



HAL
open science

On the relative weight of pitch and durational cues in the perception of accentuation by French listeners

Amandine Michelas, Sophie Dufour

► **To cite this version:**

Amandine Michelas, Sophie Dufour. On the relative weight of pitch and durational cues in the perception of accentuation by French listeners. *Journal of the Acoustical Society of America*, 2024, 156 (5), pp.3169-3182. 10.1121/10.0034368 . hal-04789786

HAL Id: hal-04789786

<https://hal.science/hal-04789786v1>

Submitted on 17 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/385769939>

On the relative weight of pitch and durational cues in the perception of accentuation by French listeners

Article in *The Journal of the Acoustical Society of America* · November 2024

DOI: 10.1121/10.0034368

CITATIONS

0

READS

20

2 authors:



Amandine Michelas
Aix-Marseille University

49 PUBLICATIONS 385 CITATIONS

SEE PROFILE



Sophie Dufour

29 PUBLICATIONS 77 CITATIONS

SEE PROFILE

On the relative weight of pitch and durational cues in the perception of accentuation by French listeners

Amandine Michelas^{1,a} and Sophie Dufour¹

¹ *Aix Marseille Univ, CNRS, LPL, 13 100 Aix-en-Provence, France*

Accentuation is encoded by both durational and pitch cues in French. While previous research agree that the sole presence of pitch cues is sufficient to encode accentuation in French, the role of durational cues is less clear. In four cue-weighting accent perception experiments, we examined the role of pitch and durational cues in French listeners' perception of accentuation. French listeners were tested on acoustic continua ranging from an unaccented first syllable [depla'se] 'moved' to an accented first syllable ['de] [pla'se] 'a dice placed'. The continua were obtained by manipulating pitch and durational cues either separately or conjointly. In Experiments 1 and 2, participants identified each step of the continuum by choosing between two possible segmentations while in Experiments 3 and 4, participants performed an AX discrimination task on two adjacent steps. Results indicate that participants' performance was better when pitch cues were varied compared to when durational cues were varied in both the identification and discrimination tasks. In addition, while participants always benefited from the additional presence of pitch cues, participants did not consistently benefit from the additional presence of durational cues. Taken together, these results suggest that pitch cues are a better marker of French accentuation than durational cues.

^a Email: amandine.michelas@univ-amu.fr

1 I. INTRODUCTION

2 When listeners process the speech signal, they identify the sounds that compose it with an
3 impressive ease. However, this identification process involves considering a very large number of
4 different acoustic information simultaneously. For instance, to differentiate the English sounds /b/
5 and /p/, English listeners have to process simultaneously VOT information (the amount of time
6 between the stop release and onset of voicing,) duration of the stop closure and fundamental
7 frequency among many other acoustic cues (Lisker, 1986). Whereas all these acoustic cues play a role
8 in the consonant identification process, they are not perceptually equivalent: VOT information plays
9 a more important role than fundamental frequency for English listeners (Abramson & Lisker, 1985).
10 This has been referred to as perceptual cue weighting (e.g., Francis et al., 2008; Schertz & Clare,
11 2020; Idemaru et al. 2012). Perceptual cue weighting does not apply only to segmental information
12 (i.e., phonemes) but also to suprasegmental information such as stress.

13
14 In stress-accent languages such as English or Dutch, polysyllabic words contain a syllable which
15 is more acoustically salient than the other syllables and which can change the meaning of that word
16 (e.g., the English word pair *present* ‘gift’ vs. *present* ‘to give, provide, or make something known’). The
17 acoustic realization of this ‘stressed’ syllable results of a combination of several acoustic parameters
18 (including vowel quality, intensity and durational indices¹) and the relative weight of each parameter
19 in lexical stress marking differs according to languages. For instance, while in English, the stressed
20 syllable is primarily cued by vowel quality (Gay, 1978; Lindblom, 1963), in Dutch, syllable
21 lengthening is the most important cue to lexical stress (Sluijter & van Heuven, 1996). This language-
22 dependent weighting of cues affects the perception of stress, with English listeners’ stress perception
23 depending more on vowel quality (Zhang & Francis, 2010). while Dutch listeners rely primarily on
24 syllable lengthening to perceive stress (van Heuven & de Jonge, 2011). Unlike English and Dutch,

25 the acoustic highlighting of the syllable does not occur at the word level in French but occurs at a
26 larger level than the word which is called the Accentual Phrase (AP). This unit corresponds
27 minimally to a lexical word and to all the function words that this word governs (e.g., Jun &
28 Fougreon 2002; Delais-Roussarie et al. 2015). The last syllable of the AP is always accented² in
29 French. It is always longer than non-final syllables and typically associated to a fundamental
30 frequency (f₀) rise (e.g., Jun & Fougeron, 2002; Welby, 2006; Michelas & d’Imperio, 2012; Delais-
31 Roussarie et al. 2015; Smith et al. 2019). Since the ‘accented’ syllable always marks the end of APs
32 and thus the end of words occurring at the end of APs in French, French listeners are able to exploit
33 the conjoint presence of durational and pitch cues to accent to segment the speech stream into
34 words (e.g., Rietveld, 1980; Bagou et al., 2002; Christophe et al. 2004; Bagou & Frauenfelder, 2006;
35 Tremblay et al., 2012, 2016; Bagou & Frauenfelder, 2018). The sole presence of pitch cues in the
36 absence of durational cues also appears as a robust cue to segment speech into words (Tremblay et
37 al., 2012, 2016). However, the reverse is not true and the contribution of the sole presence of
38 durational cues in the absence of pitch cues is less clear. While some studies (Banel & Bacri, 1994;
39 Tyler & Cutler, 2009) showed that French speakers perceive a word boundary when cued by
40 duration alone, some other studies did not find such an effect (Bagou & Frauenfelder, 2018). In this
41 study we thus re-examine the role of pitch and durational cues in accentuation marking in French by
42 testing their contribution in French listeners’ perception of accent.

43

44 To date, most studies that have examined the contribution of durational and pitch cues in
45 French listeners’ perception of accent have done so by examining the effect of these cues in speech
46 segmentation. Among these studies, several ones indicated that French listeners use the combination
47 of the f₀ rise and the syllable lengthening located at the end of APs to segment the speech stream
48 into words (Christophe et al., 2004; Tremblay et al., 2012, 2016; Bagou & Frauenfelder, 2018). For

49 example, in a word monitoring task, Christophe et al. (2004) asked participants to detect a target
50 word (e.g., *chat* ‘cat’) in sentences like [*Le chat grincheux*] ‘the grumpy cat’ in which only the word
51 *grincheux* bore an accent cued by both an f0 rise and syllable lengthening, or in sentences like [*Le chat*]
52 [*grimpaît*] ‘the cat climbed up’ in which both the words *chat* and the word *grimpaît* were accented.
53 They found slower detection times in [*Le chat grincheux*] than in [*Le chat*] [*grimpaît*]. This slower
54 detection of *chat* in [*Le chat grincheux*] has been interpreted as resulting from competition between *chat*
55 and its competitor *chagrin* ‘sadness’. In this case, since there is no acoustic cue on *chat* to indicate its
56 end, the word *chagrin* remains active and competes with the word *chat*. The results of Christophe et
57 al.’s thus suggest that French listeners insert a word final boundary each time they encounter an
58 accented syllable marked by a combination of durational and pitch cues. The conjoint contribution
59 of durational and pitch cues to accent marking and their role in speech segmentation has been
60 replicated in several subsequent studies using various paradigms such as word-monitoring tasks
61 (Tremblay et al. 2012, 2016) or artificial-language learning experiments (Bagou & Frauenfelder,
62 2018).

63
64 Other studies have shown that French listeners are not only able to use the combination of pitch
65 and durational cues to segment speech into words, but they are also able to exploit the sole presence
66 of pitch cues to do so (Tremblay et al., 2012; Tremblay et al., 2016). In a word-monitoring task,
67 Tremblay et al. (2012) presented to French listeners sequences composed of a monosyllabic word
68 and a multisyllabic word (e.g., *rat tondu* ‘the shaved rat’) and asked them to detect a bisyllabic word
69 which corresponds to the combination of the monosyllabic word and the first syllable of the
70 following multisyllabic word (e.g., *raton* ‘young rat’). In this experiment, the stimuli were
71 resynthesized so that the monosyllabic word (e.g., *rat*) was unaccented or received an accent marked
72 by (1) pitch cues but not durational cues, (2) durational cues but not pitch cues or (4) both

73 durational and pitch cues. The results revealed that French listeners use pitch cues to segment
74 speech into words in the absence of durational cues so that they detected the word *raton* to a greater
75 extent in the condition in which the monosyllabic word *rat* was unaccented in comparison to when it
76 bore an accent marked by pitch cues only.

