University of Zurich[UZH]

Zurich Open Repository and Archive
University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

Segments as carriers of prosodic information in word onsets

Torres, Catalina ; Welby, Pauline

Abstract: Research has shown that listeners exploit prosodic cues to carry out word segmentation, and that the choice of cues and how these are weighted are language specific. French accentual phrases are demarcated to the left edge with a phrase initial rising (LHi) accent. Listeners are sensitive to the start of the rise (from L to Hi) which is used for word segmentation and to disambiguate between segmentally identical pairs. However, it is unclear whether prosodic cues are only used when disambiguation is required or if they play a more general role. Additionally, it remains an open question as to how sensitive listeners are to prosodic information of the initial rise encoded at the segmental level. This study examines whether microprosodic variations influence word recognition in less well studied segmental environments such as consonant clusters. A manipulation of duration and fundamental frequency cues at the word onset was performed. Results show that lexical activation was significantly delayed for words with onsets containing voiced consonant clusters of the type /bl/. Lexical access was also delayed for onsets of the type /pl/, although the effect was weaker. These results provide preliminary evidence that French listeners are sensitive to fine-grained prosodic information on segments.

DOI: https://doi.org/10.21437/tai.2021-13

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-238430
Conference or Workshop Item
Published Version

Originally published at:
Torres, Catalina; Welby, Pauline (2021). Segments as carriers of prosodic information in word onsets. In: International Conference on Tone and Intonation, Sonderborg, Denmark, 6 December 2021 - 9 December 2021, 61-65.
DOI: https://doi.org/10.21437/tai.2021-13

1st International Conference on Tone and Intonation (TAI)
6-9 December 2021, Sonderborg, Denmark

Segments as carriers of prosodic information in word onsets

Catalina Torres[1], *Pauline Welby*[2]

[1]University of Zurich, Switzerland; University of Melbourne, Australia;
ARC Centre of Excellence for the Dynamics of Language, Australia
[2]Aix Marseille University, CNRS, LPL, Aix-en-Provence, France; University of New Caledonia, Noumea

catalina.torres@ivs.uzh.ch

Abstract

Research has shown that listeners exploit prosodic cues to carry out word segmentation, and that the choice of cues and how these are weighted are language specific. French accentual phrases are demarcated to the left edge with a phrase initial rising (LHi) accent. Listeners are sensitive to the start of the rise (from L to Hi) which is used for word segmentation and to disambiguate between segmentally identical pairs. However, it is unclear whether prosodic cues are only used when disambiguation is required or if they play a more general role. Additionally, it remains an open question as to how sensitive listeners are to prosodic information of the initial rise encoded at the segmental level. This study examines whether microprosodic variations influence word recognition in less well studied segmental environments such as consonant clusters. A manipulation of duration and fundamental frequency cues at the word onset was performed. Results show that lexical activation was significantly delayed for words with onsets containing voiced consonant clusters of the type /bl/. Lexical access was also delayed for onsets of the type /pl/, although the effect was weaker. These results provide preliminary evidence that French listeners are sensitive to fine-grained prosodic information on segments.
Index Terms: word recognition, French, segmental prosody

1. Introduction

A great deal of research has investigated how listeners process the speech stream in order to arrive at meaningful entities such as words. This has led to a rich body of psycholinguistic research into mechanisms involved in word segmentation, word recognition, and lexical access. This study investigates the role of prosodic cues during word recognition in French listeners. More precisely, we set out to examine the role of prosodic cues at the left edge of an Accentual Phrase (AP), encoded in segments of different types. A growing body of research indicates that strategies in word segmentation are language specific [1, 2, 3, 4, 5, 6]. In French, a language without lexical stress [7], lexical distinctions based on prominence marking are not stored in the mental lexicon [8] and intonation is marked at the post-lexical level.

