at least four semesters of French at the time of the study, and most of them had little exposure to French before the onset of puberty. Their proficiency in French was identified with the help of a cloze (i.e., fill-in-the-blank) test independently shown to provide a reliable estimate of proficiency in French (Tremblay, to appear; Tremblay & Garrison, 2010). The participants were evenly divided into two proficiency groups on the basis of their cloze test scores.

The participants also completed a language background questionnaire in which they specified relevant biographical information. For L2 learners, this information included their age of first exposure to French, their number of years of instruction in/on French, the number of months they spent in a French-speaking environment, and their percent weekly use of French. The L2 learners' cloze test scores and their biographical information are provided for each proficiency group in Table 1.

Table 1. L2 Learners' Cloze Test Scores and Biographical Information

Group	Cloze (/45)	Age of First Exposure	Years of Instruction	Months of Immersion	% Use
Mid (n=14)	19.4 (3.4)	11.6 (2.6)	7.5 (2.0)	0.3 (0.4)	8.0 (2.6)
High (n=14)	32.7 (3.1)	12.5 (3.8)	11.4 (4.0)	15.4 (11.1)	27.7 (13.6)

4.2 Materials

The participants heard sentences in which a target word was created phonemically between a monosyllabic word and the first syllable of the adjective following it (e.g., *chagrin* 'heartache' in *chat grincheux* 'cranky cat'). In the across-AP condition, the monosyllabic word received a pitch accent, and the target word crossed an AP boundary (e.g., *[Ce chat]$_{AP}$ [grincheux et bedonnant]$_{AP}$...* 'this cranky and chubby cat ...'); in the within-AP condition, the pitch accent instead fell on the last syllable of the post-nominal adjective (e.g., *[Ce chat grincheux]$_{AP}$* ...). The sentences in the two conditions shared the first three words (article, monosyllabic word, and adjective). The two experimental conditions thus compared words that differed only in their prosody. In the across-AP condition, an additional modifier was added in the second AP so that the prosodic boundary after the monosyllabic word would

sound natural. The experiment also included a control condition in which the target word was in the sentence; prosodically, this condition was identical to the within-AP condition (e.g., *[Ce chagrin fou]$_{AP}$* ... 'this crazy heartache'). The target words in the experimental and control sentences were all the subject of the sentence.

The experiment included a total of 36 critical triplets. The participants were assigned to one of three lists and saw each experimental item in only one condition (total: 12 items per condition). A complete list of the noun-adjective (-adjective) sequences in the experimental items is provided in the Appendix. The experimental items were pseudo-randomized with 72 distracter items, of which 10 were practice items. The target words in the distracter items were located in different syntactic positions (e.g., subject, object). Half of the test items in the experiment contained the target word, and half did not.

The auditory stimuli were recorded by a female native speaker of French from Bordeaux (France) using a Marantz PMD 750 solid state recorder and head-mounted condenser microphone. The speaker was trained to produce the stimuli such that a pitch accent would fall on the monosyllabic noun in the across-AP condition but on the last syllable of the post-nominal adjective in the within-AP and control conditions. The pitch accent produced on the monosyllabic noun in the across-AP condition was not followed by a pause so that the disyllabic target word could accidentally be detected.

The recordings were then normalized for intensity, and acoustic analyses of the monosyllabic noun were performed in PRAAT (Boersma & Weenink, 2007). The average F0, duration, and intensity of the monosyllabic nouns in the experimental and control conditions are provided in Table 2.

Table 2. Acoustic Analysis of Monosyllabic Nouns in the Stimuli

	F0 (Hz)	Duration (ms.)	Intensity (dB)
Across-AP	249 (5)	231 (9)	66.4 (0.2)
Within-AP	189 (4)	192 (5)	65.8 (0.2)
Control	191 (2)	177 (5)	65.9 (0.4)

Note. Mean (standard deviation)

