

# An intonational cue to word segmentation in phonemically identical sequences

ELSA SPINELLI

CNRS and Université Pierre Mendès France, Grenoble, France

NICOLAS GRIMAULT

CNRS and Université Claude Bernard Lyon 1, Lyon, France

FANNY MEUNIER

CNRS and Université Lumière Lyon 2, Lyon, France

AND

PAULINE WELBY

CNRS and Université de Provence, Aix-en-Provence, France

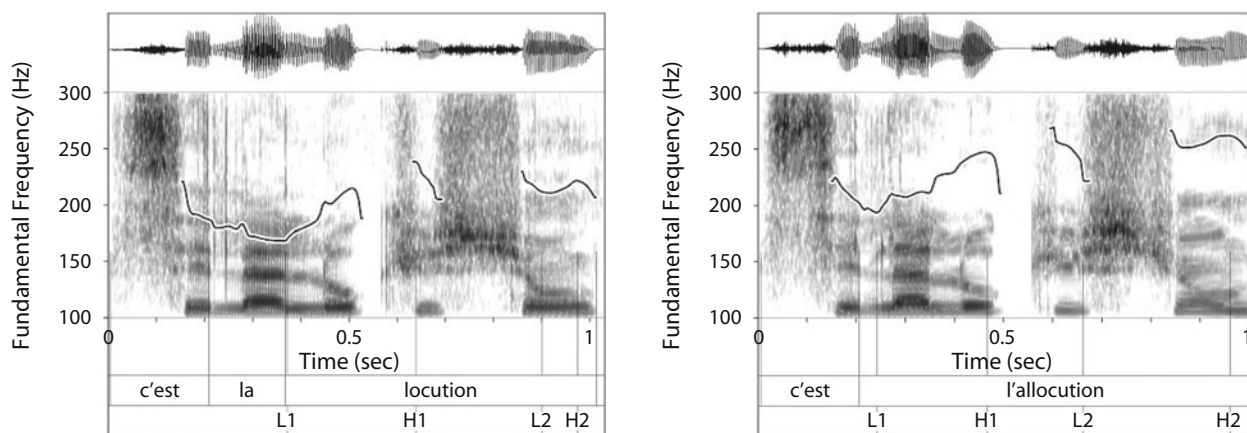
We investigated the use of language-specific intonational cues to word segmentation in French. Participants listened to phonemically identical sequences such as /sɛlafif/, *C'est la fiche/l'affiche* "It's the sheet/poster." We modified the  $f_0$  of the first vowel /a/ of the natural consonant-initial production *la fiche*, so that it was equal to that of the natural vowel-initial production *l'affiche* (resynth-consonant-equal condition), higher (resynth-consonant-higher condition), or lower (resynth-consonant-lower condition). In a two-alternative forced choice task (Experiment 1), increasing the  $f_0$  in the /a/ of *la fiche* increased the percentage of vowel-initial (*affiche*) responses. In Experiment 2, participants made visual lexical decisions to vowel-initial targets (*affiche*) following both the natural consonant-initial production (*la fiche*) and the resynth-consonant-equal version. Facilitation was found only for the resynth-consonant-equal condition, suggesting that raising the  $f_0$  allowed online activation of vowel-initial targets. The recognition system seems to exploit intonational information to guide segmentation toward the beginning of content words.

How do listeners accomplish the task of word segmentation, given that, in spoken language, there are no clear and obvious cues associated with word beginnings and ends? There is now a vast body of evidence showing that listeners use their tacit knowledge of a wide range of patterns in their native language to help them segment speech, including cues from allophonic variation, phonotactic constraints, transitional probabilities, and lexical stress (e.g., Cutler & Norris, 1988; McQueen, 1998; Quené, 1992; Saffran, Newport, & Aslin, 1996). Here, we examine the possibility that cues from intonation or the melodic structure of a language can help listeners find word beginnings in the speech stream.

A given stretch of speech can be consistent with multiple lexical hypotheses, and these hypotheses can begin at different points in the input. In the French sequence *l'abricot* /labʁiko/ "the apricot," segmental information could be compatible with several competing hypotheses, such as *l'abri* /labʁi/ "the shelter," *la brique* /labʁik/ "the brick," and *la brioche* /labʁijɔʃ/ "the brioche." Listeners are routinely confronted with such transient segmentation ambiguities, and some ambiguities are total, as in *Il m'a*

*donné la fiche/l'affiche* /ilmadonelafif/ "He gave me the sheet/the poster," where there is no contextual information favoring one hypothesis over another, and the lexical hypotheses *fiche* "sheet" and *affiche* "poster" both seem to be equally supported by the (lack of) contextual information in the input.

Traditional psycholinguistic models, such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), have considered the processes underlying the mapping of sensory information from the acoustic input to the stored entries in the lexicon from such a phonemic approach. In these models, segmentation can be considered as a by-product of lexical competition, in the sense that it is achieved by a process of competition between candidate words. Lexical hypotheses that are consistent with the input are activated at any moment in time, regardless of their location in the input. Bottom-up activation and lateral inhibition among competing candidates allow listeners to resolve transient ambiguities that arise by selecting candidates that account for the entire input (such as *l'abricot* rather than *l'abri* or *la brique* in /labʁiko/). In addition, a "modified version" of the 1994 Shortlist model incorpo-



**Figure 1.** Two-rise (early rise, late rise) LHLH pattern (taken from the experimental materials) in a consonant-initial item (*la location*; left panel) and a vowel-initial item (*l'allocation*; right panel). Note that the early rise (L1H1) begins at the content word. The word and tone labels are time aligned to the waveform and spectrogram. When an accentual phrase begins with clitic function words, there is often an inflection point in the  $f_0$  curve at the function-word–content-word boundary (e.g., between *la* and *location* and between *c'est* and *l'allocation*), corresponding to L1. This inflection point is clear when there are sonorant or voiced consonants at the critical region (here [l]) but is not present (at least physically) when there are obstruents.

rates lexical stress information: “Candidates starting at a strong syllable onset are given a boost if they themselves have a strong onset” (Norris, 1994, p. 229).

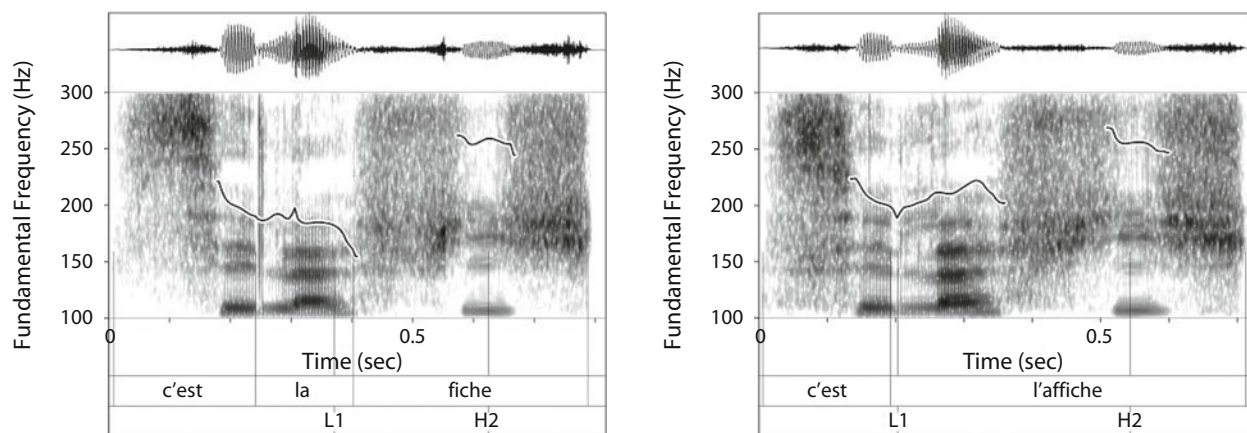
Such a phonemic approach, however, does not account for the correct segmentation of sequences such as *l’affiche/la fiche*, given that the phonemic transcriptions of both items of a pair are strictly identical, differing only in the placement of the word boundary. Phonemically identical phrases such as *c’est la fiche/c’est l’affiche* thus pose a problem for these types of models, because both lexical candidates could account for the entire sequence. To account for correct segmentation of this type of stimulus, we must consider that the recognition system exploits fine-grained sources of information contained in the speech signal. In this study, we focus on the potential contribution of language-specific intonational cues.

There is a growing body of evidence that French listeners are sensitive to the prosodic patterns of their language and use these patterns in word segmentation. Before presenting these studies, we will briefly outline some basics of French prosody. It is generally agreed that French prosodic structure includes a unit of phrasing below the level of the utterance, which we refer to as the *accentual phrase* (AP; Jun & Fougeron, 2000, 2002; for reviews of other terms, see Jun & Fougeron, 2000; Welby, 2003, 2006). A French AP includes one or more content words (nouns, verbs, etc.; typically, one or two) and any preceding clitic function words (determiners, conjunctions, etc.), and its last syllable is lengthened and has special prominence. If the AP is not utterance final, this prominence is typically marked not only by lengthening, but also by a rise in fundamental frequency ( $f_0$ ). This late (final) rise is a marker of what has been called the *primary stress* or *primary accent*. An early  $f_0$  rise is also sometimes found near the beginning of the AP; this early rise is a marker of *secondary stress* or *secondary accent* but does not typically change meaning.