French intonational structure is thus defined post-lexically and the Accentual Phrase represents the lowest tonally marked prosodic constituent [9, 10, 7, 11]. The AP consists of one or more content words, optionally preceded by one or more proclitic function words and has been reported to contain approximately 2.3 to 2.6 words (or 1.2 content words), and 3.5 to 3.9 syllables in read speech [10]. The notation /LHiLH*/ represents the underlying tonal pattern of the canonical AP. To mark a distinction between the two low tones, we will use /L_1HiL_2H*/ in this description, following a similar notation by [12]. A phrase final rise (L_2H*) is typically found on the metrically strongest and last full syllable (non-reduced, non-schwa) of the AP whereas a non-accentual rise can occur phrase initially (L_1Hi) [7, 11]. The position of the primary prominence is fixed at the word level but its realisation relies upon the location of a word within the phrase. Tonal targets within the AP can be undershot or omitted leading to variations in the tonal configuration of the AP [10, 11]. Shorter APs (of three or fewer syllables), for example, are more likely to be produced with simpler patterns, such as a /L_1H*/ pattern [11]. In addition to this intonational structure, an AP is marked by a lengthening of the last full syllable. The syllable to which H* is associated is significantly longer than non-final ones.

French listeners do rely on prosodic cues for word segmentation, as a number of studies have shown [13, 2, 14, 15, 16, 17, 1, 18]. Studies using an artificial language experiments have concluded that (Swiss) French listeners rely on word-final prominences for word segmentation [1, 18]. However, these studies did not include function words in the artificial language which limits the conclusions that can be made. Prominence, in these studies, is understood as a high tone or a lengthened syllable. As a result, it has been argued that French listeners do not rely on word-initial prominence when segmenting the speech stream. One study demonstrated that in non-word sequences like [me.la.mɔ̃.din] when a fundamental frequency (F0) rise (early rise) started at the second syllable, listeners were more likely to perceive the sequence as two words (*mes lamondines*) but when the early rise started at the first syllable the sequence was perceived as a single content (non)word (*melamondine*) [13]. In contrast to English, where statistical regularities in stress pattern (predominantly word-initial stress) can be used in word segmentation by signalling potential word boundaries [6], the French early rise signals actual rather than potential word boundaries. [14] showed an even finer use of F0 as cue in which the beginning of the rise in F0 at the onset of a segmentally identical pair of words (*l'affiche* vs. *la fiche*) helped to disambiguate them and facilitated lexical access.

Another line of study has shown that speakers are sensitive to very fine differences at the segmental level when activating lexical items [19, 20]. It has been shown that French listeners can detect word boundaries based on duration cues of voiceless plosives, whereby a longer stop closure is interpreted as a word boundary [21, 22]. Interestingly, studies on German and French have demonstrated that the acoustic-phonetic characteristics of voiceless fricatives are systematically affected by the F0 context [23, 24, 25], showing that aperiodic pitch impressions of voiceless fricatives reflect the pitch level of the adjacent F0 contour. This means that voiceless fricatives show higher spectral energy in environments with high F0 and lower values in environments with low F0. Moreover, it was found that listeners of German

10.21437/TAI.2021-13

are sensitive to these differences [26]. This suggests listeners are sensitive to intonation cues which are encoded at the segmental level. This study investigates the influence of F0 and duration as prosodic cues to word recognition while taking into consideration segmental prosodic effects. An experiment with an online lexical decision task was conducted. We are interested in examining the effect of a manipulation of duration and F0 on the detection of content word onsets. The experiment examines the relevance of F0 and duration as cues at the left edge of French words, by further examining the relevance of the F0 rise at the content word onset and taking into consideration the word-initial segments [13, 19, 20].

1.1. Predictions

It is expected that the manipulation of the cues will play out differently depending on segmental material. Taking into account previous descriptions of the intonational phonology and characteristics of consonants in French, as well as the literature on word segmentation in French we postulate the following exploratory hypotheses:

- H1 LACK OF ELBOW CAUSES DELAY: A flat F0 will lead to delayed word recognition in items starting with a nasal consonant or a voiced plosive + liquid. This is because voiced segments are carriers of F0 cues, and so the absence of an elbow and rise will be interpreted as the absence of a word beginning.
- H2 INFLUENCE OF SEGMENTAL ENVIRONMENT: Words starting with voiceless fricatives will be affected by the manipulation of F0 since their spectral energy is not manipulated and their acoustic cues will not match the surrounding pitch context. This effect should not be present when duration is manipulated instead.
- H3 INCREASED INITIAL DURATION SIGNALS WORD BOUNDARY: A lengthened content word-initial voiced or voiceless plosive will facilitate word recognition. This is because longer stop closure duration is interpreted as a word onset in French.