As can be seen from the acoustic measurements, the monosyllabic words had higher F0, longer duration, and higher intensity in the across-AP condition than in the within-AP and control ones. Paired-samples *t*-tests performed on the F0 values in the three conditions, with the alpha-level adjusted to .0167, revealed significant differences between the across-AP condition and both the within-AP and control conditions (respectively: $t(35)=11.876$, $p<.001$; $t(35)=13.868$, $p<.001$), but not between the within-AP and control conditions ($t<|1|$). Similar paired-samples *t*-tests conducted on the duration values yielded significant differences between the three conditions (across-AP-within-AP: $t(35)=4.722$, $p<.001$; across-AP-control: $t(35)=6.026$, $p< 001$; within-AP-control: 4.481, $p<.001$). A last set of paired-samples *t*-tests performed on the intensity values revealed significant differences between the across-AP and

within-AP conditions ($t(35)=2.579$, $p<.014$), but not between the control condition and either the across-AP condition or the within-AP condition (respectively: $t(35)=1.304$, $p<.201$; $t<|1|$). These results indicate that the monosyllabic words in the across-AP condition had significantly higher F0, longer duration, and higher intensity than those in the within-AP condition, and the monosyllabic words in the within-AP condition had similar a F0 and amplitude but a longer duration than the first syllable of the disyllabic words in the control condition.

Another prosodic cue that could influence the participants' false alarm rates is the presence of a phrase accent (LHi) at the left edge of APs, here potentially on the first syllable of the adjective in the across-AP condition and on the monosyllabic noun in the within-AP condition. The intonational patterns of the stimuli were thus examined closely to determine whether phrase accents had unintentionally been produced. For the stimuli in the across-AP condition, this inspection revealed a steep F0 fall on the first syllable of the adjective, and no evidence of a second rise in the adjective. Phrase accents in French typically begin with an L tone, and the fall following the Hi tone is usually gradual rather than steep (Jun & Fougeron, 2002). This suggests that this F0 fall on the first syllable of the adjective in the across-AP condition was due to the pitch accent on the previous syllable rather than to a phrase accent. For the stimuli in the within-AP condition, the F0 in the monosyllabic noun (M: 191 Hz, SD: 27) was similar to that of the article preceding it (M: 193 Hz, SD: 19), and it was slightly higher than the F0 in the first syllable of the adjective (M: 184 Hz, SD: 19), with no significant difference found between the monosyllabic word and either the determiner ($t < |1|$) or the first syllable of the adjective ($t(35)=1.451$, $p<.156$). This suggests that the monosyllabic word in the within-AP condition did not contain a phrase accent. We are therefore confident that the main cue to prosodic boundaries in our stimuli is the pitch accent on the monosyllabic noun in the across-AP condition.

The present experimental design can potentially introduce a confounded variable, that of speech rate: the presence of a pitch accent on the monosyllabic noun in the across-AP condition can result in slower speech rate, potentially leading the participants to have fewer false alarms in the across-AP conditions than in the within-AP one. To determine whether this was indeed the case, we measured the duration of the noun-adjective sequences in the across-AP and within-AP conditions. These additional analyses revealed that the noun-adjective sequences were in fact shorter in the across-AP condition (M: 667 ms., SD: 14) than in the within-AP condition (M: 701 ms., SD: 13), a difference which is statistically significant ($t(35)= –2.713$, $p<.010$). Hence, if speech rate influenced the results, it would likely be in the opposite direction to what is predicted from the experimental manipulation.

4.3 Procedure

The experiment was administered with E-Prime (Psychology Software Tools, Inc.; Schneider, Eschman, & Zuccolotto, 2002). In each trial, the participants saw a target word printed in the center of the computer display. Five-hundred milliseconds after the onset of the visual presentation of the word, they heard (through headphones) a sentence containing or not containing that word. They were asked to press "o" (for *oui* 'yes') if they heard the word in the display and do nothing if they did not hear the word. Their accuracy rates were measured. In order to increase the likelihood that the participants would incorrectly detect the target words that were not in the sentences in the experimental conditions, the participants were put under time pressure when completing the task. Fourteen of the distracter items that contained the target word were followed by a reminder screen that the participants should try to respond faster. The order of item type (e.g., experimental, control, distracter) was the same for all the participants, but the particular test item appearing under each type was randomized across participants.

4.4 Data Analysis and Predictions

Two experimental triplets were excluded, because the wrong recordings had accidentally been used as stimuli in one of the two experimental conditions. This resulted in the exclusion of 5.5% of the data.