77
78 Studies using artificial-learning experiments similarly showed that French listeners are able to use
79 pitch cues in the absence of durational cues to segment an artificial language into words (Bagou &
80 Frauenfelder 2002, 2006, 2018; Tyler & Cutler, 2009; Tremblay et al. 2017). For example, Bagou &
81 Frauenfelder (2018) presented to participants continuous sequences of consonant-vowel syllables
82 forming artificial words (e.g., /ʁun.sət.fɛb#.ʁyf.dap#.ʁɛb.nyv.zak#.../) and asked them to
83 remember the words of the artificial language. The last syllable of the target artificial word was either
84 (1) completely unaccented, (2) marked by syllable lengthening (the syllable's intrinsic duration was
85 increased by 30%), (3) marked by an f₀ rise (the end of the syllable was raised of 30Hz compared to
86 the beginning of the syllable), or (4) marked by both durational and pitch cues. After having exposed
87 to the artificial language, participants were presented with pairs of items. For each pair, one member
88 was a word from the artificial language while the other was a sequence of syllables that crossed a
89 word boundary in the artificial language. Participants had to decide whether the word of the artificial
90 language was the first or the second member of the pair. The results show that participants gave
91 more correct responses when an f₀ rise was associated to the last syllable of artificial words
92 compared to when the artificial words were unaccented. By contrast, participants did not perform
93 better when the last syllable of artificial words were lengthened compared to when the words were
94 unaccented. The results of this study indicate that the presence of a final f₀ rise facilitated word
95 learning whereas the presence of durational cues did not.

96

97 The effect of durational cues in French speech segmentation are not so clear and different
98 studies come to contradictory conclusions. While the study of Bagou & Frauenfelder (2018) suggests
99 that French listeners did not exploit durational cues in the absence of final pitch rise to segment an
100 artificial language into words, other studies using the same paradigm revealed that French listeners
101 appear to use vowel lengthening as a cue to word-final boundaries. In their artificial learning
102 experiment, Tyler & Cutler (2009) presented to participants continuous sequences of consonant-
103 vowel syllables composed of three and four-syllable artificial words. While the pitch contour was
104 kept flat for all words (f_0 was set to 120Hz for all words), these words were demarcated by (1) no
105 cue, (2) the lengthening of their initial vowel or (3) the lengthening of the final vowel. In each
106 lengthening condition, the vowel was lengthened by 60 ms. After having been exposed to the
107 artificial language, participants had to perform a forced choice test during which they have to decide
108 which member of a pair of items was a word of the artificial language. The results show that
109 participants gave more correct responses when the final vowel was lengthened rather when there
110 was no cue. The results of Tyler & Cutler (2009) thus suggest that French listeners interpret vowel
111 lengthening in the absence of pitch rise as a cue to word-final boundary.

112

113 Such a conclusion was also reached in an older study examining the effect of syllable lengthening
114 on speech segmentation in French (Banel & Bacri, 1994). In this study, the authors presented
115 phonemically ambiguous sequences of two syllables to French listeners and asked them to determine
116 whether they heard “one word” (e.g., *bagage* ‘luggage’) or “multi-word” (e.g. *bas gage* ‘low pledge’).
117 While the pitch contour was kept flat and intensity fixed for all stimuli, these ambiguous sequences
118 differed in the relative duration of the first syllable compared to the second syllable (i.e, for *bagage*: 1st
119 short syllable – 2nd long syllable; for *bas gage*: 1st long syllable – 2nd short syllable) with a duration
120 difference between the first syllable and the second syllable of 35% in the two versions. The authors

121 found that French listeners were more likely to interpret these sequences as containing one word
122 (e.g., *bagage*) when the second syllable was lengthened, while they were more likely to interpret them
123 as containing two words (e.g., *bas gage*) when the first syllable was lengthened. Hence, even in the
124 absence of pitch cues, French listeners were able to use syllable lengthening to segment the speech
125 stream into words.

126

127 Given that some studies (Banel & Bacri, 1994; Tyler & Cutler, 2009) showed that French
128 speakers perceive a word boundary when cued by duration alone while other studies did not show
129 such an effect (Bagou & Frauenfelder, 2018), in the present study, we tested the hypothesis that
130 pitch cues would be a better marker than durational cues to accentuation in that language. To do so,
131 we used a series of four cue-weighting accent perception experiments during which participants
132 heard acoustic continua ranging from an unaccented first syllable [depla'se] 'moved' to an accented
133 first syllable ['de] [pla'se] 'a dice placed'. The stimuli were manipulated in eight steps from the
134 unaccented syllable to the accented one, manipulating the two main acoustic correlates of
135 accentuation in French: syllable lengthening and f₀ rise. In experiments 1 and 2, participants
136 completed an identification task in which they had to determine whether the sequence they heard
137 correspond to one word (i.e., *deplacé* 'moved') or two words (i.e., *dé#placé* 'dice placed'). In
138 experiments 3 and 4, participants completed a discrimination task in which they heard two
139 continuous steps on the continuum and had to determine whether the two steps were identical or
140 different. In both types of tasks, duration and pitch cues were manipulated either separately to test
141 their independent contribution in accentuation marking (Experiments 1 and 3), or at the same time
142 in order to determine whether French listeners benefit from their conjoint presence (Experiments 2
143 and 4). If pitch cues are a better marker of accentuation than durational cues, French listeners would
144 rely more on this type of cue to identify and discriminate lexical sequences such as [depla'se] vs. ['de]

145 [pla'se]. Moreover, if French listeners benefit from the co-presence of durational and pitch cues to
146 perform identification and discrimination tasks, in Experiments 3 and 4, we expect better
147 performance when the two cues will be varied compared to when only one cue will be varied.
148

149 **II. EXPERIMENT 1**

150 Experiment 1 used a two forced-choice identification task in which pitch and durational cues
151 were manipulated separately.
152

153 **A. Method**

154 ***1. Participants***

155
156 Sixty monolingual native speakers of French (45 females, 15 males) between 18 and 45 years old
157 (mean=24.2, SD=3.4) participated in the experiment. They all reported having no
158 neurological/hearing impairment or language disorders. None of them were professional musicians.
159 Since English (or German) are taught in the French educational system, they all reported having
160 basic knowledge in one of these languages. Some of them also reported having basic knowledge in
161 Italian, Spanish or Mandarin Chinese, but none spent more than 3 months abroad, and all have been
162 exposed to these languages after 10 years of age during middle school. The four experiments
163 developed in this article followed local ethical guidelines³.

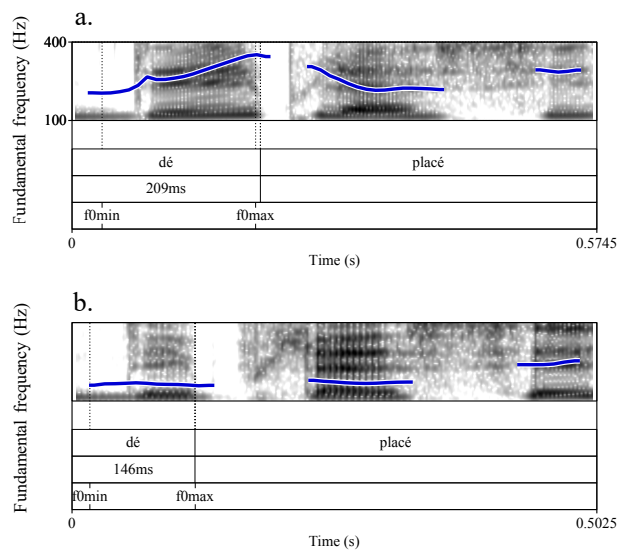
164

165

166 ***2. Speech materials***

167

168 The material used was derived from natural versions of the phonemic sequence /deplase/ which
 169 is lexically ambiguous between *déplacé* ‘moved’ and *dé#placé* ‘a dice placed’. A native speaker of French
 170 read both sequences one time each within carrier sentences. Within these sentences, the syllable
 171 /de/ was either accented (*[On a remarqué]_{AP} [le petit dé]_{AP} [placé sur cette case]_{AP}* ‘We noticed the small
 172 dice placed on this square’) or unaccented (*[Le pion]_{AP} [est déplacé]_{AP} [sur cette case]_{AP}* ‘The pawn is
 173 moved to this square’) depending on the location of the syllable within the AP. Acoustic analysis
 174 conducted on syllable /de/ confirmed that the syllable was longer and associated with an f0 rise in
 175 the *dé#placé* sequence (syllable length: 209 ms; f0 min: 204 Hz; f0 max: 351 Hz; % of f0 slope: 72%)
 176 than in the *déplacé* sequence (syllable length: 146 ms; f0 min: 159 Hz; f0 max: 159 Hz; % of f0 slope:
 177 0%)⁴. Figure 1 reports spectrograms, f0 curves and textgrids for the two natural versions of the
 178 /deplase/ sequences.
 179