1.2. Method

Participants were invited to a forced choice lexical decision task where for each sequence they heard, they had to decide whether or not this was a real French word. The experiment was carried out using the software PsychoPy [27], version 3.2.1.

1.2.1. Participants

64 French listeners (age range 18-28 years) were recruited from the Aix Marseille University community in Aix-en-Provence, France.

1.2.2. Materials

A native female speaker of Hexagonal French was recorded producing all items in a carrier phrase and target APs were excised for acoustic manipulation. Stimuli consisted of a short AP containing the definite article followed by the target word e.g.: *la mission* ('the mission'). Stimuli were divided into four subgroups according to a sonority distinction related to the word-initial phonemes. The groups formed consisted of (i) voiced nasals [n], (ii) voiced plosive + liquid [bl], (iii) voiceless plosive + liquid [pl] and (iv) voiceless fricatives [s]. Note that in French phonologically voiced plosives has voicing in the closure, as opposed to other languages such as English or German

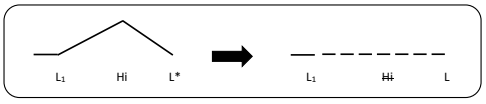

Figure 1: *Schematic representation of resynthesis manipulation carried out on fundamental frequency. The initial LHiL* pattern is resynthesised into a flat pattern with no peaks or elbows.*

where the phonological contrast between e.g. /b/ and /p/ is due to a difference between short and long lag in voice onset time and there is no voicing in either case. Target items consisted of five disyllabic, five trisyllabic, and five quadrisyllabic French words that together formed an AP together with a preceding article[1]. The set of stimuli further included non-word items for disyllables, trisyllables, and quadrisyllables (3x30), as well as an additional 87 fillers. The experiment included three conditions corresponding to the type of resynthesis manipulation carried out:

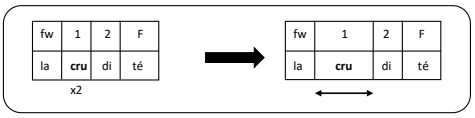

Figure 2: *Schematic representation of resynthesis manipulation carried out on duration values. The word-initial syllable is resynthesised into a longer syllable while no other syllables are manipulated.*

(i) **Manipulation of fundamental frequency (ResynthF0)**, the aim of this manipulation was to create a continuous flat F0 pattern with no elbows or peaks. Figure 1 shows a schematic representation of the resynthesis process applied to all stimuli (words, non-words, and fillers). Resynthesis was carried out using the PSOLA technique in Praat [28] using scripts to automate the procedure. This procedure led to F0 mean values that were higher for onsets of the type voiced plosive + liquid (approx. +11 Hz) but lower for nasal (approx. -8 Hz). (ii) **Manipulation of duration (ResynthDur)**, the aim of this manipulation was to change syllabic duration patterns. The first syllable of the content word, which L_1 consistently straddles, was lengthened to make word-initial syllable the longest in the AP. This manipulation specifically targeted the duration of the content word-initial syllable, increasing it by a stretch factor of two, whereas the preceding article and remaining syllables were not modified. A schematic representation is given in Figure 2. This procedure ensured that the portion of the consonant and the vowel were increased proportionally to the original duration values. In this case, the original F0 pattern was retained and only duration was modified. (iii) **Combination of F0 and duration manipulation (ResynthF0Dur)**; the third manipulation included the lengthening of the word-initial syllable as well as the flattening of the F0 curve.

1.3. Data analysis

To control for frequency effects that could affect response time (RT) we used *Lexique* an online French lexical database (lexique.org), which provides lexical frequency measures based on

[1] One AP of the type /pl/ was excluded because it was produced with the wrong article.

film subtitles [29, 30]. Seven words for which no frequency scores information was available were excluded (2 = voiced plosive + liquid; 1 = voiceless plosive + liquid; 2 = nasal; 2 = voiceless fricative). Responses given before the end of the first content word syllable, or after 5000 ms were considered unrealistic and discarded. Responses given with RTs greater than three times the standard deviation were also discarded. A total of 3% of the data were excluded from the analyses. Since stimuli duration varied depending on the manipulation carried out, we calculated the RT excluding the duration of the stimuli, and used these values in statistical analyses.