The results will be reported as accuracy rates for the control condition and as false alarm rates for the experimental conditions. Since the former were at ceiling, no statistical analyses were performed on them. For the latter, mixed analyses of variance (ANOVAs), with prosodic information as within-subject variable and, for L2 learners, with proficiency as between-subject variable, were conducted on the arcsine-square-root-transformed subject (F_1) and item (F_2) means. The native and non-native listeners' data were analyzed separately due to the uneven number of participants in each group. To examine the relationship between the participants' ability to use prosodic information and the acoustic cues in the stimuli, linear regression analyses were also performed on the participants' proportion of false alarms with these cues as predictors.

If the participants use prosodic information to segment French speech into words, they should show significantly lower false-alarm rates in the across-AP condition than in the within-AP condition. Given the findings of previous studies (e.g., Tyler & Cutler, 2009), we might also expect to find significant relationships between the native French listeners' proportion of false alarms and the average F0 in the monosyllabic noun, and between the English listeners' proportion of false alarms and the duration of the monosyllabic noun.

5. Results

The proportion of correct responses in the control condition indicated that all three groups were successful at detecting the target word when it was in the stimuli (mid-level L2 learners: .95, SD: .09; high-level L2 learners: .95, SD: .08; native listeners: .98, SD: .04). This suggests that the L2 learners' listening skills were sufficiently good to detect French words in continuous speech.

Figure 1 shows the native and non-native listeners' mean proportion of false alarms (and standard errors) in the across-AP and within-AP conditions. As can be seen from these results, the participants incorrectly detected the target word fewer times in the across-AP condition than in the within-AP condition, and this difference tended to increase with improved proficiency in French. Mixed ANOVAs on L2 learners' proportion of false alarms revealed a significant effect of prosodic information ($F_1(1,26)=39.042$, $p<.011$; $F_2(1,66)=7.966$, $p<.006$) and a significant effect of proficiency ($F_1(1,26)=5.901$, $p<.022$; $F_2(1,66)=18.528$, $p<.001$), but no interaction between prosodic information and proficiency (F's<1), indicating that the high-level L2 learners did not show a significantly larger effect of prosodic information than the mid-level L2 learners. Similar ANOVAs on native speakers' proportion of alarm rates also revealed a significant effect of prosodic information ($F_1(1,10)=20.282$, $p<.001$; $F_2(1,33)=17.978$, $p<.001$). These results confirm that both the native and non-native French listeners used pitch accents to detect word-final boundaries.

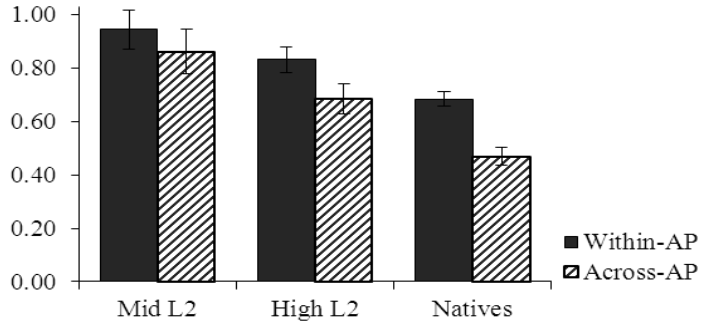

Figure 1. Proportion of False Alarms in the Experimental Conditions

In order to investigate the relationship between the participants' proportion of false alarms and the acoustic cues in the monosyllabic nouns, three linear regressions were performed on the proportion of false alarms, with acoustic cue (F0, duration, and average intensity, one for each regression), group (mid L2, high L2, natives), and the interaction between the prosodic cue and group as predictors. For the three models, the only predictor that reached significance is the acoustic cue × group interaction (F0: $r^2=.283$, $p<.001$; duration: $r^2=.310$, $p<.001$; intensity: $r^2=.214$, $p<.001$). Subsequent linear regressions were thus computed separately for each group. These linear regressions are plotted in

Figure 2. They yielded significant relationships between the native listeners' proportion of false alarms and the F0 values (r^2=.121, p<.003), between the native listeners' and high-level L2 learners' proportions of false alarms and the duration values (respectively, r^2=.145, p<.001; r^2=.117, p<.004), and between the mid-level L2 learners' proportions of false alarms and the intensity values (r^2=.076, p<.023).