180
 181 Figure 1: spectrograms, f0 curves and textgrids for the *dé#placé* sequence in a. and the *déplacé*
 182 sequence in b. F0 min and f0 max labels are only appropriate for *dé#placé* sequences in which the

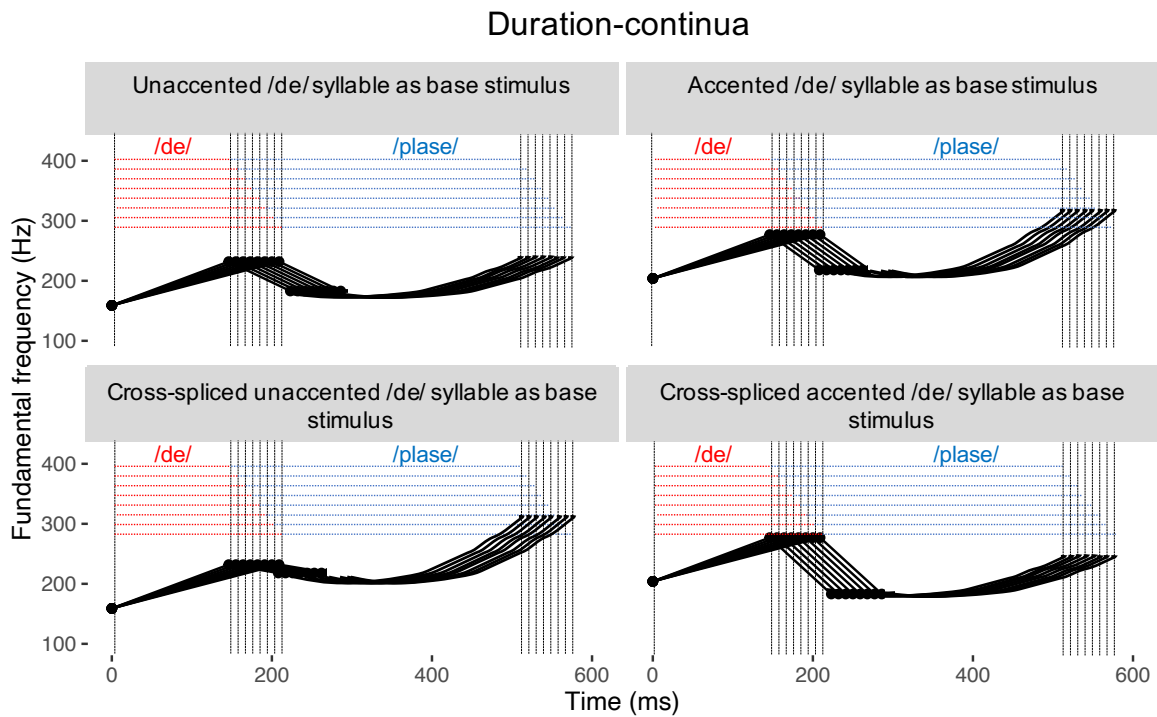
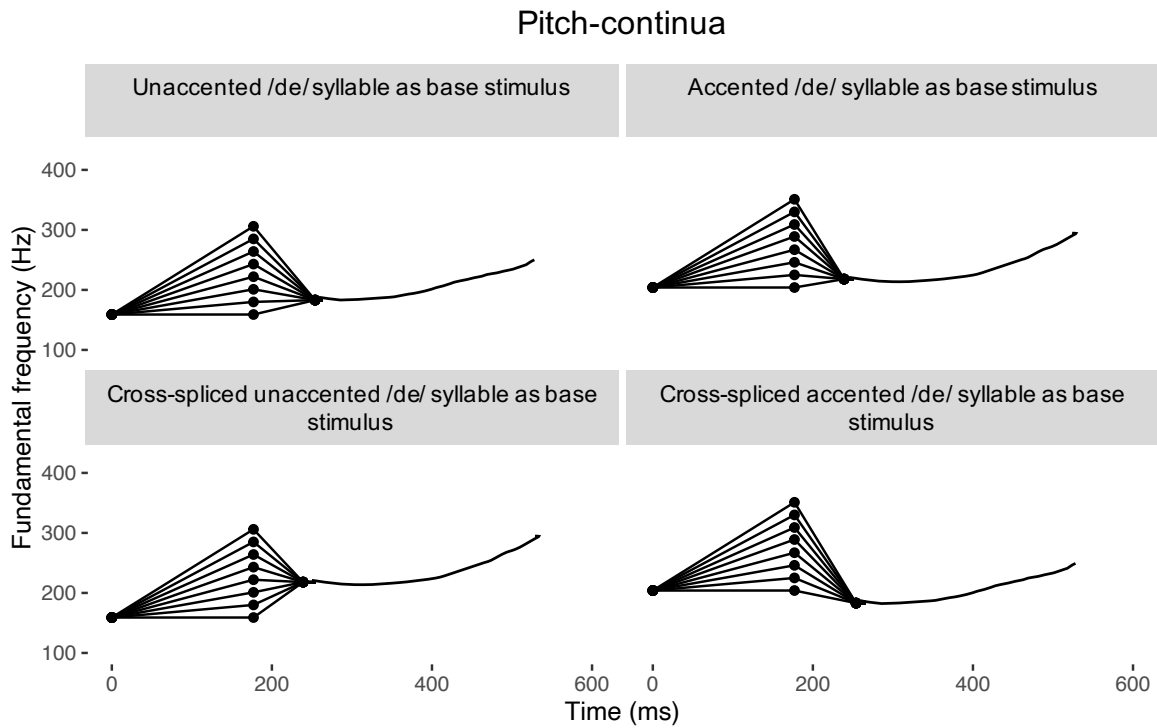
183 /de/ syllable is accented. For *déplacé* sequences in which the /de/ syllable is unaccented, it
184 corresponds to the first f0 value observed at the onset and the offset of the syllable.

185

186 Height continua of 8 steps (i.e., 4 f0-continua and 4 duration-continua) were generated from two
187 base stimuli in order to manipulate pitch cues and durational cues separately.

188

189 F0-continua: For the four 8-step f0-continua, the f0 rise associated to the /de/ syllable was
190 manipulated to create an acoustic continuum ranging from an unaccented first syllable [depla'se]
191 'moved' to an accented first syllable ['de] [pla'se] 'a dice placed' while the duration of this first
192 syllable was altered to make it ambiguous between an unaccented and accented syllable (see Table I
193 and Figure 2).



194

195 Figure 2. F0-continua and duration-continua from the four base stimuli.

196 TABLE I. Acoustic characteristics of the /de/ syllable for the f0-continua and duration-
 197 continua for steps 1-8.
 198 f0-continua

f0-continua																
Base stimuli in which the /de/ syllable was originally unaccented									Base stimuli in which the /de/ syllable was originally accented							
Steps	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
f0 min (Hz)	159	159	159	159	159	159	159	159	204	204	204	204	204	204	204	204
f0 max (Hz)	159	180	201	222	243	264	285	306	204	225	246	267	289	309	330	351
Duration (ms)	177	177	177	177	177	177	177	177	177	177	177	177	177	177	177	177
Duration-continua																
Base stimuli in which the /de/ syllable was originally unaccented									Base stimuli in which the /de/ syllable was originally accented							
Steps	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
f0 min (Hz)	159	159	159	159	159	159	159	159	204	204	204	204	204	204	204	204
f0 max (Hz)	232	232	232	232	232	232	232	232	277	277	277	277	277	277	277	277
Duration (ms)	146	155	164	173	182	191	200	209	146	155	164	173	182	191	200	209

199
 200 For the first continuum, we used as base stimulus the natural *déplacé* sequence (see Figure 1b).
 201 For the second continuum we used as base stimulus the natural *dé#placé* sequence (see Figure 1a).
 202 Because as illustrated in Figure 1, the two /plase/ sequences had not the same acoustic
 203 characteristics, they can influence participants' decisions. To avoid that, we generated two others f0-
 204 continua in which the base stimulus was created by cross-splicing. For the third continuum, we
 205 cross-spliced the unaccented /de/ syllable with the /plase/ sequence of the accented /de/ syllable.
 206 In the same way, for the fourth f0-continuum, we cross-spliced the accented /de/ syllable with the
 207 /plase/ sequence of the unaccented /de/ syllable.