This two-rise (early rise, late rise) pattern is illustrated in Figure 1; the rises are labeled LHLH (L1H1 L2H2)

according to an autosegmental-metrical model, in which intonation patterns are modeled as a series of (H)igh and (L)ow tones, with intermediate values determined by interpolation (see Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988). This rise typically appears on content words and only very rarely on function words. In short APs, where there is not enough time to realize all four tones of the LHLH pattern, other patterns are realized, such as the LH (L1H2) pattern, a rise from an early L to a late H, as illustrated in Figure 2.

Turning to the role of these patterns as cues to word segmentation, Banel and Bacri (1994) showed that listeners could exploit their knowledge of the rhythmic patterns of French in word segmentation. Given ambiguous sequences such as /ba.gaʒ/ (*bagage* “luggage” or *bas gage* “low pledge”), listeners were more likely to hear one word (*bagage*) when the second syllable was lengthened and two words (*bas gage*) when the first was lengthened. This finding corresponds to the expectation that a phrase-final syllable will be lengthened and that a phrase boundary will not occur in the middle of a word.<sup>2</sup> Rietveld (1980) found phonetic differences between minimal pairs such as *le comtat saccagé* “the devastated county” and *le comte a saccagé* “the count has laid waste,” including differences in  $f_0$ , intensity, and duration corresponding to a primary accent and late rise on the final syllable /ta/ of *le comtat* (/lə.kɔ̃.ta/) and a primary accent and late rise on the /kɔ̃/ of *le comte a* (/lə.kɔ̃.ta/). The study also showed that French listeners could retrieve the intended meaning of such minimal pairs and that durational differences were the most reliable cues. Christophe (1993) tested whether French adults and French infants were sensitive to prosodic differences in their native (or to-be-native) language (see also Christophe, Dupoux, Bertoncini, & Mehler, 1994). Both French-speaking adults and French-acquiring infants were able to distinguish between a bisyllabic sequence such as /ma.ti/ excised from intraword contexts such as *mathématicien* “mathematician” versus interword contexts such as *panama tisane* “panama



**Figure 2.** One-rise L1H2 pattern (taken from the experimental materials) in a consonant-initial item (*la fiche*; left panel) and a vowel-initial item (*l'affiche*; right panel). The word and tone labels are time aligned to the waveform and spectrogram. Note that the rise begins at the content word (*fiche* or *affiche*), although there is no clear elbow or inflection point, since the  $f_0$  curve is perturbed by the voiceless fricative [f].

(hat) herbal tea.” Tokens in the intraword and interword conditions differed along several dimensions: In the interword condition, the /t/ closure was longer, and the /a/ was longer and had higher intensity and higher  $f_0$ .

More recently, Bagou, Fougeron, and Frauenfelder (2002) showed that Swiss French listeners could use lengthening of AP-final syllables and  $f_0$  rises on these syllables to detect “words” in an artificial language. Welby (2003, 2007) showed that French listeners could use the presence of an early rise in  $f_0$  or even a simple “elbow” (inflection) in the  $f_0$  curve not followed by a rise as a cue to a content word beginning, building on suggestions in the literature (Di Cristo, 2000; Vaissière, 1997). Listeners interpreted nonsense sequences such as /me.la.mɔ̃.din/ as a single nonword *mélamondine* when the  $f_0$  rise began at the first syllable (/me/) and as two words *mes lamondines* “my lamondines” when it began at the second syllable (/la/). Bagou and Frauenfelder (2006), however, found that whereas a two-rise (LHLH) pattern facilitated “word” identification in an artificial language paradigm, an early-rise-only pattern did not, interpreting this result to mean “initial prominences [do] not fill the demarcative role that has been attributed to them” and that “the prosodic segmentation strategy [is] primarily based on final prominences, the presence of which constitutes a necessary condition for the potential use of initial prominences” (p. 574; our translation). Welby (2007), however, suggested that this conclusion might be premature, since the artificial language used in the study contained only content words and no function words (and there was, therefore, no function-word/content-word alternation) and since the early-rise-only pattern is rare (see Jun & Fougeron, 2000, 2002; Welby, 2006).

In addition to these studies of French, a small number of recent studies have examined the role of intonational cues to word segmentation in other languages. Kim and colleagues used the word-spotting task to examine the role of prosodic regularities in word segmentation in Korean. Kim (2004a, 2004b) investigated the contribution of intonation pattern type and location of the word within the AP, finding an effect of location of the word within the AP and

a limited effect of intonation pattern type. She suggested that this may be due either to a facilitatory role for the AP-final H tone in the AP preceding the target AP or to domain-initial strengthening.

In a further word-spotting study, Kim and Cho (2009) studied the role of intonation pattern—in particular, intonation pattern frequency—and duration in Korean word segmentation. Participants spotted words more accurately when the intonation pattern across the AP boundary was frequent (H#L; “#” here indicates an AP boundary) than when the pattern was less frequent (L#), and they were more sensitive to this cross-boundary information than to prosodic patterns within the AP. When lengthening of the AP-final syllable was added, the participants’ accuracy improved only when the tone preceding the AP boundary was less frequent (L#). In contrast, and somewhat surprisingly, when this tone was more frequent (H#), the participants performed less accurately, suggesting that these two cues do not operate in a straightforward cumulative way in word segmentation. Due to particularly high error rates, which the authors noted are common in word-spotting studies, Kim and Cho were unable to analyze their response time (RT) data.

In a recent study of the contribution of intonational cues to segmentation in Japanese, Warner, Otake, and Arai (2010) investigated the contribution of  $f_0$  rises at AP boundaries (AP boundary rises, or APRs). Their study of a spontaneous speech corpus showed that 68% of the words began with an  $f_0$  rise and that the overwhelming majority of all  $f_0$  rises (97%) occurred at the beginning of a word. They used a word-spotting task comparing the detection of words in AP-medial position without an APR with that of words in AP-initial position with an APR. Listeners spotted words in a nonword sequence (e.g., *kazari* “decoration” in *rekerenikazari*) more quickly and more accurately when the word was in an AP-initial position, produced with an  $f_0$  rise (an APR). The authors noted that this  $f_0$  rise was accompanied by “slight lengthening”<sup>3</sup> and concluded that “the results show an effect of prosody overall rather than specifically the  $f_0$  rise, but the  $f_0$  rise is likely to be the largest perceptual cue to the AP boundary.”



We now have growing evidence for the use of intonational cues in word segmentation. The time course of the use of these cues—that is, whether intonational cues are used online in lexical access—remains an unresolved question. Although an online task (word spotting) was used in the studies of Korean, Kim (2004b) and Kim and Cho (2009) reported few significant results for their RT measures. And although the RT results reported in Warner et al. (2010) likely reflected the early use of an intonational cue in Japanese word segmentation, as the authors themselves noted, we cannot exclude the possibility that other prosodic cues present in their natural speech stimuli played a role. Similarly, although the results of one of our earlier studies on segmentation of phonemically identical sequences in French (Spinelli, Welby, & Schaegis, 2007) were compatible with the use of intonational cues in the earliest stages of processing, our natural speech stimuli contained other potential cues (e.g., from duration).

In this earlier study (Spinelli et al., 2007), listeners were well above chance level in discriminating and identifying phonemically identical sequences such as /selafif/, “It’s the sheet/poster.” There were greater priming effects for targets that matched the intended segmentations than for targets that did not. Acoustic analyses identified differences in duration, formant structure, and  $f_0$  between members of pairs such as *l’affiche* (/l#a/; here, “#” indicates a word boundary) versus *la fiche* (/la#/) and *l’allocation* (/l#a/) versus *la locution* (/la#/). Among the cues that differentiated the two members of the pairs, some cues were more likely to be used by listeners during word segmentation, as was shown by the positive correlation between the acoustic cues, including  $f_0$ , and the identification data. In cases of ambiguities such as (/l#a/)... (e.g., *l’affiche*) and (/la#/)... (e.g., *la fiche*), a higher  $f_0$  across the first vowel seems to indicate the beginning of a content word (e.g., *l’affiche* or *l’allocation*).<sup>4</sup> This difference can be seen in Figures 1 and 2, in which the  $f_0$  in *la fiche* and *la locution* (left panels) stays relatively low until the beginning of the content words *fiche* and *locution* but rises across the first syllable of *l’affiche* and *l’allocation* (right panels). Therefore, among the cues associated with each intended segmentation, there was an intonational regularity that listeners could potentially use in order to decode the speech stream.

The positive correlation between  $f_0$  values and identification data strongly suggested that listeners relied on intonational cues to segment such ambiguous sequences. There was, however, no direct evidence. But Vaissière (1997) explicitly proposed such a role for the early  $f_0$  rise of French: “It is reasonable to suppose that this acoustic marking of word beginnings [the early rise] speeds lexical access” (p. 68, our translation). Vaissière drew a parallel between lexical stress in English and intonational patterns in French: “It is the regular alternation in melodic height between content words and function words that seems to us to be an essential characteristic of French (along the same lines as the alternation between strong syllables and weak syllables in English)” (p. 66, our translation).