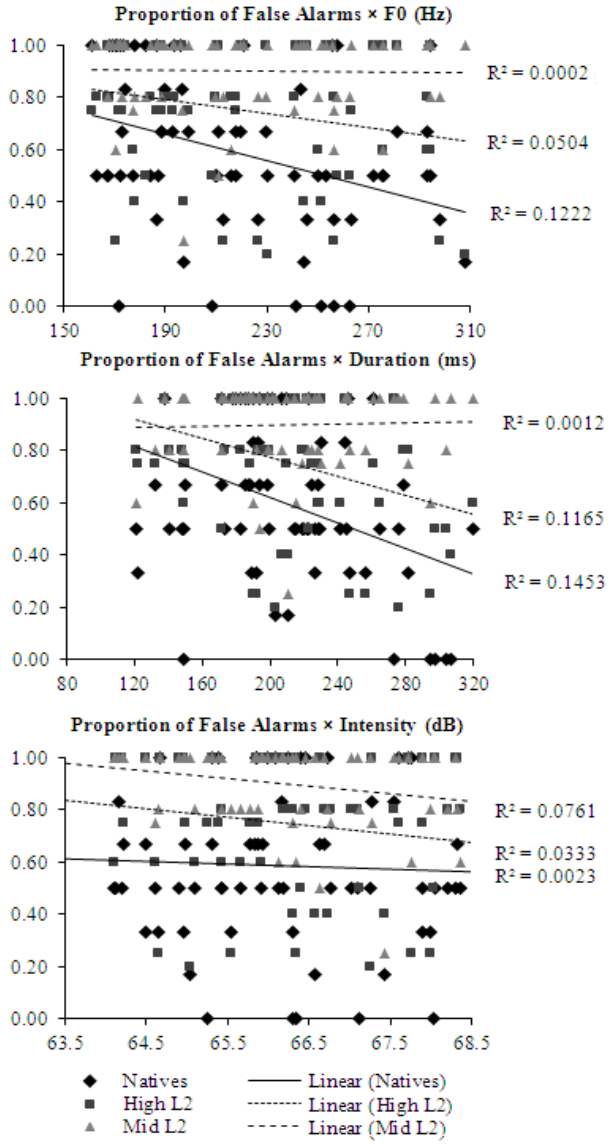

Figure 2. Relationship between the False Alarm Rates and Acoustic Cues

These results suggest that whereas the native French listeners used both F0 and duration to identify word-final boundaries in continuous speech, the high-level L2 learners used only duration, and the mid-level L2 learners used only intensity. It is unclear why the mid-level L2 learners did not also rely on duration, given that it is correlated with prominence in English and it is an important cue to word-final boundaries cross-linguistically. One possibility is that they had poorer listening skills than the high-level L2 learners, and thus had more difficulty detecting syllable boundaries in continuous speech. This could have led them to rely on intensity as a cue to accented syllables rather than on duration as a cue to word-final boundaries (in this case, only the latter required listeners to detect syllable boundaries.

Let us now turn to a discussion of these findings and their implications for understanding L2 speech segmentation.

6. Discussion

Our results showed that the L2 learners used prosodic information to recognize words in continuous speech, indicating that they were able to associate prominent syllables with word-final boundaries in French. Yet, unlike native French listeners, they did not rely on F0 to detect word-final syllables. These results are in line with those of Tyler and Cutler (2009), who showed that French listeners, but not English listeners, use F0 rises as a cue to word-final boundaries. Because both duration and intensity cues coincided with F0 cues in our stimuli, the L2 learners did not need to rely on F0 to hear word-final boundaries in French. It remains to be seen whether they would be able to use F0 in resynthesized stimuli where it is the only cue to word-final boundaries. In the absence of other cues, it is possible that L2 learners' speech segmentation would benefit from F0 rises. On the other hand, since increased duration and F0 rises both characterize AP-final syllables in French, L2 learners may not ever need to associate F0 rises with word-final boundaries. The fact that phrase accents, also signaled by an F0 rise, provide a cue to word-initial boundaries in French may further increase L2 learners' difficulty in using this cue. What is clear from our results, however, is that L2 learners favored duration and intensity over F0 as a cue to word-final boundaries, suggesting perhaps that they know F0 rises in French do not necessarily signal word-initial boundaries (unlike in English).