1.4. Statistical analysis

A linear mixed effects model was used to investigate the normalised RT using the packages `lme4` [31] and `lmerTest` [32]. The model included Condition (ResynthF0,ResynthDur,ResynthF0Dur), Onset segment type (nasal, voiceless fricative, voiceless plosive + liquid, voiced plosive + liquid), Lexical frequency (score), and Number of syllables in the content word (2,3,4) as fixed factors, and Participant and Stimulus as random intercepts, plus additional random slopes for Condition and Lexical frequency. Additionally, we used the step function for backward model selection and pairwise comparisons were obtained using `emmeans` [33].

2. Results

Figure 3 shows normalised RT values according to onset type of the content word, the three manipulations carried out, and number of syllables in the content word.

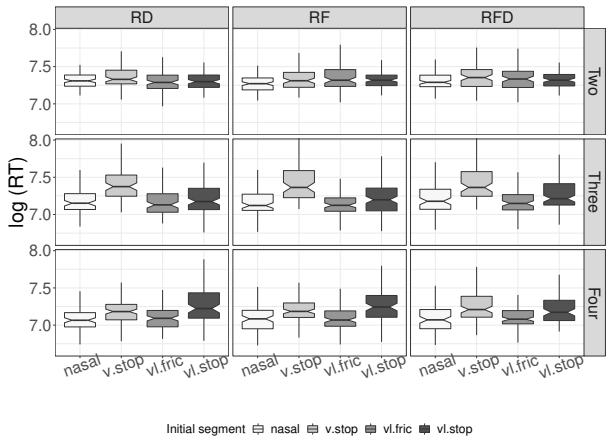

Figure 3: *Box plots of normalised RT measurements in words starting with varying onset types (voiced plosive + liquid, voiceless plosive + liquid, nasal, voiceless fricative) according to resynthesis manipulation (ResynthF0, ResynthDur, ResynthF0Dur), according to number of syllables in the content word (two, three, and four syllables)*

The log values of RT measurements were fitted to a linear mixed effects model to investigate the experimental manipulations (N=3229). The model retained after selection included Condition, Onset segment type, Number of syllables in the word, and Lexical frequency as fixed factors, and Stimulus and Participant as random intercepts, plus additional random slopes for Lexical frequency. Results show that the factors Condition (ResynthF0Dur, Est. 0.02 $\pm$ 0.01 log(RT), p= 0.04), Onset segment type (voiced plosive, Est. 0.1 $\pm$ 0.02 log(RT), $p<0.0001$), Number of syllables in the word (Four, Est. -0.2 $\pm$ 0.02 log(RT), $p<0.0001$), and Lexical frequency (Est. -0.01 $\pm$ 0.002 log(RT), $p<0.0001$) had a statistically significant effect on RT. To better understand how condition and initial segments interact factor comparisons were obtained. Table 1 summarises relevant contrasts of interest.

The results show that as expected from H1, the manipulation of F0 at the word onset had a great impact on word onsets with a voiced plosive + liquid where the cue is relevant. As shown in Table 1 in condition ResynthF0Dur, when the number of syllables is equal, words with a voiced plosive + liquid onset were recognised significantly more slowly in comparison to words starting with a nasal or voiceless fricative consonant. However, the strongest effect was noticeable when both duration and F0 were manipulated in voiced plosives. Contrary to predictions regarding nasal consonants, however, no evidence was found that the missing elbow at the word onset caused a delay in the recognition of words. In fact, as shown in Figure 3, words starting with nasal consonants were retrieved at consistently faster rates than words with other types of onsets. No significant effect was found when comparing words starting with a voiceless fricative across conditions. Contrary to expectations with respect to the influence of segmental environment in H2, we find that words starting with a voiceless fricative are not retrieved more slowly in the ResynthF0 or at faster pace in ResynthDur conditions. In addition, contrary to expectations about the role of the duration manipulation in signalling word boundaries in H3, words starting with a voiceless stop are recognised at a slower rate than words starting with e.g., a voiceless fricative. Note however, that this difference was significant when the onset voiceless plosive + liquid was in the ResynthF0Dur condition while the voiceless fricative was in ResynthDur or ResynthF0. The same comparison between onsets of voiceless fricatives compared to voiceless plosive + liquid, in equal conditions, i.e., ResynthF0, ResynthDur or ResynthF0Dur, did not yield a statistically significant difference. Thus, although we find that onsets with voiceless fricatives are not equally impacted by the manipulation of duration or F0, and are recognised faster, the type of manipulation had an impact. Finally, words with three and four syllables were recognised significantly faster than words with two syllables, when starting with the same consonant, in the same condition, and across conditions.