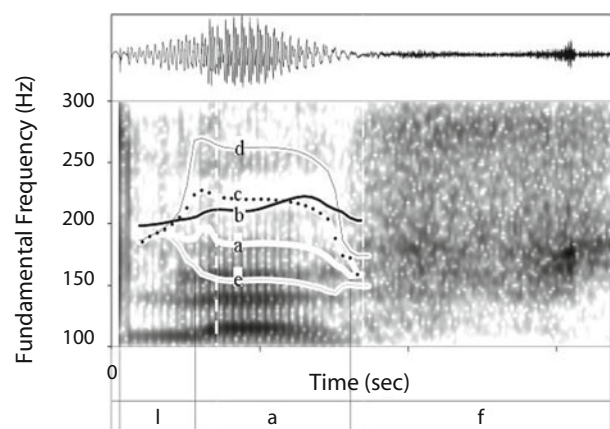
In the present study, we examined the potential role of this intonational cue in lexical access and controlled for

the confounding influence of other prosodic factors present in Spinelli et al. (2007) by resynthesizing the  $f_0$  of the natural speech stimuli from our earlier study. In this way, our aim was to provide direct evidence that listeners use intonational cues to disambiguate between phonemically identical spoken phrases such as /selafif/ *C’est l’affiche/la fiche* in offline and online segmentation.

We present two experiments in which French listeners listened to ambiguous sequences taken from the Spinelli et al. (2007) study, such as those in (1), where both competitors (*fiche*, *affiche*) are matched in lexical frequency and syntactic category:

- (1) a. C’est la fiche. /selafif/ “It’s the sheet.”  
b. C’est l’affiche. /selafif/ “It’s the poster.”

In Experiment 1, listeners performed an offline identification task, and in Experiment 2, listeners performed a visual lexical decision task after hearing phonemically identical sequences as primes (e.g., /lafif/). In Experiment 1, the participants were confronted with one of the five versions we created for each sequence, corresponding to five experimental conditions: the two versions produced by the speaker (*la fiche* and *l’affiche*) and three resynthesized versions created by modifying the  $f_0$  across the first vowel (/a/). The five conditions (see Figure 3) were (1) *natural-consonant* (e.g., *la fiche*, consonant-initial production, no resynthesis), (2) *natural-vowel* (e.g., *l’affiche*, vowel-initial production, no resynthesis), (3) *resynth-consonant-equal* (e.g., consonant-initial item *la fiche* with  $f_0$  of the first vowel /a/ resynthesized to equal that of the vowel-initial item *l’affiche*), (4) *resynth-consonant-higher* (e.g., consonant-initial item *la fiche* with the  $f_0$  of the first vowel /a/ resynthesized to be higher than



**Figure 3.** Illustration of the  $f_0$  resynthesis for the pair *la fiche/l’affiche*. The  $f_0$  curves are superimposed and time aligned with the spectrogram and waveform for the natural production of the three first phonemes (/laf/) of the consonant-initial *la fiche* and are labeled a–e. (a) Natural consonant-initial *la fiche*. (b) Natural vowel-initial *l’affiche*. (c) Resynth-consonant-equal condition resynthesized from natural *la fiche*. (d) Resynth-consonant-higher condition (resynthesized from natural *la fiche*). (e) Resynth-consonant-lower condition (resynthesized from natural *la fiche*). The dashed vertical white lines represent the boundaries of the first vowel of *l’affiche* and are relevant only for the  $f_0$  of the natural *l’affiche* production (b).

that of the vowel-initial item *l'affiche*), and (5) *resynth-consonant-lower* (e.g., consonant-initial item *la fiche* with the  $f_0$  of the first vowel /a/ resynthesized to be *lower than* that of the consonant-initial item *la fiche*). Considering the intonational phonology of the language, we hypothesized that the manipulation of  $f_0$  values would modify the listeners' segmentation. Decreasing the  $f_0$  value in the /a/ of *la fiche* should give rise to more consonant-initial segmentation (*la fiche*), whereas increasing the  $f_0$  in the /a/ of *la fiche* should give rise to more vowel-initial segmentation (*l'affiche*).

Note that we use the terms *consonant-initial* and *vowel-initial* as a short way of referring to members of pairs such as *la fichell'affiche*, but we believe that the variation in the  $f_0$  indicates, in fact, the beginning of a content word in French. A higher  $f_0$  across the first vowel should indicate the beginning of a content word, such as *l'affiche*, where the /l/ of the definite article is resyllabified (/la/) as the onset of the content word, as compared with *la fiche*, where the /la/ is not content word initial but is, rather, a clitic function word. This basic difference in intonational patterns extends beyond *consonant-initial* and *vowel-initial* pairs arising from elision. For example, in a pair such as *mésaventure* "misadventure"/*mes aventures* "my adventures," the consonant-initial *mésaventure* would likely be produced with a rise across the first syllable, whereas in the vowel-initial *mes aventures*,  $f_0$  would probably stay low before rising at the beginning of the content word.

## EXPERIMENT 1

### Method

**Participants.** Fifty undergraduate students from the Université Pierre Mendès France, Grenoble II participated in the experiment for course credit. They were all native speakers of French and reported no hearing impairment.

**Materials.** Twenty-nine pairs of phonemically identical sequences composed of a definite article followed by a noun (e.g., *la fichell'affiche* "the sheet/the poster") were selected from the 30 pairs in Spinelli et al. (2007; for the complete list, see the Appendix). The noun in one member of the minimal pair was vowel initial (e.g., *af-fiche*), and the noun in the other member of the pair was consonant initial (e.g., *fiche*). The members of the pairs did not differ in lexical frequency [ $t(28) = 1.26, p > .20$ ]. Spinelli et al. (2007) showed that there were clear intonational differences between the two conditions. There was a systematically lower  $f_0$  ( $M = 179$  Hz) for the first vowel, /a/, in the consonant-initial items (e.g., *la fiche*) than in the vowel-initial items ( $M = 195$  Hz; e.g., *l'affiche*) and no difference in  $f_0$  for the second vowel between the two conditions (e.g., the /i/ in *la fichell'affiche*;  $M = 219$  Hz for the consonant-initial condition vs. 221 Hz for the vowel-initial condition, a statistically insignificant difference;  $F < 1$ ). For the present experiment we used the 29 pairs of the 30 from Spinelli et al. (2007) for which this general intonational pattern applied—that is, the pairs for which there was a lower  $f_0$  for the first vowel in consonant-initial items (*la fiche*) than in the vowel-initial items (*l'affiche*). All of the two-syllable items (e.g., *la fichell'affiche*) in both conditions had a single-rise L1H2 pattern (see Figure 2), and all but one of the three- and four-syllable items (*la tension/l'attention*, *la locution/l'allocation*) had either the two-rise LHLH pattern (see Figure 1) or an LHH pattern, in which there was an early rise followed by a high plateau stretching to the end of the AP. All of the items, therefore, were realized with  $f_0$  rises early in the AP. We will collectively refer to these rises as *APRs*, adopting the term used in Warner et al. (2010).<sup>5</sup>

We used the natural productions of the consonant-initial (*la fiche*) and vowel-initial (*l'affiche*) versions of each pair, and for each pair, created three additional tokens by performing  $f_0$  resynthesis of the consonant-initial item. As part of the Spinelli et al. (2007) study, segment boundaries were labeled, using Praat software, following standard segmentation criteria (Peterson & Lehiste, 1960). For the present study, the relevant segment boundaries are those marking the beginning and end of the first vowel (/a/) (*la fiche*). The /l-/a/ boundary was located at the onset of vocal formant structure and the accompanying increase in amplitude, and the /a/-consonant boundary was located at the break in vowel formant structure, which was usually accompanied by a clear drop in amplitude.

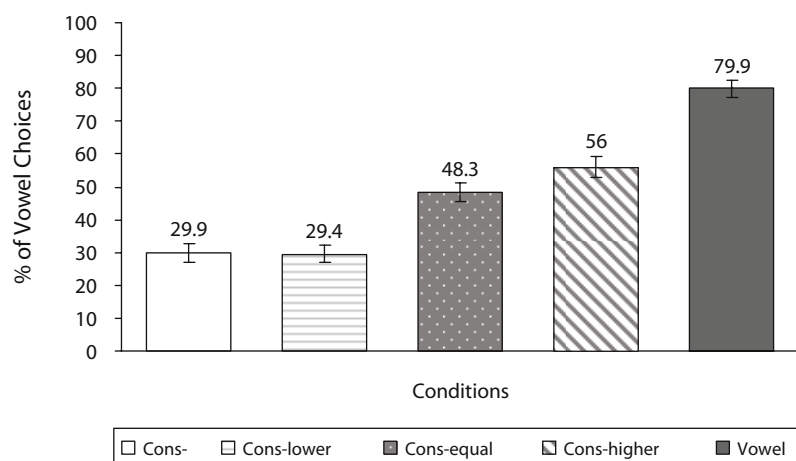
For the present study,  $f_0$  resynthesis was performed using STRAIGHT software (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999) in a MATLAB environment. To vary the  $f_0$  value of the first vowel in each of the three resynthesis conditions, the average  $f_0$  value of the first vowel was computed in raw linear hertz for each pair, using Praat software (Boersma & Weenink, 2004). These parameters were superimposed on the natural consonant-initial  $f_0$  contours (line a in Figure 3). For each pair, a scaling factor was defined as the ratio between the average  $f_0$  values of /a/ from the consonant-initial version (*la fiche* [line a]) and the vowel-initial version (*l'affiche* [line b]) and was applied in a time window corresponding to the /a/ of the consonant-initial version (*la fiche*) with 20-msec cosine ramps. The  $f_0$  curve (plotted on a hertz scale) for the vowel /a/ of the consonant-initial version (*la fiche*) was first multiplied by this scaling factor to reach the same mean  $f_0$  value as in the vowel /a/ in the vowel-initial version (*l'affiche*) (resynth-consonant-equal condition); second, it was multiplied by twice this scaling factor (resynth-consonant-higher condition); and third, it was divided by this scaling factor (resynth-consonant-lower condition). Figure 3 shows the critical early part (three first phonemes: /laf/) of the  $f_0$  contour in hertz in all the experimental conditions for the pair *la fichell'affiche*. Note that although we could have resynthesized the  $f_0$  curve using more traditional intonational landmarks, such as turning points, the technique used here shifts the *tonal center of gravity* (TCoG). According to the TCoG hypothesis (Barnes, Brugos, Shattuck-Hufnagel, & Veilleux, 2008), what is important in signal contrasts is a shift in the bulk of (relatively) high and low  $f_0$ , regardless of the exact shape of local parts of the pattern, which may be interrupted by obstruents or realized as more or less concave/convex, "peaky," and so forth. If we were to repeat the experiment, however, we would resynthesize not only the first vowel, but also the onset /l/ consonant.