One might argue on the basis of our results that the high-level L2 learners did not necessarily perceive French accented syllables as such. Since duration is a reliable cue to word-final boundaries in English, irrespective of whether the word-final syllable is stressed, the high-level L2 learners might have heard the accented monosyllabic noun as a word-final syllable simply because of its longer duration. Notice that this cannot be true of the mid-level L2 learners, however, as increased intensity is not a cue to word-final boundaries in English. This means that the mid-level L2 learners must have perceived the accented monosyllabic nouns as prominent. Since these learners were clearly less

advanced in French than the high-level L2 learners (see Table 1), we believe it is unlikely that the high-level L2 learners did not perceive the accented syllables as prominent. The task that both groups have yet to achieve, then, is to establish the correct mapping between F0 rises and prosodic prominence when the latter is word-final.

The results of this study also indicated that both English L2 learners of French and native French listeners could use prosodic information (albeit differently) even if lexical information was not degraded or absent. This may be due in part to the fact that the participants were put under time pressure while completing the experiment, thus increasing the likelihood that they would use non-segmental information for resolving temporary lexical ambiguities in the speech stream. Mattys et al. (2005) proposed that speech segmentation proceeds according to a hierarchy of cues, with cues such as lexical information overriding cues such as stress and prosody; "lower-level" cues have a stronger effect on word recognition when "higher-level" cues are not available, and they reduce (but do not eliminate) the effect of "higher-level" cues if they conflict with them. The efficiency of cues within each level depends on their reliability for identifying word boundaries, which is computed by statistical learning mechanisms (see also Saffran, 2001; Saffran et al., 1996). The less variable and more reliable nature of "higher-level" cues as compared to "lower level" cues (among other factors) contributes to explaining the nature of the proposed hierarchy. Increasing evidence in support for it is emerging from the processing literature (e.g., Mattys et al., 2005; Mattys et al., 2007; Mattys & Melhorn, 2007; White et al., 2010; see also Norris, McQueen, & Cutler, 1995). To the extent that our participants resolved temporary lexical ambiguities with prosody as a result of being put under time pressure, our findings are in line with this hierarchy, and suggest that prosodic information also constrains non-native listeners' speech segmentation.

7. Conclusion

The present study investigated English and French listeners' use of prosodic information in the segmentation of French speech. The results of a word-monitoring task under time pressure confirmed that both the native and non-native listeners used prosodic information to identify word-final boundaries in French, but they extracted different cues from the speech signal, with the L2 learners relying on duration and intensity and with the native listeners relying on F0. Further research with this particular population of L2 learners should seek to isolate these acoustic cues in order to determine whether English listeners can learn to map F0 rises to prominent syllables and segment words at the offset of these syllables.

Appendix: Experimental Items

Across-AP	Within-AP	Control
bal conventionnel et pesant	bal conventionnel	balcon arrondi
bancs dominicaux chrétiens	bancs dominicaux	bandeaux acajous
banc distinct et éloigné	banc distinct	bandit basque
boue gisante et huileuse	boue gisante	bougie blanche
cerf vorace et majestueux	cerf vorace	cerveau droit
chat grincheux et bedonnant	chat grincheux	chagrin fou
chat lépreux et légendaire	chat lépreux	chalet suisse
chat pauvre et mal avenant	chat pauvre et sale	chapeau melon
corps bossu et déformé	corps bossu	corbeau noir
corps végétatifs et ralentis	corps végétatifs	corvées ménagères
coût rentable et sécurisé	coût rentable	courant fort
coup singulier et symbolique	coup singulier	coussin en mousse
dents gélatineuses et difformes	dents gélatineuses	dangers nucléaires
phare doré et balisé	phare doré	fardeau lourd
fort maléfique et hanté	fort maléfique	format papier
fou larmoyant et navrant	fou larmoyant	foulard marron
fours miniatures et métallisés	fours miniatures	fourmis oranges
mat tournant et brisé	mat tournant	matou rond
mort surprenante et mystique	mort surprenante	morsure profonde
paons flétris et ternis	paons flétris	pamphlets roses
père militaire et carriériste	père militaire	permis de chasse
père silencieux et calme	père silencieux	persil chinois
père vertueux et enthousiaste	père vertueux	pervers sexuel
pins séduisants et géants	pins séduisants	pincées brutales
plat fondant et salé	plat fondant	plafond rond
ports ténébreux et désuets	ports ténébreux	portées des chattes
port trépidant et célèbre	port trépidant	portrait photo
rangs partiels et désordonnés	rangs partiels	remparts rouges
rat tondu et bedonnant	rat tondu	raton gris
sols dallés et brillants	sols dallés	soldats belges
temps bourgeois et catholiques	temps bourgeois	tambours blancs
tours néogothiques montantes	tours néogothiques	tournées artistiques
vers gélatineux et luisant	vers gélatineux	vergers abondants
vers séchés et ondulés	vers séchés	versets tristes
vies perdues et sacrifiées	vies perdues	vipères vertes
vie réjouissante et sereine	vie réjouissante	virée nocturne