3. Discussion

The results of this study confirm that French listeners are sensitive to duration and F0 cues present at the onset of words. Although the manipulations included in the experiment were performed at the level of the AP and the syllable the evidence shows that the recognition of words was affected differently depending on type at the onset of the word. As predicted by the LACK OF ELBOW CAUSES DELAY hypothesis, words starting with a voiced plosive + liquid were affected. Note however, that this effect appeared when both duration and F0 had been manipulated. This result was consistent when comparing word recognition of words starting with voiceless fricatives or nasal consonants to words stating with the voiced plosive + liquid consonant cluster. It may be that the perception of French voiced plosives relies on duration and F0 cues. Thus, a manipulation of these two cues had the biggest impact on words starting with voiced plosive + liquid.

Despite the evidence reported in previous research [13],

Factor 1	**Factor 2**	**Contrast**	Estimate	SE	t-value	***p*-value**
Syllables	**Condition**	**Segment**				
two	ResynthF0Dur	nasal - v.stop	-0.1	0.02	-5	= .0003
two	ResynthF0Dur	v.stop - vl.fric	0.1	0.02	5.1	= .0002
four	ResynthF0	nasal - v.stop	-0.1	0.02	-5	= .0003
four	ResynthF0 - ResynthF0Dur	nasal - v.stop	-0.1	0.02	-5.4	$<$.0001
	Segment	**Condition**				
two	v.stop	ResynthF0 - ResynthF0Dur	-0.02	0.01	-2	= .9 (n.s)
two	v.stop	ResynthDur - ResynthF0	-0.001	0.01	-0.2	= 1 (n.s)
two	vl.fric - vl.stop	ResynthF0 - ResynthDur	-0.1	0.02	-3	= 0.3 (n.s)
two	vl.fric - vl.stop	ResynthDur - ResynthF0Dur	-0.1	0.02	-4	= 0.03
two	vl.fric - vl.stop	ResynthF0 - ResynthF0Dur	-0.1	0.02	-3.9	= 0.04
Condition	**Segment**	**Syllables**				
RsesynthDur	v.stop	two - four	0.2	0.02	9.5	$<$.0001

Table 1: *Results of relevant Tukey corrected factor comparisons for log RT.*

which suggests that F0 and especially the AP initial rise L_1Hi play an important role in word segmentation in words starting with nasal consonants, no evidence was found for words starting with a nasal consonant being significantly affected by the missing word-initial elbow. In fact, it was found that words starting with nasal consonants were systematically recognised at a faster rate than words starting with a voiced plosive + liquid, all other things being equal. Thus, contrary to our expectations we did not find any evidence for the LACK OF ELBOW CAUSES DELAY hypothesis in relation to words whose onset was a nasal consonant. Note that the effect was the strongest when duration and F0 were manipulated in the voiced plosive + liquid cluster whereas only one cue (duration or F0) had been manipulated for the nasal onset. One possible explanation for this result is that although a missing elbow could impoverish recognition of words starting with nasal consonants, under the same circumstances, onsets of voiced plosives + liquid are impacted to a higher degree. This could mean that the prosodic information encoded at the segmental level in voiced plosives serves a double function, firstly, in these consonants, F0 and duration carry information relative to the initial rise, and secondly they carry information pertaining to the identity of the segment. Arguably, the same does not apply to nasal consonants which might be carriers of prosodic information at the word onset but whose segmental identity is not equally compromised by the acoustic manipulation.