**Procedure.** The participants were presented with sequences taken from the five experimental conditions (e.g., *C'est la fiche*; resynth-consonant-equal) and were asked to identify the noun by making a forced choice between the two possible nouns (e.g., *fiche* vs. *af-fiche*). The consonant-initial choice (*fiche*) was presented on the left, and the vowel-initial choice (*af-fiche*) on the right. As a control, we also included 30 pairs of non-phonemically-identical sequences composed of a definite article followed by a noun such as *le bain* "the bath"/*le pain* "the bread" (unambiguous set). The two members of these pairs differed only in voicing or in place of articulation.

The experiment was controlled by E-Prime software. The stimuli were presented over Sennheiser HD 212 Pro headphones at a comfortable listening level. The stimuli were counterbalanced in five experimental lists, so that each participant heard all the conditions but only one member of each ambiguous pair. Within each list, the order of item presentation was randomized. The participants were asked to respond by pressing one of the two response buttons, and their responses were collected.

### Results and Discussion

The results of 2 participants who were not at 100% correct responses in the unambiguous control set were discarded from the analyses. Percentages of vowel choices in the five conditions are presented in Figure 4. First of all, it



**Figure 4.** Mean percentages of vowel responses (i.e., content word initial *l'affiche*) in the five conditions of Experiment 1: natural consonant-initial *la fiche*, resynth-consonant-lower condition (resynthesized from natural *la fiche*), resynth-consonant-equal condition (resynthesized from natural *la fiche*), resynth-consonant-higher condition (resynthesized from natural *la fiche*), and natural vowel-initial *l'affiche*. Error bars show standard errors of the means.

should be noted that for natural productions—that is, productions for which  $f_0$  was not manipulated—the participants show percentages of correct responses of 75% (79.9% for vowel-initial items and 70.1% for consonant-initial items), which are comparable to those obtained in Spinelli et al. (2007; see Welby, 2003, 2007, for similar results on nonword sequences). The results were evaluated using one-way repeated measure ANOVAs with five levels of condition (natural-consonant, resynth-consonant-lower, resynth-consonant-equal, resynth-consonant-higher, natural-vowel).  $F$  values are reported for analyses with participants ( $F_1$ ) and items ( $F_2$ ) as random factors. Analyses showed a main effect of condition [ $F_1(4,188) = 67.2$ ,  $MS_e = 313.3$ ,  $p < .001$ ;  $F_2(4,112) = 42.6$ ,  $MS_e = 292.6$ ,  $p < .001$ ].

Crucially, there was a significant difference between the natural-consonant condition and the resynth-consonant-equal condition [ $F_1(1,47) = 23.2$ ,  $MS_e = 353.5$ ,  $p < .001$ ;  $F_2(1,28) = 18.2$ ,  $MS_e = 268.6$ ,  $p < .001$ ], with more vowel choices in the resynth-consonant-equal condition (48.3%) than in the natural-consonant condition (29.9%). The consonant-lower condition also differed significantly from the resynth-consonant-equal condition [ $F_1(1,47) = 31.99$ ,  $MS_e = 267.6$ ,  $p < .001$ ;  $F_2(1,28) = 12.49$ ,  $MS_e = 407.3$ ,  $p < .001$ ]. Moreover, there was also a difference between the resynth-consonant-equal condition and the resynth-consonant-higher condition, with more vowel responses (*affiche*) (56.0%). This difference is significant by participants [ $F_1(1,47) = 4.02$ ,  $MS_e = 353.7$ ,  $p < .05$ ], although it did not quite reach significance by items [ $F_2(1,28) = 3.35$ ,  $MS_e = 226.8$ ,  $p = .07$ ]. There was no significant difference between the natural-consonant condition and the resynth-consonant-lower condition (both  $F$ s  $< 1$ ).

**Correlations.** Correlation analyses were carried out between the identification data of Experiment 1 and the  $f_0$  measurements of the stimuli. Because some pairs show greater differences in the  $f_0$  of the first syllable vowel be-

tween the two conditions (e.g., /la/ in *la fiche* vs. *l'affiche*), we hypothesized that the gain in vowel-initial segmentation percentages (from the natural-consonant condition to the resynth-consonant-equal condition) should be greater for pairs with greater  $f_0$  differences than for others. There was indeed a positive correlation [ $r(27) = .67$ ,  $p < .001$ ] between the  $f_0$  differences of the pairs and the increase in percentage of vowel responses (*affiche*) from the natural-consonant condition to the resynth-consonant-equal condition.

Our results show that decreasing the  $f_0$  of the first vowel in items such as *la fiche* (resynth-consonant-lower condition) does not induce more consonant-initial segmentation. However, there is a gradient effect of increased  $f_0$  on the vowel-initial segmentation. Increasing the  $f_0$  value of the /a/ of /la#fiʃ/ induced a gain of 18.4% (from 29.9% to 48.3%) of vowel segmentation when the value was raised to meet that of the natural vowel segmentation (resynth-consonant-higher) and of 26.1% (from 29.9% to 56.0%) when it was higher. It should be noted, however, that in the resynth-consonant-equal condition, and even in the resynth-consonant-higher condition, the percentages of vowel segmentation did not reach those of natural vowel-initial segmentation, confirming that there are other cues responsible for the correct disambiguation of such sequences, including durational and formant cues (see Spinelli et al., 2007).<sup>6</sup>

One might wonder, however, whether the increase in  $f_0$  is really linked to vowel-initial segmentation. In the resynth-consonant-equal conditions, the participants' responses gave rise to a percentage (48.3%) that did not differ from chance. It could be argued that resynthesis does not increase vowel-segmentation responses but, rather, disturbs listeners so that they respond at chance. In order to ensure that the 48% of the responses in the resynth-consonant-equal condition did, in fact, correspond to a real increase in vowel-initial segmentations (rather than chance), we looked at online activation of vowel-initial



words in the resynth-consonant-equal condition. If raising the  $f_0$  to simulate an early rise did, in fact, bias listeners to vowel-initial responses, the modified version of the ambiguous sequence (resynth-consonant-equal) should show a priming effect on the vowel-initial target words, as compared with the natural, unmodified (*la fiche*) consonant-initial segmentation.

## EXPERIMENT 2

In this experiment, we used a cross-modal identity priming paradigm with a lexical decision task. The participants heard short French sequences such as /lafif/. At the offset of the acoustic stimulus, they saw a target letter string (e.g., *affiche* “poster”) on a computer screen and were asked to decide whether or not these letter strings were real French words.

### Method

**Participants.** Thirty-seven undergraduate students from Université Louis Lumière, Lyon II participated in the experiment for course credit. They were all native speakers of French and reported no hearing impairment. None of them had taken part in Experiment 1.

**Stimuli and Procedure.** The 29 vowel-initial nouns in Experiment 1 (e.g., *affiche*) served as experimental visual targets. Three auditory primes were associated with each target. One related prime consisted of the article + noun sequence corresponding to the consonant condition in the previous experiments, with the neutral context (*c'est* “it is”) excised from the carrier phrase, using Adobe Audition software (e.g., *la fiche*; natural-consonant condition). Another related prime consisted of the article + noun sequence corresponding to the resynth-consonant-equal condition in the previous experiments, with the neutral context (*c'est*) excised from the carrier phrase (e.g., *la fiche*; resynth-consonant-equal condition). A third prime, consisting of an article + noun sequence, served as a baseline (e.g., *l'amiral* “the admiral”). Half of the unrelated primes were elided sequences of article + vowel-initial nouns (e.g., *l'amiral* “the admiral”), and the other half were sequences of article + consonant-initial nouns (e.g., *la route*, “the road”).