208
 209 For each of the four base stimuli, we generated a re-synthesis file that contains duration and
 210 pitch points at the beginning and the end of the /de/ syllable. For each base stimulus, the duration

211 of the /de/ syllable was set to 177 ms which corresponds to an ambiguous value between the
212 unaccented /de/ syllable (146 ms, see Figure 1) and the accented /de/ syllable (209 ms). To
213 generate the pitch manipulation and because f0 minima were not the same in the accented version
214 and in the unaccented version of the stimuli, the different steps were generated separately for the
215 accented base stimuli and for the unaccented base stimuli. For the two base stimuli with the
216 accented /de/ syllable, a continuum was created between 204 Hz and 351 Hz (see Table I and
217 Figure 2). For the two base stimuli with the unaccented /de/ syllable, a continuum was created
218 between 159 Hz and 306 Hz. As a result, for both the accented and unaccented /de/ syllable, we
219 obtained an increase of 147 Hz between the f0 min and the f0 max by step of 21 Hz (= 8 steps).
220 These 21Hz steps thus correspond to the difference between the f0 max and the f0 min obtained in
221 the accented /de/ syllable divided by the 8 steps of the continuum.

222 Pitch contours were obtained via interpolation between the first and the second pitch point.

223
224 Duration-continua: For the four 8-step duration-continua, the duration of the /de/ syllable was
225 manipulated to create an acoustic continuum ranging from an unaccented first syllable [depla'se]
226 'moved' to an accented first syllable ['de] [pla'se] 'a dice placed' while the f0 rise associated to this
227 syllable was altered to make it ambiguous between an unaccented and accented syllable (see Table I
228 and Figure 2).

229
230 As for f0-continua, for the first duration-continuum we used the natural *déplacé* sequence as base
231 stimulus (see Figure 1b) while the natural *dé#placé* sequence was used as base stimulus for the second
232 duration-continuum (see Figure 1a). Because as illustrated in Figure 1, the two /plase/ sequences
233 had not the same acoustic characteristics, they can influence participants' decisions. To avoid that,
234 we generated two others duration-continua in which the base stimulus was created by cross-splicing.

235 For the third continuum, we cross-spliced the unaccented /de/ syllable with the /plase/ sequence of
236 the accented /de/ syllable. In the same way, for the fourth duration-continuum, we cross-spliced the
237 accented /de/ syllable with the /plase/ sequence of the unaccented /de/ syllable.

238

239 For each of the four base stimulus, we generated a re-synthesis file that contains duration and
240 pitch points at the beginning and the end of the /de/ syllable. For each base stimulus, the f0
241 movement associated to the /de/ syllable was manipulated to correspond to an ambiguous f0
242 movement between the f0 rise observed for the accented /de/ syllable (see Figure 1a) and the f0
243 plateau observed for the unaccented /de/ syllable (see Figure 1b). To generate this pitch
244 manipulation and because f0 minima were not the same in the accented version and in the
245 unaccented version of the stimuli, we adjusted f0 maxima separately for the accented base stimuli
246 and for the unaccented base stimuli. For the two base stimuli with the accented /de/ syllable for
247 which the f0 minima corresponds to a value of 204 Hz, f0 maxima were set at a value of 277 Hz.
248 For the two base stimuli with the unaccented /de/ syllable for which the f0 minima corresponds to
249 a value of 150 Hz, f0 maxima were set at a value of 232 Hz. As a result, for both the accented an
250 unaccented /de/ syllable, we obtained an f0 rise of 73 Hz which corresponds to an ambiguous
251 movement between the f0 plateau observed for the natural unaccented /de/ syllable and the f0 rise
252 of 147 Hz observed for the natural accented /de/ syllable (see Table I). Pitch contours were
253 obtained via interpolation between the first and the second pitch points. To generate the durational
254 manipulations, the duration of the /de/ syllable was manipulated by creating a continuum between
255 146 ms (corresponding to the length of the unaccented /de/ syllable; see Figure 1b) and 209 ms
256 (corresponding to the length of the /de/ syllable; see Figure 1a) by step of 9 ms (= 8 steps). As a
257 result, these 9 ms steps correspond to the difference of duration between the unaccented syllable
258 and the accented syllable divided by the 8 steps of the continuum. Note that this 9 ms duration is

259 really close to the just-noticeable difference (JND) for segment duration which has been estimated
260 to 10ms in the literature (Klatt, 1976; Koffi, 2018).

261

262 All acoustic manipulations were done using the Pitch Synchronous Overlap and Add method
263 (PSOLA, Moulines & Charpentier, 1990; see also Michelas & d'Imperio, 2015; Petrone et al., 2017)
264 on Praat (Boersma & Weenink, 2024) which is known for the high degree of naturalness achieved.
265 RMS amplitude of all stimuli was equated to 50 dB SPL.

266

267 **3. Procedure**

268

269 Participants were tested in a sound attenuated booth of the *Laboratoire Parole et Langage*, and the
270 stimuli were presented over headphones at the sound level of 60 dBA for all participants.
271 Participants were informed that they would hear sequences containing either one word (*déplacé*) or
272 two words (*dé#placé*) and their task was to decide whether the sequence was extracted to the
273 sentence *Le pion est déplacé sur cette case* 'The pawn is moved to this square' or to the sentence *On a*
274 *remarqué le petit dé placé sur cette case* 'We noticed the small dice placed on this square'. They give their
275 responses on a button box placed in front of them. Half of them heard the stimuli from the four 8-
276 step f0-continua and the other half the stimuli from the four 8-step duration-continua. For each type
277 of continuum (f0 and duration) half of the participants were asked to press the left button for
278 "déplacé" and the right button for "dé#placé" and the reverse was true for the other half. For each
279 continuum, each step was presented ten times. As a result, each participant heard 320 trials
280 randomly presented (4 base stimulus * 8 steps * 10 repetitions). The experiment lasted
281 approximatively 30 minutes. An inter-trial interval of 1000 ms elapsed between each participant's

282 response and the beginning of the next stimulus. The participants began the experiment with a block
283 of 8 practice trials and no feedback was given to participants.

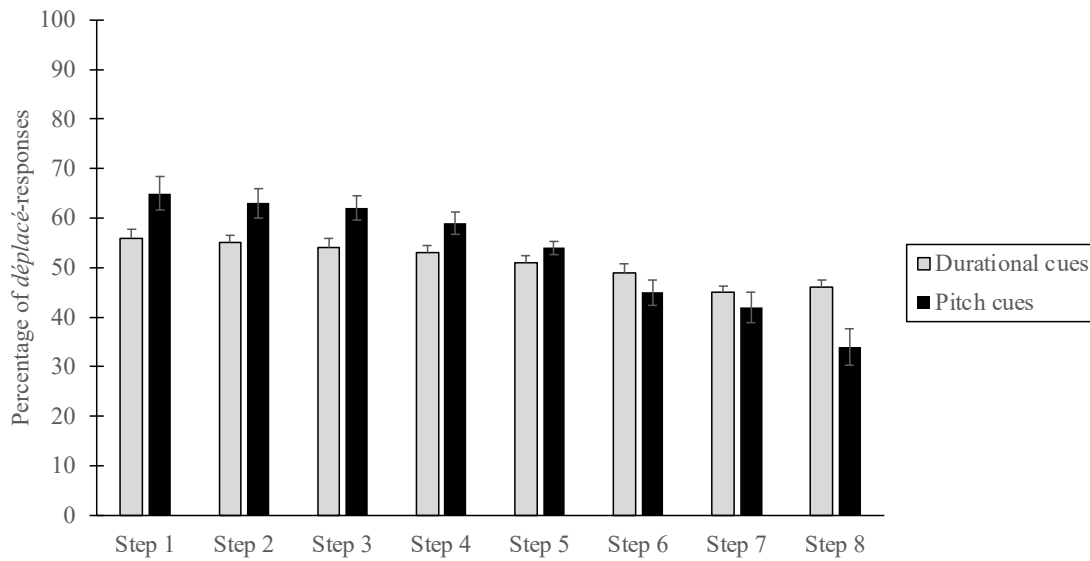
284

285 **B. Results and discussion**

286

287 The mean percentage of *déplacé* responses for pitch and durational cues are displayed in Figure 3.
288 The type of response (1 = *déplacé* response, 0 = *dé#placé* response) was first analyzed using a mixed-
289 effect regression model (lme4 package, Douglas et al., 2015, in R-studio statistics Version
290 2022.12.0+353, R core team, 2022) with a logistic linking function (Baayen et al., 2008). The model
291 included steps (1-8), type of cues (pitch vs. durational cues) and their interaction as fixed effects.
292 Participants were included as random intercept and a random slope was included for the within-
293 participant factor steps⁶. Since our data set was too large for the default fitting methods in glmer.nb,
294 it was necessary to implement the function options `nAG0=0` and `optimizer="nloptwrap"` for the
295 model-fitting process to converge (Bates et al., 2015). Significance of the main effect was assessed
296 using the Anova function from the car package (Fox, 1996). The model revealed a significant step x
297 cue interaction ($X^2 = 14.30$, $p < .05$) showing that the number of *déplacé* responses decreased more
298 from one extremity to the other of the continuum when pitch cues were varied than when
299 durational cues were varied. To have a better understanding of this interaction, two mixed-effect
300 models were run, one with pitch cues and one with durational cues. These two models included
301 steps as fixed factors, random intercepts by participant. The random slope for the within-participant
302 factor steps was not included because it does not allow the model to converge (Barr et al., 2013).
303 Both models revealed a significant main effect of steps (f0 cue: $X^2 = 20.47$, $p < .01$; duration cue: X^2
304 $= 24.66$, $p < .0001$). Pairwise comparisons were performed and summarized in Table II. As shown in