No evidence was found in favour of the INFLUENCE OF SEGMENTAL ENVIRONMENT hypothesis which predicted that in the ResynthF0 condition words starting with a voiceless fricative would be more difficult to identify than in the ResynthDur condition. Some evidence was found that words starting with voiceless plosives were recognised at a slower rate than e.g., words starting with voiceless fricatives, speaking against of the INCREASED INITIAL DURATION SIGNALS WORD BOUNDARY hypothesis. However, this was only the case when the word-initial consonant cluster voiceless plosive + liquid was in a condition where both duration and F0 had been manipulated, whereas only duration or F0 had been manipulated for the voiceless fricative. This result suggests that a manipulation of duration does not facilitate word recognition of stimuli starting with a voiceless plosive + liquid. It is likely that the combined manipulation of duration and F0 masked the identity of the initial consonant cluster (voiceless plosive + liquid) and this led to the delayed word recognition. Since word recognition did not significantly differ between words starting with voiceless plosive + liquid and voiceless fricatives when the acoustic manipulation was equal, this suggests that prosodic factors other than cues related to segmental identity hindered word recognition. Possibly the missing rise L_1Hi and the exaggerated word-initial syllable in the stimuli had an equally strong impact on phrase-level prosody and word recognition making it impossible to disentangle whether one segment type or the other is easier to recover. We find that the INCREASED INITIAL DURATION SIGNALS WORD BOUNDARY hypothesis can neither be clearly confirmed nor discarded. To clearly determine the impact of a duration manipulation on segments such as voiceless plosives, an experiment targeting only the word-initial segment rather than the syllable could be employed. However, this would not address the research question of this study which seeks to shed light on the relationship between segments and the intonational pattern of the AP initial rise present in French phrase (word) level prosody. Arguably, in the given segmental context, prosodic cues at the word level, such as the initial rise or syllabic duration, take precedence over prosodic information like duration encoded at the segmental level of the word-initial voiceless plosive.

4. Conclusion

An online experiment was carried out with French listeners to determine whether a manipulation of acoustic cues to word onsets would affect the recognition of words. The results of this study show that manipulations had an effect on word recognition depending on the type of manipulation and segment type at word onset. It was found that consonant clusters of the type voiced plosive + liquid [bl] were affected the most when both duration and F0 cues had been manipulated. Contrary to predictions, a manipulation of F0 cues on nasal consonants in word onsets did not have a significant effect on word recognition. We attribute this result to a double function of acoustic cues for onsets of the type [bl], namely F0 and duration are cues to both the segmental identity and prosodic information of the initial rise (L_1Hi). Finally, it was also established that the effects of acoustic manipulation were stronger on short words of two syllables than on words of three or four syllables.

5. References

[1] M. D. Tyler and A. Cutler, "Cross-language differences in cue use for speech segmentation," *The Journal of the Acoustical Society of America*, vol. 126, no. 1, pp. 367–376, 2009.

[2] A. Tremblay, C. E. Coughlin, C. Bahler, and S. Gaillard, "Differential contribution of prosodic cues in the native and non-native segmentation of French speech," *Laboratory Phonology*, vol. 3, no. 2, pp. 385–423, 2012.

[3] A. Tremblay, M. Broersma, C. E. Coughlin, and J. Choi, "Effects of the native language on the learning of fundamental frequency in second-language speech segmentation," *Frontiers in psychology*, vol. 7, p. 985, 2016.

[4] J. M. Toro, N. Sebastián-Gallés, and S. L. Mattys, "The role of perceptual salience during the segmentation of connected speech," *European Journal of Cognitive Psychology*, vol. 21, no. 5, pp. 786–800, 2009.

[5] S. Kim, M. Broersma, and T. Cho, "The use of prosodic cues in learning new words in an unfamiliar language," *Studies in Second Language Acquisition*, vol. 34, no. 3, pp. 415–444, 2012.

[6] A. Cutler, "Lexical stress," in *The Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds. Oxford: Blackwell, 2005, ch. 11, pp. 264–289.

[7] S.-A. Jun and C. Fougeron, "Realizations of accentual phrase in French intonation," *Probus*, vol. 14, no. 1, pp. 147–172, 2002.

[8] A. Michelas and S. Dufour, "Are prosodic variants stored in the French mental lexicon?" *Experimental Psychology*, 2019.