For the lexical decision task, 29 nonwords (e.g., *drofut*) were created for presentation as visual targets. Three auditory primes were associated with each nonword target. Two related primes consisted of an article + noun ambiguous sequence (e.g., *l'adroite* “the skillful one”; *la droite* “the right”). Such phonologically related pairs were associated with nonword targets to discourage the participants from associating phonological overlap with a *yes* response. In one prime, the consonant-initial segmentation item was used (e.g., *la droite*; consonant-initial intended condition), and in the other, the vowel-initial segmentation item was used (e.g., *l'adroite*; vowel-intended condition). A third prime, consisting of an article + noun sequence, served as a baseline (e.g., *l'araignée* “the spider”). Half of the unrelated phrases were elided sequences of article + vowel-initial noun (e.g., *l'araignée* “the spider”) and the other half were sequences of article + consonant-initial noun (e.g., *la tisane* “the herbal tea”).

In order to reduce the proportion of related pairs to 25%, 100 targets (50 real words, 50 nonwords) were presented in unrelated conditions. Fifty of the primes (25 associated with existing word targets and 25 associated with nonword targets) consisted of the masculine singular definite article + a masculine singular noun (e.g., *le bureau*—LARDON “the desk—BACON”), and the other 50 primes (25 associated with existing word targets and 25 associated with nonword targets) consisted of the plural definite article + a plural noun (e.g., *les traces*—PLATEAU “the traces—TRAY”).

The experiment was controlled by E-Prime software. Primes were presented auditorily over Sennheiser HD 212 Pro headphones at a comfortable listening level. Fifty milliseconds after the offset of the

prime, the target was visually displayed in lowercase letters at the center of a computer screen. The participants were asked to indicate whether the visual target was a real word or a nonword by pressing one of two response buttons as quickly and as accurately as possible. They were required to press the *yes* button with the index finger of their preferred hand and the *no* button with the index finger of the other hand. The stimuli were counterbalanced across three experimental lists, so that each participant received all three priming conditions (consonant, consonant-equal, and baseline) but saw each target only once. Within each list, the order of item presentation was randomized. The computer clock was triggered at the presentation of the target on the screen and stopped when the participant responded. Response latencies and errors were collected. The session began with 10 practice trials.

### Results and Discussion

RTs were calculated from onset of visual target presentation to the time of the buttonpress. Those longer than 1,300 msec and shorter than 200 msec (a total of 1.6% of the responses) were removed. Errors were also removed (2.8% of the responses), as well as data for two items that led to an extremely high percentage of incorrect responses [*avaleur* “(sword) swallower,” 73%; and *amarre* “mooring rope,” 49% of incorrect responses]. Mean RTs, standard deviations (*SDs*), and error rates for word targets in the three priming conditions are given in Table 1. The results were evaluated using one-way repeated measure ANOVAs with three levels (consonant, consonant-equal, and baseline). *F* values are reported for the analyses with participants ( $F_1$ ) and items ( $F_2$ ) as random factors.

Analyses of RTs revealed a main effect of priming condition that was significant by participants [ $F_1(2,72) = 4.02$ ,  $MS_e = 2,433.6$ ,  $p < .05$ ;  $F_2(2,52) = 1.77$ ,  $MS_e = 3,205.1$ , n.s.]. Planned comparisons showed a facilitation for the resynth-consonant-equal condition, relative to the baseline condition, that was significant by participants [ $F_1(1,36) = 5.87$ ,  $MS_e = 3,285.9$ ,  $p < .05$ ] and marginally significant by items [ $F_2(1,26) = 3.74$ ,  $MS_e = 2,982.3$ ,  $p = .06$ ].<sup>7</sup> There was no significant facilitation for the natural-consonant condition relative to the baseline condition [ $F_1(1,36) = 1.37$ ,  $MS_e = 2,138.6$ , n.s.;  $F_2(1,26) = 2.09$ ,  $MS_e = 2,022.3$ , n.s.]. Analyses conducted on errors revealed no effect of priming condition [ $F_1(2,72) = 2.88$ ,  $MS_e = 22.6$ ,  $p = .07$ ;  $F_2(2,52) = 1.61$ , n.s.].

Hence, we found no evidence of activation of the targets (e.g., *affiche*) when this segmentation was not intended (auditory stimulus, *la fiche*; natural-consonant condition). There was however evidence of activation of the targets (e.g., *affiche*) when the  $f_0$  was increased to meet that of

**Table 1**  
Mean Reaction Times (in Milliseconds), With Standard Deviations (*SDs*), and Percentages of Errors to the Vowel-Initial Targets (e.g., *affiche*) in the Three Priming Conditions in Experiment 2

Condition	Reaction Time		Errors (%)
	<i>M</i>	<i>SD</i>	
Natural-consonant	643	86	1
Resynth-consonant-equal	623	100	3
Baseline	656	96	3

the intended vowel segmentation. This result thus suggests that increasing the  $f_0$  value did, in fact, increase the vowel-initial segmentation and allows us to rule out the possibility that listeners were responding only at chance in the resynth-consonant-equal condition in Experiment 1.

## GENERAL DISCUSSION

There is now a growing body of evidence suggesting that information beyond the phonemic level is available to the lexicon and that, during spoken word recognition, fine-grained differences in the speech signal influence processing at the lexical level and, thus, modulate lexical selection (e.g., Cutler & van Donselaar, 2001; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Spinelli et al., 2007). Furthermore, there is converging evidence from a variety of paradigms that language-specific intonational cues play a role in word segmentation. Spinelli et al. (2007) suggested that French listeners relied on early intonational cues to segment ambiguous sequences such as /selafif/ but provided no direct evidence of such a role. In the present study, we showed that manipulating the  $f_0$  of the /a/ vowel in sequences such as /lafif/ to correspond to intonational patterns found in natural speech influenced perceived segmentation in expected ways. Raising the  $f_0$  in the natural consonant-initial production of *la fiche* /la#fi/ led to more vowel-initial segmentation (*l'affiche*) (Experiment 1) and also increased activation of the lexical representation of the vowel-initial target (*affiche*). To our knowledge, this is the first direct evidence of the use of an intonational cue to lexical access.<sup>8</sup> In contrast, no activation of the vowel-initial target (*affiche*) was found for the natural production of *la fiche* /la#fi/ (Experiment 2).

Our results confirm and extend the results of Welby (2003, 2007), demonstrating that there are reliable, language-specific intonational cues to French content word beginnings. Welby (2003, 2007) provided evidence from offline identification tasks that the early  $f_0$  rise of the two-rise LHLH pattern and even a simple "elbow" (inflection) in the  $f_0$  curve not followed by a rise acted as cues to content word beginnings. The present study confirms that influence of the early rise (L1H1) of the two-rise LHLH pattern in an online task measuring lexical activation and provides the first evidence that the single-rise pattern L1H2, which is very common in short APs (see Figure 2), and the LHH pattern also signal a content word beginning. The recognition system was able to exploit this intonational information to provide differential bottom-up support to competitors, guiding segmentation toward the beginning of content words.

The finding of intonational cues at function-word/content-word boundaries is compatible with models in which function words play a special role in infant and adult word segmentation (e.g., Christophe, Guasti, Nespor, Dupoux, & Van Ooyen, 1997; on infants' sensitivity to function words, see also Gerken, Jusczyk, & Mandel, 1994; Gerken & McIntosh, 1993; Shafer, Gerken, Shucard, & Shucard, 1992). Some assumptions, however, must be reconsidered in light of the results of the present experiments. Christophe et al. (1997), for example, write that:

one could assume that function words are somehow phonetically marked (an assumption, however, that seems rather implausible for French). In this view, subjects would be able to identify real function words just by listening to them and no content word syllable would ever be truly homophonous to a function word . . . Each of these proposals makes different predictions that can be tested experimentally. [Note: For instance, the prosodic hypothesis predicts that subjects will be misled when encountering a syllable homophonous to a function word at the beginning of a prosodic unit (such as the French verb "décider" where "des" is the plural [in]definite article).] (p. 608)

Our results, however, clearly demonstrate that French proclitic function words are, in fact, phonetically marked; for example, they are shorter than the phonemically identical content-word-initial syllables, show systematic differences in formant values, and have markedly distinct intonational properties (the present study; Spinelli et al., 2007; Welby, 2003, 2006). These markers are effective, although imperfect, in disambiguating segmentally identical sequences in both offline and online tasks. That such robust differences should exist is not surprising; a number of studies have demonstrated that purportedly homophonous words in fact show systematic differences in spontaneous speech. For example, Lavoie (2002) found that the English function word *for*, but never its phonemic homophone, the content word *four*, was often realized as [fə]. And Gahl (2008) showed that high-frequency words (e.g., *time*) are shorter than their less frequent phonemic homophones (e.g., *thyme*) (see also Fosler-Lussier & Morgan, 1999, on the relationship between frequency and pronunciation). Since function words are among the most common words in any language, the shorter durations observed in our previous data seem likely to be, at least in part, due to frequency. In addition, Local (2003) argued that "[g]rammatical or function words have systematically different phonetic exponents from nonfunction words" (p. 333), noting, for example, that English function words behave differently from content words with respect to nasal assimilation (see also Selkirk, 1995).