305 Table II, three comparisons among seven were significant when pitch cues were varied while no
 306 comparison was significant when durational cues were varied.



307
 308 Figure 3. Mean percentage of *déplacé* responses as a function of type of cues (pitch vs. durational
 309 cues) and steps. Error bars represent standard errors.

310
 311 TABLE II. Pairwise comparisons for both types of cues

	Pitch cues		Durational cues	
	z-value	p-value	z-value	p-value
Step 1 vs. 2	1.859	= .06	0.568	> .20
Step 2 vs. 3	0.943	> .20	0.243	> .20
Step 3 vs. 4	1.448	= .15	0.283	> .20
Step 4 vs. 5	2.253	< .05 *	1.271	> .20
Step 5 vs. 6	2.962	< .01 *	0.527	> .20
Step 6 vs. 7	1.553	= .12	1.626	= 0.10
Step 7 vs. 8	4.143	< .0001 *	0.191	> .20

312
 313 To sum up, Experiment 1 showed a significant interaction between the steps and the type of
 314 cues (pitch vs. durational cues) indicating that the percentage of *déplacé* responses decreased more

315 when pitch cues were varied compared to when durational cues were varied. Such results suggests
316 that French listeners use pitch cues more than durational cues to identify lexical sequences. A closer
317 look at the results reveals that the percentage of *déplacé* responses differed from the chance level
318 when pitch cues were varied, going from 65% for step 1 to 33% for step 8. By contrast, when
319 durational cues were varied, the percentage of *déplacé* responses approached the chance level for all
320 steps going from 56% for step 1 to 46% for step 8 thus suggesting a limited role of durational cues
321 to identify lexical sequences.

322

323 Given the specificities of accentuation in the French language in which the final f0 rise is always
324 accompanied by syllable lengthening, one possibility is that French listeners exploit durational cues
325 only when pitch cues are already present in the stimuli. To test this possibility, we conducted a
326 second experiment in which pitch and durational cues were manipulated conjointly. If our reasoning
327 is correct, we expect the percentage of *déplacé* responses to decrease more when the two cues are
328 varied compared to when pitch cues only are varied. If it is not the case and if French listeners do
329 not exploit durational cues even when pitch cues are present in the stimuli, the percentage of *déplacé*
330 responses should be identical when the two cues are varied and when durational cues only are
331 varied. Finally, the comparison between the percentage of *déplacé* responses when the two cues are
332 varied and the percentage of *déplacé* responses when durational cues only are varied constitutes an
333 additional test of the role of pitch cues in accentuation marking. If, as suggested by the results of
334 Experiment 1, pitch cues are a good marker of accentuation, the percentage of *déplacé* responses
335 should be identical when the two cues are varied and when durational cues only are varied.

336

337 **III. EXPERIMENT 2**

338 Experiment 2 consisted of a two forced-choice identification task in which pitch and durational
339 cues were manipulated conjointly on the same syllable. In order to determine whether French
340 listeners benefit from the co-presence of the two cues, we compared participants' responses in
341 Experiment 2 (when the two indices were combined) to participants' responses in Experiment 1
342 (when only one cue was varied)⁵.

343

344 **A. Method**

345

346 **1. Participants**

347

348 Thirty monolingual native speakers of French (28 females, 2 males) between 18 and 38 years old
349 (mean=22.7, SD=3.2), with the same characteristics as participants from Experiment 1, participated
350 in the experiment.

351

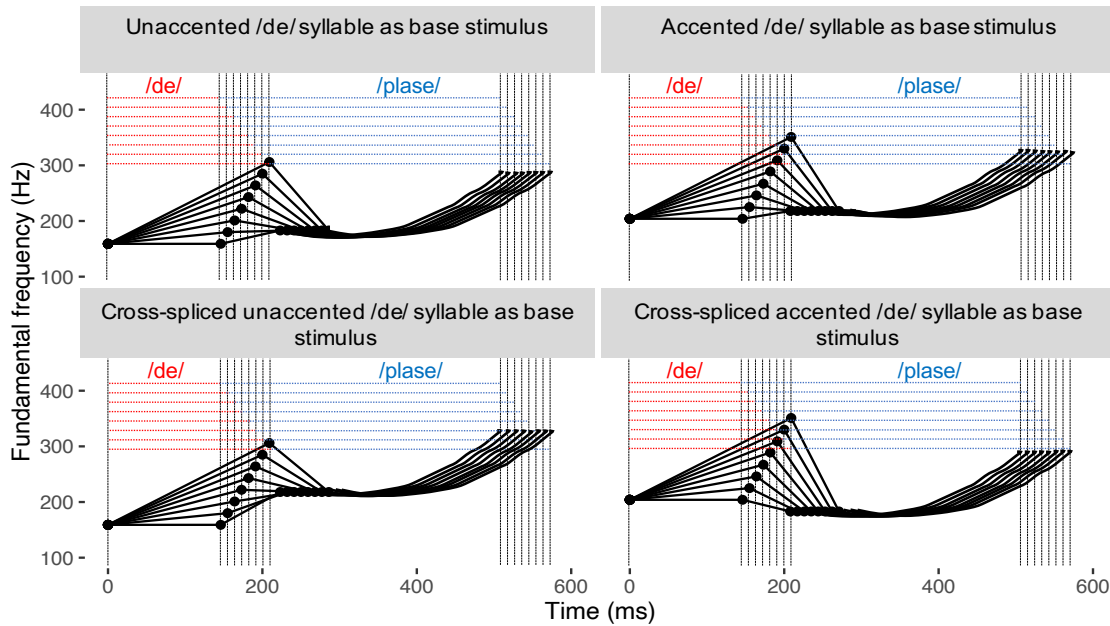
352 **2. Speech materials**

353

354 Similar to what was done for Experiment 1, four continua of 8 steps were generated from
355 natural versions of the phonemic sequence /deplase/. The natural *déplacé* sequence was used for the
356 first continuum while the natural *dé#placé* sequence was used to create the second continuum. As for
357 experiment 1, two other continua were created by cross-splicing to avoid the potential influence of
358 the following /plase/ sequence on the target syllable. We used the same procedure as in
359 Experiment 1 to do the pitch and durational manipulations except that the we manipulated both
360 indices at the same time on the same syllable. Figure 4 and table III illustrate the acoustic

361 manipulation for steps 1-8 for each base stimulus. RMS amplitude of all stimuli was equated to 50
 362 dB SPL.

363



364

365 Figure 4. Acoustic continua from the four base stimuli.

366

367 TABLE III. Acoustic characteristics of the /de/ for steps 1-8.

	Base stimuli in which the /de/ syllable was originally unaccented								Base stimuli in which the /de/ syllable was originally accented							
Steps	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
f0 min (Hz)	159	159	159	159	159	159	159	159	204	204	204	204	204	204	204	204
f0 max (Hz)	159	180	201	222	243	264	285	306	204	225	246	267	289	309	330	351
Duration (ms)	146	155	164	173	182	191	200	209	146	155	164	173	182	191	200	209

368

369 **3. Procedure**

370

371 It was the same procedure as in Experiment 1.

372 **B. Results and discussion**

373

374 Figure 5 shows the participants' proportion of *déplacé* responses as a function of steps when
375 accentuation was cued by a combination of cues compared to when it was cued by only one cue
376 (pitch cues in a. and durational cues in b.). Two mixed-effect regression models were used. The first
377 model compared the responses obtained when durational and pitch cues were varied to the
378 responses obtained when only pitch cues were varied. The second model compared the responses
379 obtained when durational and pitch indices were varied to the responses obtained when only
380 durational cues were varied. Both models included steps (1-8), number of cues (combined cues vs.
381 independent cues) and their interaction as fixed effects. Participants were included as random
382 intercept and a random slope was included for the within-participant factor steps.