References

Bagou, O., Fougeron, C., & Frauenfelder, U. H. (2002). Contribution of prosody to the segmentation and storage of “words” in the acquisition of a new mini-language. In

B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002* (pp. 159-162). Aix-en- Provence: Laboratoire Parole et Langage.

Bagou, O., & Frauenfelder, U. H. (2006). Stratégie de segmentation prosodique: Rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle. In *Actes des XXVIèmes Journées d'Étude sur la Parole* (pp. 571-574). Dinard, France.

Banel, M.-H., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication, 15,* 115–126.

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris.

Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labeling (version 3.0). Available at www.ling.ohio-state.edu/_tobi/ame_tobi/labelling_guide_v3.pdf.

Boersma, P., & Weenink, D. (2004). Praat: Doing phonetics by computer (Version 5.2.11) [Computer program]. Available at www.praat.org.

Carroll, S. E. (2004). Segmentation: Learning how to 'hear words' in the L2 speech stream. *Transactions of the Philological Society, 102,* 227–254.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics, 31,* 585–598.

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.

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

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45,* 207–228.

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

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition, 51,* 237–265.

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

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

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1–23.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159–207.

Jusczyk, P. W., Kemler-Nelson, D. G., Hirsh-Pasek, K., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24,* 252–293.

Kemler-Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55–68.

Kim, S. (2004). The role of prosodic cues in word segmentation in Korean. In *Proceedings of the 8h International Conference on Spoken Language Processing* (pp. 3005-3008), Jeju Island, Korea.

Kim, S., Broersma, M., & Cho, T. (to appear). The use of prosodic cues in processing an unfamiliar language. *Studies in Second Language Acquisition.*

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.

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

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 397–408.

Mattys, S. L., & Melhorn J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America, 122,* 554–567.

Mattys, S., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 960–977.

Mattys, S. L., White, L., Melhorn, J. F. (2005). Integration of multiple speech segmentation cues. *Journal of Experimental Psychology: General*. *134,* 477–500.

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.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology.* Dordrecht: Foris.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1209–1228.

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.

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research, 45,* 519–530.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002) *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools Inc.

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

Spinelli, E., Welby, P., & Schaegis, A. L. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision. *Language and Cognitive Processes, 22,* 828–859.

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29,* 553–584.

Tremblay, A. (to appear). Proficiency assessment in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition.*

Tremblay, A., & Garrison, M. D. (2010). Cloze tests: A tool for proficiency assessment in research on L2 French. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected proceedings of the Second Language Research Forum 2008* (pp. 73–88). Somerville, MA: Cascadilla Press.

Tremblay, A., & Owens, N. (2010). The role of acoustic cues in the development of (non-)target-like second-language prosodic representations. *Canadian Journal of Linguistics, 55,* 85–114.

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.

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53-66). Berlin: Springer.

Verluyten, S. P. (1982). *Recherches sur la prosodie et la métrique du français.* Unpublished Ph.D. dissertation, University of Antwerpen.

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word-boundary cue in Tokyo Japanese. *Language & Speech, 53,* 107-131.

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

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