Recall that we found a difference between the resynth-consonant-equal and resynth-consonant-higher conditions, but not between the natural consonant-initial (*la fiche*) condition and the resynth-consonant-lower condition. This contrast in results can be explained by a number of factors. First, we used the natural *la fiche* productions as the basis for all three resynthesized conditions (resynth-consonant-equal, resynth-consonant-higher, and resynth-consonant-lower). In the natural *la fiche* condition, listeners were, of course, provided with all acoustic cues to correct word segmentation. Our results here and in previous work (Spinelli et al., 2007; Welby, 2007) and observations of missegmentations in natural speech suggest that, in the absence of contextual cues, listeners do not reach 100% segmentation accuracy but, rather, reach the ceiling at around 70%. In the resynth-consonant-lower condition, the listener had access to all the acoustic cues to the correct segmentation (duration, formant cues, etc.), with a (potentially) enhanced intonational cue,



but this cue did not seem to be strong enough to move the ceiling and raise the percentage of *la fiche* responses. That we did not find a difference between the natural-consonant condition (*la fiche*) and the resynth-consonant-lower condition, in which the  $f_0$  of the vowel of first syllable /a/ was lowered, may also be due in part to the differing functions of raising and lowering  $f_0$  in French and across languages. It is common, perhaps even universal, to raise  $f_0$  to increase the salience of high tones (e.g., to convey emphasis), but it seems to be rarer to lower  $f_0$  to emphasize low tones, and there is no evidence in the literature that French speakers employ this strategy.

Another question that remains is how the recognition system uses intonational information to guide segmentation. Although we varied the  $f_0$  value in the first syllable of our experimental sequences, at one level or another, interpretation of  $f_0$  (and other acoustic) cues is, of course, dependent on context. To give a trivial example, it cannot be the case absolutely that an  $f_0$  value of 200 Hz signals the beginning of a content word and an  $f_0$  value of 180 Hz the beginning of a function word. Fundamental frequency (e.g., the  $f_0$  of the /a/ in *la fiche*) is interpretable only in context—for example, by considering speaker-related factors (e.g.,  $f_0$  range and sex) and the intonation pattern of the surrounding context. It may be important to take into account, directly or indirectly, the difference between the  $f_0$  of V1 (e.g., /a/ in *la fiche*) and that of V2 (e.g., /i/ in *la fiche*). In Spinelli et al. (2007), we noted that “there was a lower  $F_0$  for the first vowel [V1] in the two-word (*la fiche*) items (mean: 179 Hz) than in the one-word (*l’affiche*) items (mean: 195 Hz) [ $F(1,29) = 106.7$ ;  $p < .001$ ] and no difference in  $f_0$  for the second vowel in the two-word and one-word pairs (219 Hz for the two-word items vs. 221 Hz for the one-word items;  $F < 1$ )” (p. 837) and that this pattern was due to an earlier timing of the rise to a peak in the second syllable in items such as *l’affiche* than in items such as *la fiche*. It seems likely that listeners compute relative measures of  $f_0$  in order to detect word boundaries and to interpret  $f_0$  patterns more generally.

We assume that intonational information will be used very early in processing, as it unfolds, in line with what we know about how other types of cues to word segmentation are used. As van Donselaar, Koster, and Cutler (2005) noted, “listeners exploit incoming information in speech signals in a continuous manner, making use of cues as soon as they become available in order to modulate the activation of potential candidate words that have received some support from the speech signal” (p. 267). In their study, fragments taken from the first syllable of Dutch words such as *Octopus* primed *Octopus*, whereas fragments taken from words with second-syllable stress such as *October* did not. Reinisch, Jesse, and McQueen (in press) confirmed and extended these results in an eyetracking study. They observed more fixations on first-syllable-stress targets such as *Octopus* than on second-syllable-stress competitors such as *October*, well before disambiguating segmental information became available. It seems entirely plausible that  $f_0$  information available to French listeners across the first couple of phones of /la#f/ or /la#af/, taken from *la fiche* or *l’affiche*, will boost lexical activation of the ap-

propriate content word. Listeners very quickly (probably within the first couple of syllables) evaluate the pitch range and other speaker characteristics of their interlocutor and have an idea of what counts as high or low  $f_0$  for that talker. They also know the proclitic function words of their language and the prosodic properties of the words in this closed class (see, e.g., Selkirk, 1995). And for French listeners, this includes specific intonational properties. The fragment-priming study of Spinelli et al. (2007) did not allow us to directly examine this question, since we used natural speech stimuli that varied along dimensions other than  $f_0$ . But we plan follow-up studies to examine more directly the time course of the use of intonational cues.

There is a growing body of research on listeners’ perception of intonation patterns (e.g., Barnes et al., 2008; Dilley & McAuley, 2008; Niebuhr, 2007). Barnes et al. proposed the notion of a TCoG, a measure of where the “mass of the high tonal region” is located. They hypothesized that “[p]erception of tonal alignment is driven not by the location of  $F_0$  turning points per se, but by the location of the Tonal Center-of-Gravity (TCoG) within the relevant region.” With respect to the minimal pairs in our study, the TCoG is located consistently farther to the right in consonant-initial (*la fiche*) items than in vowel-initial items (*l’affiche*) (see Figure 2). TCoG allows listeners to perceive as “the same” intonation patterns that differ on the surface—for example, rises with simple peaks and rises with high plateaux (D’Imperio, 2000; Welby, 2003) and, crucially for our study, cases with and without clear turning points in the  $f_0$  curve. For example, in Figure 1, the critical region contains only sonorant segments, but in Figure 2, there are breaks in the  $f_0$  curve due to segmental perturbations at the obstruent (e.g., the /f/ in *la fiche*/*l’affiche*).<sup>9</sup> If it is to account for our results, the TCoG must be calculated incrementally as the intonation pattern unfolds. There is also evidence that listeners take into account intonational characteristics of preceding, nonlocal context (*distal* context; Dilley & McAuley, 2008), but that is not a factor here.

The extent to which listeners abstract away from the  $f_0$  information in the signal remains an open research question. It is clear that speakers/listeners learn and reproduce the intonational patterns of their language—for example, producing early rises at the beginning of content words that are new to them (or at the beginning of nonwords; see Welby, 2003, 2007). But it does not necessarily follow that listeners need to recover the exact identity of an intonational pattern to use intonational cues. Indeed, the exact intonational pattern may not be clear until a number of syllables downstream; in Figure 1 (*la locution*/*l’allocation*), for example, listeners may not “know” that they are dealing with a two-rise LHLH pattern, rather than a single-rise L1H2 pattern, until the third syllable of the content word. The relationship between the computation of intonational patterns (as in the TCoG model discussed in the preceding paragraph) and other cues to word segmentation is also an open research question.

Our results are incompatible with models of word recognition in which variability in the acoustic signal is factored out before lexical access to recover sublexical units such as phonemes, because phonemic transcriptions of

our stimuli are all identical. There is, therefore, a need for the current models of spoken word recognition to incorporate information such as intonational information in the process of lexical competition. So far, most of the current models of spoken word recognition, such as TRACE or Shortlist, could not account for our results—not even Shortlist B, a recent Bayesian model of continuous speech recognition (Norris & McQueen, 2008). In this model, the concept of lexical activation is abandoned for that of likelihood and probability. A search algorithm evaluates multiple hypotheses, searching for the best path through a word lattice. In its current form, Shortlist B could not simulate our results, because the input the model uses are sequences of multiple phoneme probabilities derived from a large-scale diphone gating task. According to the developers of Shortlist B, this is a limitation of using the diphone database as input, rather than a limitation of the theory (D. Norris, personal communication, 2009). According to this view, if the model had access to the level of phonetic detail, *f*0 information could be used to shift the bias between the alternative parses.

In fact, there is a recent attempt to model the way prosodic information could bias the competition process. Salverda, Dahan, and McQueen (2003) and Cho, McQueen, and Cox (2007) have proposed a *prosody analyzer* that computes prosodic structure in parallel with segmental analysis in order to provide cues as to the likelihood of the presence of a word boundary. The locations of boundaries in the prosodic structure could then be used to modulate the lexical competition process. We suggest that, for French, the content-word-initial rises (APRs) that influenced segmentation in our study could be one of these boundary cues.

An alternative account is the exemplar-based view of word recognition (Goldinger, 1998, *inter alia*), in which it is assumed that each word in the mental lexicon is associated with a set of detailed exemplars of that word and that the variability present in the speech signal such as voice, *f*0, and intonation contour (Church & Schacter, 1994; Johnson, 1997) forms part of the speaker/listener's lexical knowledge. Listeners may use fine-grained acoustic (and other) information present in the input in a comparison with this stored information. Our results would then suggest that this acoustic information would include content-word-initial rise (APR) information.

It is also important to consider our results in light of recent work on the relative contribution of contextual versus acoustic information to lexical access. Under normal listening circumstances, listeners might also be helped by other types of cues provided by the sentential context. Numerous studies have evaluated the role of linguistic context on the activation of lexical candidates (see Simpson, 1994, for a review). Our results show that in the total absence of pragmatic context, listeners succeed in segmenting 75% of the time, suggesting that acoustic cues are reliable guides to segmentation. Mattys and Melhorn (2007) recently showed that in “multicue listening environments, listeners adopt compensatory segmentation strategies whereby reliance on sentential lexical information is inversely related to the strength of the acoustic cues.” When listeners had to choose between *pie* and *eye* in phrases

such as *plump eye/plum pie*, the bias toward the contextually appropriate segmentation was greater when the acoustic cues were weak, and the effect of context was reduced when acoustic cues toward one version were strong. In other words, when acoustic cues are weaker, listeners rely more on context. Moreover, it should be noted that in cases of strong conflicting contextual/acoustic cues (e.g., in sentences such as *The surgeon looked at the drawing of a plum ## pie*, where context favored *eye* but the acoustics favored *pie*), there were around 80% of acoustically biased responses, suggesting that acoustic cues outweighed contextual ones. This suggests that acoustic cues are very powerful, although the authors suggested that this effect might have been due to the task's focusing on form rather than on meaning. The task in our study also focused on word form; further research is needed to investigate the weighting of the use of acoustic cues in natural listening conditions where listeners are focused on meaning.