383

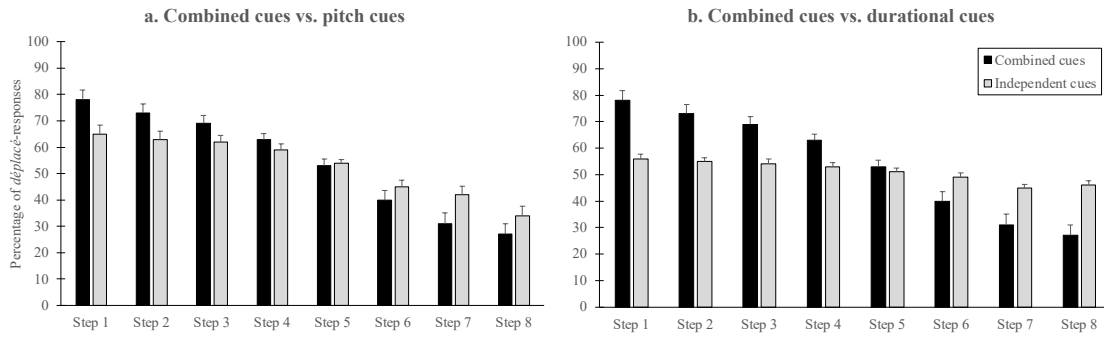
384 For the model comparing pitch cues with the combination of cues, a significant effect of steps
385 was found ($X^2 = 52.45$, $p < .0001$), showing that the percentage of *déplacé* responses significantly
386 decreased with steps. No significant effect of the number of cues was found ($X^2 = 0.25$, $p > .20$).
387 The steps x number of cues interaction was not significant ($X^2 = 10.87$, $p = 0.1442$).

388

389 For the model comparing durational cues with the combination of cues, a significant effect of
390 steps was also found, showing that the percentage of *déplacé* responses significantly decreased with
391 steps ($X^2 = 41.72$, $p < .0001$). More interestingly and contrary to what happens for pitch cues, the
392 effect of the number of cues ($X^2 = 4.37$, $p < .05$) and the steps x number of cues interaction were
393 significant ($X^2 = 25.56$, $p < .001$). As shown in Table IV, this interaction was due to the fact that
394 almost all steps significantly differ from each other for combined cues but not for durational cues.

395

396



397

398 Figure 5. Mean percentage of *déplacé* responses as a function of steps and number of cues

399 (combined vs. independent cues). Error bars represent standard errors.

400

401 TABLE IV. Pairwise comparisons between adjacent steps for both number of cues

	Combined cues		Durational cues	
	z-value	p-value	z-value	p-value
Step 1 vs. 2	4.303	< .0001	0.568	> .20
Step 2 vs. 3	2.155	< .05	0.243	> .20
Step 3 vs. 4	2.937	< .01	0.283	> .20
Step 4 vs. 5	3.617	< .001	1.271	> .20
Step 5 vs. 6	4.145	< .0001	0.527	> .20
Step 6 vs. 7	4.338	< .0001	1.626	= .10
Step 7 vs. 8	4.223	< .0001	0.191	> .20

402

403 To sum up, Experiment 2 reveals that the percentage of *déplacé* responses similarly decreased

404 when the two cues were varied compared to when pitch cues only were varied. Thus, in contrast

405 with our hypothesis, French listeners did not exploit durational cues to disambiguate lexical

406 sequences even when pitch cues were present. By contrast, the percentage of *déplacé* responses

407 decreased more when the two cues were varied compared to when durational cues only were varied.

408 This result indicates that French listeners benefit from pitch cues even when durational cues are

409 present. Taken together, these results corroborate the results of Experiment 1 by showing that pitch

410 cues are more efficient than duration cues in identification tasks. Pairwise comparisons of
411 Experiment 1 also showed that participants were able to use almost all f0 fine-grained differences
412 distinguishing two adjacent steps while this was not the case for fine-grained durational differences.
413 We thus conducted a third experiment to strengthen our claim that French listeners are more
414 sensitive to fine-grained pitch differences compared to fine-grained durational differences. If this is
415 confirmed, participants should be better to discriminate two adjacent steps differing in pitch
416 compared to two adjacent steps differing in duration.

417

418 **IV. EXPERIMENT 3**

419 Experiment 3 used the /de/ syllables extracted from the f0-continua and duration-continua of
420 Experiment 1. Participants performed an AX discrimination task in which A and X were identical or
421 differed either in pitch or in duration.

422

423 **A. Method**

424

425 **1. Participants**

426

427 Sixty monolingual native speakers of French (47 females, 13 males) between 19 and 38 years old
428 (mean=23.4, SD=3.6) with the same characteristics as participants from Experiments 1 and 2,
429 participated in the experiment. Each participant gave informed consent prior to experiment and was
430 remunerated for his/her time. As in Experiment 1, half the participants heard the stimuli from the
431 f0 continuum, while the other heard the stimuli from the duration continuum.

432

433

434 **2. *Speech materials***

435

436 For the purpose of the discrimination task, the /de/ syllables were extracted from the two f0-
437 continua and the two duration-continua without cross-splicing. Hence, this experiment concerned
438 only two continua for each type of cues among the four used in Experiment 1.

439

440 **3. *Procedure***

441

442 Participants were tested in a sound attenuated booth of the *Laboratoire Parole et Langage*, and the
443 stimuli were presented over headphones at the sound level of 60 dBA for all participants. Each
444 experimental trial consisted of a pair of two stimuli (A and X) separated by an interval of 500 ms⁶.
445 The participants were told that they would hear pairs of stimuli in which the second stimulus would
446 be either identical or different from the first stimulus. They were required to press a button on their
447 left or on their right to indicate whether X was identical or different from A. No further instructions
448 were given to participants, neither on the way the stimuli might differ, nor on the sentences from
449 which they were extracted. Half the participants heard the stimuli from the two 8-step f0-continua
450 and the other half the stimuli from the two 8-step duration-continua. For each type of continuum
451 (f0-continuum or duration-continuum) and thus for each participant, there were 7 different pairs (1-
452 2, 2-3, 3-4, 4-5, 5-6, 6-7, 7-8) and 8 same pairs. Moreover, for each type of continuum, the 7
453 different combinations were presented 10 times. In order to have the same number of same and
454 different responses, 6 pairs of the same combination were presented 9 times and 5 pairs of the same
455 combination were presented 8 times. As a result, each participant heard 280 trials randomly
456 presented (2 base stimulus * 140 combinations). The experiment lasted approximately 20 minutes.
457 An inter-trial interval of 1000ms elapsed between each participant's response and the beginning of

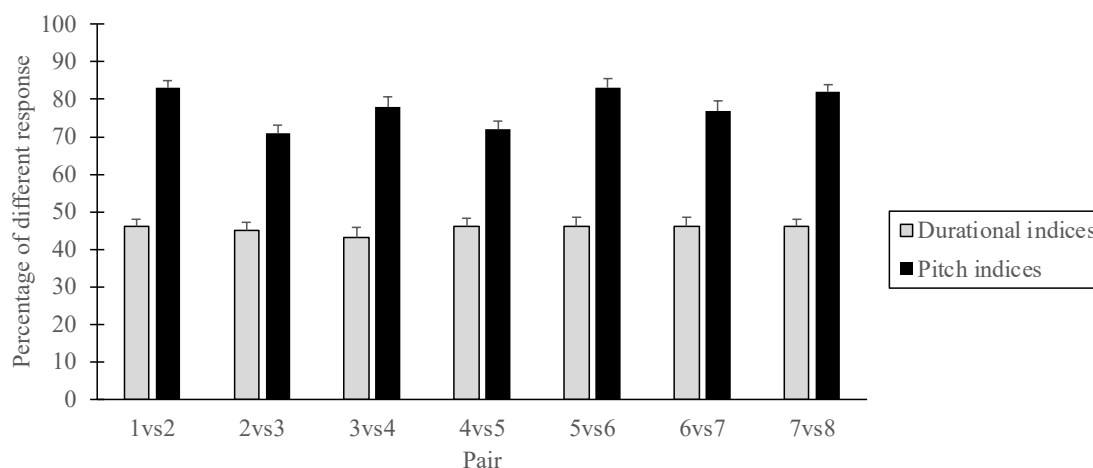
458 the next stimulus. The participants began the experiment with a block of 8 practice trials and no
459 feedback was given to participants.