[9] S.-A. Jun and C. Fougeron, "The accentual phrase and the prosodic structure of French," in *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 2, 1995, pp. 722–725.

[10] ——, "A phonological model of French intonation," *Intonation: Analysis, modeling and technology*, pp. 209–242, 2000.

[11] P. Welby, "French intonational structure: Evidence from tonal alignment," *Journal of Phonetics*, vol. 34, no. 3, pp. 343–371, 2006.

[12] P. Welby and H. Lœvenbruck, "Anchored down in Anchorage: Syllable structure and segmental anchoring in French," *Italian Journal of Linguistics/Rivista di linguistica*, vol. 18, pp. 74–124, 2006.

[13] P. Welby, "The role of early fundamental frequency rises and elbows in French word segmentation," *Speech Communication*, vol. 49, no. 1, pp. 28–48, 2007.

[14] E. Spinelli, N. Grimault, F. Meunier, and P. Welby, "An intonational cue to word segmentation in phonemically identical sequences," *Attention, Perception, & Psychophysics*, vol. 72, no. 3, pp. 775–787, 2010.

[15] A. C. M. Rietveld, "Word boundaries in the French language," *Language and Speech*, vol. 23, no. 3, pp. 289–296, 1980.

[16] M.-H. Banel and N. Bacri, "On metrical patterns and lexical parsing in French," *Speech Communication*, vol. 15, no. 1-2, pp. 115–126, 1994.

[17] M.-H. Banel, U. H. Frauenfelder, and P. Perruchet, "Contribution des indices métriques à l'apprentissage d'un langage artificiel," *XXIIèmes Journées d'Etude sur la Parole*, 1998.

[18] O. Bagou and U. H. Frauenfelder, "Stratégie de segmentation prosodique: rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle," in *Acte des XXVIèmes Journées d'Étude de la Parole*, 2006.

[19] J. E. Andruski, S. E. Blumstein, and M. Burton, "The effect of subphonetic differences on lexical access," *Cognition*, vol. 52, no. 3, pp. 163–187, 1994.

[20] B. McMurray, M. K. Tanenhaus, and R. N. Aslin, "Gradient effects of within-category phonetic variation on lexical access," *Cognition*, vol. 86, no. 2, pp. B33–B42, 2002.

[21] A. Christophe, "Role de la prosodie dans la segmentation en mots," Ph.D. dissertation, Paris, EHESS, 1993.

[22] A. Christophe, E. Dupoux, J. Bertoncini, and J. Mehler, "Do infants perceive word boundaries? an empirical study of the bootstrapping of lexical acquisition," *The Journal of the Acoustical Society of America*, vol. 95, no. 3, pp. 1570–1580, 1994.

[23] O. Niebuhr, "At the edge of intonation: The interplay of utterance-final F0 movements and voiceless fricative sounds," *Phonetica*, vol. 69, no. 1-2, pp. 7–27, 2012.

[24] P. Welby and O. Niebuhr, "The influence of F0 discontinuity on intonational cues to word segmentation: a preliminary investigation," in *Speech Prosody*, 2016.

[25] ——, "Segmental intonation information in French fricatives," in *19th International Congress of Phonetic Sciences*, 2019.

[26] O. Niebuhr, "On the perception of "segmental intonation": F0 context effects on sibilant identification in German," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2017, no. 1, p. 19, 2017.

[27] J. W. Peirce, "PsychoPy—psychophysics software in Python," *Journal of neuroscience methods*, vol. 162, no. 1-2, pp. 8–13, 2007.

[28] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 6.0.26)[computer program]," 2017.

[29] B. New, C. Pallier, M. Brysbaert, and L. Ferrand, "Lexique 2: A new French lexical database," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 3, pp. 516–524, 2004.

[30] B. New, M. Brysbaert, J. Veronis, and C. Pallier, "The use of film subtitles to estimate word frequencies," *Applied psycholinguistics*, vol. 28, no. 4, pp. 661–677, 2007.

[31] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[32] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: tests in Linear Mixed Effects Models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[33] R. Lenth and M. Herve, "Emmeans: Estimated marginal means, aka least-square means. R package version 1.1. 2," 2019.