Finally, it is important to consider the broader cross-linguistic implications of our results, beyond French. Warner et al. (2010) suggested that *f*0 rises might serve as a universal word boundary cue, writing that

it is cross-linguistically very common to use a basic rise–fall *f*0 pattern for some prosodic unit (Vaissière 1983a, 1983b; Hirst & Di Cristo 1998), and nearly all the languages in Hirst and Di Cristo's, (1998) broad cross-linguistic survey of intonation show this pattern. . . . In languages where rises are timed to stresses, listeners would have to infer word boundaries based on what they know about stress position.

A similar claim was made by Shukla, Nespor, and Mehler (2007), who showed that Italian listeners were able to use Japanese intonational phrase boundaries to detect “words” while learning an artificial language. Kim and Cho (2009) cautioned against interpreting this result as evidence for universality, although “Italian listeners are likely to have exploited Japanese IP boundaries not because the particular prosodic cues that were available to them were universal, but because Italian and Japanese IP happen to share some prosodic cues” (p. 3383). They also noted that their results for Korean challenge the Warner et al. (2010) claim: Korean listeners were less accurate in detecting words when they began with an H tone surrounded by L tones (L#HL). In line with Kim and Cho, we would argue that although the set of potential cues to word segmentation may be universal, the details and the weighting of these cues is likely to be language specific (and context specific; see Mattys, White, & Melhorn, 2005). We also would argue that although tonal alignment differences are likely to signal differences cross-linguistically, the nature of the differences conveyed will vary. As we have noted in earlier work, “we might expect, for example, that text-to-tune alignment is an important dimension in all spoken languages, although it plays different roles in different languages. For example, text-to-tune alignment signals semantic and pragmatic distinctions in some languages and word boundaries in others” (Welby, 2003, p. 31). As Warner et al. noted, more study is needed to establish the generality of intonational cues with respect to word segmentation.

In conclusion, we find converging evidence from offline and online paradigms that confirm and extend the results of Welby (2003, 2007) and Spinelli et al. (2007). Early  $f_0$  rises in the AP (APRs) acted as cues to word segmentation and lexical access. Vaissière (1997) and Di Cristo (2000) suggested such a role for the French early  $f_0$  rise. Our materials included not only items with this early rise in the two-rise (LHLH) intonational pattern examined in Welby (2003, 2006, 2007), but also items with the early-rise–high-plateau LHH pattern and the single-rise L1H2 pattern. This finding is particularly important because the two-rise LHLH pattern and the single-rise L1H2 pattern are the most common intonational patterns in French; these two patterns along with the LHH pattern—that is, APRs—are found in approximately 75% of the Welby (2003, 2006) data. Since this early intonational cue is so common, we expect that it will be used by French listeners. Moreover, this cue is particularly powerful. An intonational rise early in the French accentual phrase (an APR) almost always marks an *actual* word boundary, whereas other cues to word segmentation, such as transitional probabilities and lexical stress, mark *potential* word boundaries. For example, in the preceding sentence, there are six content words containing stressed syllables in the middle of the word (*into'national*, *ac'centual*, *segmen'tation*, *tran'sitional*, *proba'bilities*, and *po'tential*). The recognition system exploits intonational information to provide differential bottom-up support to competitors, guiding segmentation toward the beginning of content words. Early intonational cues thus play a robust role in word segmentation in French and, quite probably, in other languages.

#### AUTHOR NOTE

We thank Lynne Nygaard and two anonymous reviewers for their valuable comments on an earlier version of the manuscript. We also thank Ben Munson and Steve Winters for their help. Correspondence concerning this article should be addressed to E. Spinelli, Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès France, BP 48, 38040 Grenoble Cedex 9, France (e-mail: elsa.spinelli@upmf-grenoble.fr).

#### 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.
- BARNES, J., BRUGOS, A., SHATTUCK-HUFNAGEL, S., & VEILLEUX, N. (2008, April). *Turning points, tonal targets, and the English L-phrase accent*. Paper presented at the Conference on Experimental and Theoretical Advances in Prosody, Cornell University.
- BECKMAN, M. E., & PIERREHUMBERT, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255-310.
- BOERSMA, P., & WEENINK, D. (2004). Praat: Doing phonetics by computer (Version 4.2.20) [Computer program]. Available at [www.praat.org](http://www.praat.org).
- CHO, T., MCQUEEN, J., & COX, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, **35**, 210-243.
- CHRISTOPHE, A. (1993). *Rôle de la prosodie dans la segmentation en mots*. PhD dissertation, École des Hautes Etudes en Sciences Sociales.
- CHRISTOPHE, A., DUPOUX, E., BERTONCINI, J., & MEHLER, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, **95**, 1570-1580.
- CHRISTOPHE, A., GUASTI, T., NESPOR, M., DUPOUX, E., & VAN OYEN, B. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language & Cognitive Processes*, **12**, 585-612.
- CHURCH, B. A., & SCHACTER, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 521-533.
- CUTLER, A., & NORRIS, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 113-121.
- CUTLER, A., & VAN DONSELAAR, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language & Speech*, **44**, 171-195.
- DI CRISTO, A. (2000). Vers une modélisation de l'accentuation du français: Seconde partie. *French Language Studies*, **10**, 27-44.
- DILLEY, L. C., & MCAULEY, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory & Language*, **59**, 294-311.
- D'IMPERIO, M. (2000). *The role of perception in defining tonal targets and their alignment*. PhD dissertation, Ohio State University.
- FOSLER-LUSSIER, E., & MORGAN, N. (1999). Effects of speaking rate and word predictability on conversational pronunciations. *Speech Communication*, **29**, 137-158.
- GAHL, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, **84**, 474-496.
- 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.
- GERKEN, L., & MCINTOSH, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, **29**, 448-457.
- GOLDINGER, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**, 251-279.
- HIRST, D., & DI CRISTO, A. (1998). A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 1-44). Cambridge: Cambridge University Press.
- JOHNSON, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.
- JUN, S.-A., & FOUGERON, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), *Intonation: Analysis, modeling and technology* (pp. 209-242). Boston: Kluwer.
- JUN, S.-A., & FOUGERON, C. (2002). The realizations of the accentual phrase in French intonation. *Probus*, **14**, 147-172.
- KAWAHARA, H., MASUDA-KATSUSE, I., & DE CHEVEIGNÉ, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based  $f_0$  extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, **27**, 187-207.
- KIM, S. (2004a). The role of prosodic cues in word segmentation in Korean. In *Proceedings of the 8th International Conference on Spoken Language Processing* (pp. 3005-3008), Jeju Island, Korea.
- KIM, S. (2004b). *The role of prosodic phrasing in Korean word segmentation*. PhD dissertation, University of California, Los Angeles.
- 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.
- LAVOIE, L. (2002). Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association*, **32**, 175-202.
- LOCAL, J. (2003). Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics*, **31**, 321-339.
- MATTYS, S. L., & MELHORN, J. F. (2007). Sentential, lexical, and acous-