460

461 **B. Results and discussion**

462

463 The mean percentage of different responses for pitch and durational cues are displayed in Figure
464 6. We analyzed accuracy data (1= different responses, 0 = identical responses) on different pairs
465 using the same procedure as in Experiment 1. A mixed-effect regression model with a logistic
466 linking function was used. This model included pairs, the type of cues (durational vs. pitch) and their
467 interaction as fixed effects. Participants were included as random intercept and a random slope was
468 included for the within-participant factor pairs. The model revealed a significant effect of type of
469 cues ($X^2 = 970.21$, $p < .0001$) showing that participants gave more different responses when the
470 accented syllable was marked by pitch cues than when it was marked by durational cues⁸. The effect
471 of the pair ($X^2 = 22.14$, $p < .01$) was also significant with more different responses for the pairs 1-2,
472 4-5, 5-6 and 7-8 compared to the pair 2-3. The pairs x type of cues interaction was also significant
473 ($X^2 = 202.04$, $p < .01$). As shown in Table V, participants gave more different responses when
474 accentuation was cued by pitch cues compared to durational cues and this is particularly true for the
475 pairs 1-2, 3-4, 4-5, 5-6, 7-8.



476

477 Figure 6. Discrimination performance of participants as a function of pairs and type of cues.

478 Errors bars represents standard errors.

479

480 TABLE V. Pairwise comparison between pitch and durational cues for each pair

	z-value	p-value
Pitch pair 1-2 vs. Durational pair 1-2	13.200	> .0001
Pitch pair 2-3 vs. Durational pair 2-3	9.291	> .0001
Pitch pair 3-4 vs. Durational pair 3-4	12.327	> .0001
Pitch pair 4-5 vs. Durational pair 4-5	12.571	> .0001
Pitch pair 5-6 vs. Durational pair 5-6	12.898	> .0001
Pitch pair 6-7 vs. Durational pair 6-7	10.967	> .0001
Pitch pair 7-8 vs. Durational pair 7-8	12.508	> .0001

481

482 To sum up, Experiment 3 showed a significant effect of the type of cues (durational vs. pitch
 483 cues) showing that participants gave more different responses when pitch cues were varied rather
 484 than when duration cues were varied. This result reveals that French listeners were more sensitive to
 485 fine-grained pitch cues than to fine-grained durational cues to discriminate syllables differing in
 486 accentuation. Similarly to what we found in Experiment 1, participants' performance were above
 487 chance level when accentuation was cued by fine-grained pitch cues while participants' performance

488 were at the chance level when accentuation was cued by fine-grained durational cues. Again, due to
489 the specificities of accentuation in French in which the final f0 rise is always accompanied by syllable
490 lengthening, it is possible that French listeners are able to use fine-grained durational differences to
491 discriminate syllables only when these differences are accompanied by pitch differences. To test this
492 possibility, we conducted an additional discrimination experiment in which fine-grained pitch and
493 durational cues were manipulated conjointly. If our reasoning is correct, participants' performance
494 should be better when the two cues were varied compared to when fine-grained pitch cues were
495 varied. Moreover, as for Experiment 2, the comparison between participants' performance when the
496 two cues are varied and participants' performance when fine-grained durational cues are varied
497 constitutes an additional test of French listeners' sensitivity to fine-grained pitch differences.

498

499 **V. EXPERIMENT 4**

500

501 Experiment 4 used the /de/ syllables extracted from the continua of Experiment 2. Participants
502 performed an AX discrimination task in which A and X were identical or differed on the two cues
503 (pitch + durational cues). In order to determine whether French listeners benefit from the co-
504 presence of both cues, we compared the performance of participants in Experiment 4 (when the
505 two cues were varied) to their performance in Experiment 3 (when only cue was varied).

506

507 **A. Method**

508

509 **1. Participants**

510

511 Thirty monolingual native speakers of French (25 females, 5 males) between 18 and 32 years old
512 (mean=23.1, SD=3.1) with the same characteristics as participants from Experiment 1 to 3
513 participated in the experiment.

514

515 2. *Speech materials*

516

517 For the purpose of the discrimination task, the /de/ syllables were extracted from the two
518 continua of Experiment 2 without cross-splicing.

519

520 3. *Procedure*

521

522 It was the same procedure as in Experiment 3.

523

524 B. **Results and discussion**

525

526 The mean percentage of different responses for pitch and durational cues are displayed in Figure
527 7. We analyzed accuracy data (1 = different responses, 0 = identical responses) on different pairs
528 using the same procedure as in Experiment 2. Two mixed-effect regression models were used. The
529 first model compared the responses obtained when durational and pitch cues were varied to the
530 responses obtained when only pitch cues were varied. The second model compared the responses
531 obtained when durational and pitch cues were varied to the responses obtained when only durational
532 cues were varied. Both models included pairs, the number of cues (combined vs. independent) and
533 their interaction as fixed effects. Participants were included as random intercept and a random slope
534 was included for the within-participant factor pairs.

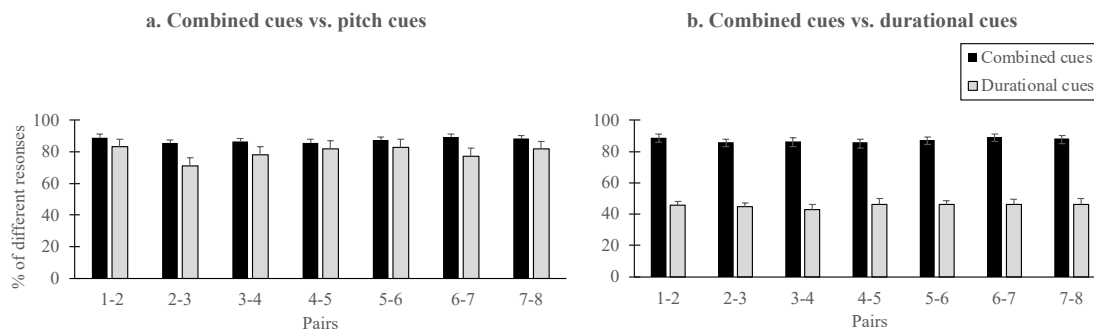
535

536 For the model including pitch cues, a significant effect of the number of cues was found ($X^2 =$
537 30.93, $p < .0001$), showing that participants gave more different responses when the accented
538 syllable was marked by combined cues compared to when it was marked by pitch cues only. The
539 effect of pairs was also significant ($X^2 = 29.56$, $p < .0001$) with more different responses for the
540 pairs 1-2 and 6-7 compared to the pair 2-3. The pairs x number of cues interaction was also
541 significant ($X^2 = 12.76$, $p < .05$). As shown in Table VI, participants gave more different responses
542 when accentuation was cued by combined cues than by pitch cues for pairs 1-2,2-3,3-4,6-7,7-8 but
543 not for pairs 4-5 and 5-6.

544

545 For the model including durational cues, a significant effect of the number of cues was also
546 found ($X^2 = 252.23$ $p < .0001$), showing that participants gave more different responses when the
547 accented syllable was marked by combined cues compared to when it was marked by durational cues
548 only.⁹ The effect of the pairs was not significant ($X^2 = 5.56$, $p > .20$). Crucially and contrary to what
549 happens for pitch cues the pairs x number of cues interaction was not significant ($X^2 = 3.48$, $p >$
550 $.20$), showing better performance when cues are combined for all pairs.

551



552

553 Figure 7. Discrimination performance of participants as a function of pairs and number of cues
554 (combined vs. independent cues). Error bars represent standard errors.

555 TABLE VI. Pairwise comparison between combined and independent cues for each pair

	z-value	p-value
Combined vs. independent cues pair 1-2	5.775	< .0001
Combined vs. independent cues pair 2-3	5.142	< .0001
Combined vs. independent cues pair 3-4	2.581	< .01
Combined vs. independent cues pair 4-5	1.292	> .20
Combined vs. independent cues pair 5-6	1.371	= .17
Combined vs. independent cues pair 6-7	4.353	< .0001
Combined vs. independent cues pair 7-8	2.523	< .05

556

557 To sum up, results of Experiment 4 showed a significant effect of the number of cues
 558 (combined vs. independent cues) for both pitch cues and durational cues, showing that in both cases
 559 participants benefit from cue redundancy to discriminate syllables differing in accentuation.