- tic effects on the perception of word boundaries. *Journal of the Acoustical Society of America*, **122**, 554-567.
- MATTYS, S. L., WHITE, L., MELHORN, J. F. (2005). Integration of multiple speech segmentation cues. *Journal of Experimental Psychology: General*, **134**, 477-500.
- MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.
- MCQUEEN, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory & Language*, **39**, 21-46.
- MEUNIER, C., & ESPESER, R. (2010). *Vowel reduction in conversational speech in French: The role of lexical factors*. Manuscript submitted for publication.
- MILLOTTE, S., RENÉ, A., WALES, R., & CHRISTOPHE, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 874-885.
- MUTHUSAMY, Y. K., COLE, R. A., & OSHIKA, B. T. (1992). The OGI multi-language telephone speech corpus. In *Proceedings of the 3rd International Conference on Spoken Language Processing* (pp. 895-898). Banff, AB.
- NIEBUHR, O. (2007). The signalling of German rising-falling intonation categories—The interplay of synchronization, shape, and height. *Phonetica*, **64**, S. 174-193.
- NORRIS, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.
- NORRIS, D., & MCQUEEN, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, **115**, 357-395.
- PETERSON, G. E., & LEHISTE, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, **32**, 693-703.
- PIERREHUMBERT, J. B. (1980). *The phonology and phonetics of English Intonation*. PhD dissertation, Massachusetts Institute of Technology. Distributed 1987 by Indiana University Linguistics Club.
- PIERREHUMBERT, J. B., & BECKMAN, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
- QUENÉ, H. (1992). Integration of acoustic-phonetic cues in word segmentation. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 349-356). Berlin: Mouton de Gruyter.
- REINISCH, E., JESSE, A., & MCQUEEN, J. M. (in press). Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately. *Quarterly Journal of Experimental Psychology*.
- RIETVELD, A. C. M. (1980). Word boundaries in the French language. *Language & Speech*, **23**, 289-296.
- SAFFRAN, J., NEWPORT, E., & ASLIN, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory & 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.
- SELKIRK, E. (1995). The prosodic structure of function words. In J. Beckman, L. Walsh Dickey, & S. Urbanczyk (Eds.), *Papers in optimality theory* (pp. 439-470). Amherst, MA: GLSA Publications. Also in J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187-214). Mahwah, NJ: Erlbaum.
- SHAFFER, V., GERKEN, L. A., SHUCARD, J., & SHUCARD, D. (1992). *"The" and the brain: An electrophysiological study of infants' sensitivity to English function morphemes*. Paper presented at the Boston University Conference on Language Development, Boston. [Cited in Christophe et al. (1997)]
- SHATZMAN, K. B., & MCQUEEN, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, **68**, 1-16.
- SHUKLA, M., NESPOR, M., & MEHLER, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, **54**, 1-32.
- SIMPSON, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359-374). San Diego: Academic Press.
- SPINELLI, E., MCQUEEN, J., & CUTLER, A. (2003). Processing resyllabified words in French. *Journal of Memory & Language*, **48**, 233-254.
- 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 & Cognitive Processes*, **22**, 828-859.
- VAISSIÈRE, J. (1983a). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53-66). Berlin: Springer.
- VAISSIÈRE, J. (1983b). The search for language-independent prosodic features. In *La percezione del linguaggio* (pp. 311-388). Florence: Presso l'Accademia Della Crusca.
- VAISSIÈRE, J. (1997). Langues, prosodies et syntaxe. *Traitement Automatique des Langues*, **38**, 53-82.
- VAN DONSELAAR, W., KOSTER, M., & CUTLER, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, **58A**, 251-273.
- WARNER, N., OTAKE, T., & ARAI, T. (2010). Intonational structure as a word-boundary cue in Tokyo Japanese. *Language & Speech*, **53**, 107-131.
- WELBY, P. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. PhD dissertation, Ohio State University.
- 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., & LÆVENBRUCK, H. (2006). Anchored down in Anchor-age: Syllable structure and segmental anchoring in French. *Rivista di Linguistica/Italian Journal of Linguistics*, **18**, 74-124.

## NOTES

1. In the model adopted here, the early rise (L1H1) is a bitonal phrase accent whose low tone is associated with the beginning of the first content word syllable. The late rise (L2H2) is a bitonal pitch accent whose high tone is associated with the last syllable of the AP. The indices (1, 2) after the tones indicate that L1 and H1 (and L2 and H2) "go together," following the notation introduced by Welby and Lævenbruck (2006).

2. We interpret the lengthening in Banel and Bacri (1994) as phrase-final lengthening, for two reasons. First, the stimuli were developed from words and two-word sequences read in isolation, and the final syllables of these sequences were thus produced in a phrase-final context. Second, the degree of lengthening is comparable to that found for AP-final lengthening in other studies (mean difference of 35% between the short and long versions [see, e.g., Bagou, Fougeron, & Frauenfelder, 2002, and Bagou & Frauenfelder, 2006]; mean duration of 400 msec for syllables in pair-first position and 600 msec for those in pair-final position or isolation).

3. Warner et al. (in press) reported that "the current materials show significant lengthening at the medial AP boundary, which may also provide a perceptual cue. However, the *f*<sub>0</sub> rise at AP boundaries is quite noticeable, while a duration difference of 16 msec is small relative to most linguistically meaningful duration distinctions." We agree that it is very likely that the intonational rise is the major cue here. But it is not at all clear that the duration difference does not also play a role. There are many reports in the phonetics literature of robust duration differences comparable in magnitude (see, e.g., Meunier & Espesser [2010] on syllable position and vowel reduction in French, and the duration differences between the first syllable of pairs like *la fiche*/*l'affiche* in Spinelli, Welby, & SchaeGIS [2007]). On some level, these differences must be perceptible to listeners, or they would not maintain them in production. This is one of the main reasons we chose to conduct the present study using *f*<sub>0</sub> resynthesis—to address the potential confounding influence of these durational differences.

4. Note that intonationally, elided forms such as *l'affiche* and *l'allocation* behave as if they were content-word initial, although structurally, they are composed of the elided form of the definite article and a noun. See the discussion in Welby (2003, pp. 198-199).

5. Note that only items with the LHLH and LHH patterns have the "early rise" discussed in the literature under various names (initial rise, initial accent, *accent secondaire*, etc.). The term *APR* refers to LHLH and LHH, as well as L1H2. Warner et al. (in press) correctly pointed out that the French AP is realized with a greater variety of tonal patterns than is the Japanese

AP. However, comparing a French read speech corpus (Welby, 2003, 2006) and the Japanese spontaneous speech corpus used in Warner et al. (2010) (a subset of the OGI corpus; Muthusamy, Cole, & Oshika, 1992), APRs are actually realized at a comparable rate in the two languages.

6. Spinelli et al. (2007) found that a lower second formant ( $F_2$ ) in the first vowel of *la fichel'affiche* sequences was associated with a vowel-initial segmentation (*l'affiche*), corresponding to a backer articulation of the [a]. Since fundamental frequency resynthesis can often slightly change formant values, we checked that  $F_2$  values at the midpoint for all the items in our resynthesised experimental conditions did not vary (resynth-consonant-equal, 1857 Hz; resynth-consonant-higher, 1857 Hz; resynth-consonant-lower, 1862 Hz; all pairwise comparisons not significant). We are therefore confident that the observed effects are indeed due to the  $f_0$  manipulation and are not confounded by any slight  $F_2$  variation introduced by the  $f_0$  resynthesis.

7. The fact that this effect does not reach significance in the by-items analyses seems to be due to one pair of items ([lavi] *la vie* "the life"/*l'avis* "the opinion") for which the consonant-initial competitor is extremely frequent (*vie* "life"; 711.5 occurrences per million). The activation of this very high frequent competitor in the ambiguous sequence [lavi] could

have blocked that of the vowel-initial target. When this pair is removed from the analysis, the effect is also significant by items ( $p < .05$ ).

8. Millotte, René, Wales, and Christophe (2008) examined the role of prosody in disambiguating syntactically ambiguous (phonemic) homophones, but this study did not specifically manipulate or control intonation independently of other prosodic factors. Similarly, as was discussed earlier, prosodic factors other than intonation were not directly controlled in the natural speech data of Warner et al. (in press).

9. The present study used materials containing a variety of segments, including many obstruent consonants in the critical region (e.g., the [f] in *la fichel'affiche* and the [t] in *la tension/l'attention*, which caused perturbations in the  $f_0$  curve. The Welby (2003, 2007) experiments, by contrast, included only materials with sonorant or at least voiced consonants in the critical regions, a restriction common in experimental intonational research. As was noted in Welby (2003, p. 222), that listeners would use intonational cues even when the  $f_0$  curve was perturbed seemed likely but not entirely obvious. Note also that turning points must be perceptually available to listeners, at least in some cases, since speakers produce  $f_0$  peaks and elbows neatly aligned with segmental material, if they can (i.e., if the segmental string is sonorant and if they have enough time; see Welby, 2003, 2006).

## APPENDIX

### Stimulus Sets Used in Experiments 1 and 2

Ambiguous Pairs (Experiment 1)					
Vowel Initial		Consonant Initial		Unrelated Primes	
Targets (Experiment 2)	Frequency	Related Primes (Experiment 2)	Frequency	(Experiment 2)	Frequency
alarme	5.32	larme	6.48	Vowels	
avarice	3.16	varice	N.A.	amande	3.16
aversion	1.77	version	11.19	allure	39.58
attraction	10.65	traction	11.74	altitude	12
attention	111.03	tension	25.68	amiral	11.94
amarre	1.16	mare	14.97	année	162.84
attache	17.84	tâche	49.29	allée	33.13
avenue	26.06	venue	52.42	annulation	2.23
arène	3.23	reine	37.9	armure	4.84
amie	33.55	mie	4.45	artère	3.26
attente	39.35	tente	24.52	ascenseur	12.39
affiche	6.77	fiche	12.26	assassin	10.52
haleine	14.19	laine	7.81	atome	16
annotation	0.13	notation	8.48	avancement	11.9
allocation	3.74	location	5.06	aventure	40.48
apesanteur	0.45	pesanteur	8.32	Consonants	
avis	70.16	vie	711.55	bénédictio	5.35
atout	2	toux	6.94	cheville	5.16
allocution	1.84	locution	0.58	route	168.42
appareil	51.77	pareille	31.26	mention	10.26
appel	65.1	pelle	6.55	messagerie	0.1
apposition	0.68	position	85.48	migration	3.68
arrêt	38.65	raie	9	nomination	7.06
ajout	0.16	joue	79.16	philosophie	42.06
attribut	2.42	tribu	11.45	pluie	69.74
accroche	9.1	croche	0.9	rançon	2.9
avaleur	0.23	valeur	152.32	relecture	0.71
affiliation	0.97	filiation	3.52	volée	11.19
approbation	11.03	probation	0.16	table	200.58
<i>M</i>	18.36	<i>M</i>	49.27	tendinite	N.A.
				légende	18.84
				<i>M</i>	32.51