560

561 **V. GENERAL DISCUSSION**

562 The aim of this study was to examine the relative contribution of pitch and durational cues to
 563 accent marking in French. To do so, French participants were tested on acoustic continua ranging
 564 from an unaccented first syllable [depla'se] ‘moved’ to an accented first syllable ['de] [pla'se] ‘a dice
 565 placed’ while durational and pitch cues were manipulated either separately or conjointly. Given the
 566 previous contradictory results about the role of durational cues on speech segmentation in French,
 567 we predicted that pitch cues would be a better marker to accentuation than durational cues in that
 568 language. Moreover, given the specificities of the French language in which the final f0 rise is always
 569 accompanied by syllable lengthening, we expected durational cues to be exploited only when pitch
 570 cues were already present in the stimuli.

571

572 Consistent with our first prediction, we found better performance in both the identification and
 573 discrimination tasks when pitch cues were varied compared to when durational cues were varied.

574 When the two lexical sequences differed on the sole presence of an f_0 rise, participants' performance
575 to identify lexical sequences differed from the chance level for all steps of the acoustic continua.
576 Moreover, participants were also able to discriminate all adjacent steps that differed in pitch. Such
577 results corroborate previous studies (Tremblay et al., 2012; Bagou & Frauenfelder, 2018) by showing
578 that French listeners are sensitive to very fine-grained f_0 differences and are able to exploit them to
579 segment speech into words even when durational cues were not found in the stimuli.

580

581 The situation is more complicated for durational cues. When durational cues were located at the
582 extremities of the acoustic continua, French participants were able to exploit the sole presence of
583 these cues to identify lexical sequences. By contrast, when locating at the center of the continua,
584 participants' performance approach the chance level and participants were not able to exploit these
585 cues to decide whether the lexical sequences they heard constituted one single word (*déplacé*) or two
586 separated words (*dé#placé*). Moreover, although the duration difference between two adjacent steps
587 of the continuum is really close to the just-noticeable difference (JND), participants were not able to
588 use this duration difference to discriminate adjacent steps. This suggests that French listeners'
589 sensitivity to durational cues is more limited than their sensitivity to pitch cues.

590

591 Taken together, these results thus indicate a more limited role of durational cues compared to
592 pitch cues to accent marking in French. This limited role of durational cues is corroborated by the
593 results of our identification task (Experience 2) showing that French participants did not benefit
594 from durational cues when pitch cues were already present in the stimuli. In fact, contrary to our
595 second prediction, we did not observe any benefit of the use of durational cues in lexical sequences
596 identification when the stimuli already differed in pitch. Given that the difference between the
597 unaccented syllable and the accented syllables in the original stimuli from which the speech was

598 manipulated was relatively small (of 63 ms), one alternative hypothesis could be that any listeners
599 (and not just French listeners) would perceive the pitch difference as being far more salient than the
600 durational difference. However, in a study by Tremblay et al. (2012) the reverse pattern was
601 observed and English listeners used durational cues (of about 40ms) but not pitch cues to segment
602 speech into words. Hence our results corroborate a previous studies suggesting that the co-
603 occurrence of pitch and durational cues is not necessary for French speakers and that the sole
604 presence of pitch cues is sufficient for them to segment the speech stream into words (Tremblay et
605 al., 2012).

606

607 The greater sensibility of French listeners to pitch cues compared to durational cues does not
608 mean however that they do not exploit durational cues at all. The fact that in the absence of pitch
609 cues, our participants were able to use durational differences located at the extremities of our
610 acoustic continua confirms the idea that French listeners are able to use the sole presence of these
611 cues provided that they are sufficiently brief or long. In our case, participants exploited them to
612 identify lexical sequences when the difference between the unaccented syllable and the accented one
613 was about 40 %. This is in line with a previous study showing that French speakers were more likely
614 to interpret phonemically ambiguous sequences of two syllables as containing one word (e.g., *bagage*
615 ‘luggage’) than two words (e.g. *bas gage* ‘low pledge’) when the first syllable was 35% longer than the
616 second syllable (Banel & Bacri, 1994). The claim that French speakers are to some extent able to
617 exploit the presence of durational cues is also corroborated by the results of the discrimination task
618 (Experience 4) in which participants were found, for some adjacent pairs, to be able to discriminate
619 two syllables that already differed in pitch.

620

621 It thus remains to explain why our participants benefit to some extent from the additional
622 presence of durational cues to discriminate syllables in the discrimination task while they do not
623 benefit from this information to identify lexical sequences in the identification task. A possibility is
624 that, in the discrimination task, participants give their response while durational cues are yet retained
625 in memory since in this kind of task only the syllable of interest is presented to the participants. In
626 contrast, in the identification task, the entire /deplase/ sequences are presented to participants.
627 Hence as discussed just above, it remains possible that durational cues are not sufficiently salient
628 either in terms of shortness or length, and one consequence is that the durational cues present in the
629 first syllable of the /deplase/ sequence are overwhelmed by the phonetic/phonemic information
630 extracted during the processing of the second syllable. As a result, information relative to the first
631 syllable (i.e. /de/ in *déplacé* or *dé#placé*) can no longer be used to perform the identification task.

632

633 To conclude, our study show that French listeners rely more on pitch cues than on durational
634 cues to differentiate an accented syllable from an unaccented one. Even if the vast majority of
635 production studies have highlighted two main acoustic correlates of accentuation in French, our
636 results therefore suggest that pitch cues are the most robust accentual cues in perception, at least in
637 identification and discrimination tasks.

638

639

640 **ACKNOWLEDGMENTS**

641 This work was supported by the *Agence Nationale de la Recherche* (ANR) under Grant ANR-23-
642 CE28-0021-01 to Amandine Michelas and the French government under the France 2030
643 investment plan, as part of the *Initiative d'Excellence d'Aix-Marseille Université* (A*MIDEX) under Grant
644 AMX-22-RE-AB-178.

645

646

647 **FOOTNOTES**

648 ¹ While pitch is sometimes described as a correlate of lexical stress in English (Fry, 1958), there is an
649 increasing consensus that there is no one-to-one relationship between pitch and lexical stress (e.g.,
650 Beckman & Edward, 1994; Sluijter & van Heuven, 1996). Instead, stressed syllables are “docking
651 sites” for intonational patterns that affects the sentences level. Stressed syllables can thus have
652 different intonational pitch accents (e.g., H*, L + H*, L*, L*+H) and thus different fundamental
653 frequency (F0) patterns depending on discourse information (Beckman & Pierrehumbert, 1986).
654 Hence, pitch can only be interpreted as a cue to lexical stress if sentence intonation (and specifically
655 the pitch accent associated to the stressed syllable) is taken into consideration (Brown et al., 2015).

656 ² Because in French the acoustic prominence given to a syllable is determined post-lexically and not
657 lexically, the term “accent” rather than “stress” is usually used (see for instance Delais-Roussarie et
658 al., 2015).

659 ³ For the four experiments, each participant gave informed consent prior to experiment and was
660 remunerated for his/her time. All the studies were conducted according to the ethics procedure of
661 Aix Marseille University and the principles expressed in the Declaration of Helsinki.

662 ⁴ Note that vowel quality differed in the original stimuli across the two sequences with a clearer,
663 better articulated vowel in the *dé#placé* sequence than in the *déplacé* sequence. Since our continua
664 were created from the /de/ syllables extracted from the two *déplacé* and *dé#placé* sequences, it is
665 unlikely that these differences of vowel quality will impact our results.

666

667 ⁵ Since our aim is to determine whether French listeners benefit from the co-presence of the two
668 cues, the comparison between results of Experiment 2 (when the two indices were combined) and
669 the results of Experiment 1 (when only one cue was varied) is crucial whereas the results of
670 Experiment 2 in themselves are not essential. Nevertheless, a mixed-effect regression model run on
671 participants' responses in experiment 2 (without comparison with Experiment 1) showed a
672 significant effect of steps ($X^2 = 32.65$, $p < .0001$) showing that the percentage of *déplacé* responses
673 significantly decrease with steps.

674 ⁶ In our study, the four base stimuli were used in the same way to generate our continua in both the
675 pitch cues and the durational cues conditions. It is thus unlikely that base stimulus will impact our
676 results. For this reason, base stimulus was not included as random factor in our statistical model.

677 ⁷ We used a 500ms ISI that leaves the time for the stimulus A to be fully processed so that both
678 phonetic and phonological information are available when the stimulus X is presented (see Tyler and
679 Fenwick, 2012).

680 ⁸ d' values from the accuracy data were also calculated using the function $(hit, fa) \{qnorm(hit) -$
681 $qnorm(fa)\}$ of the statistical software R (R core team, 2022; see Pallier, 2002). Responses were coded
682 as hits when participants correctly responded different, and as false alarms when participants
683 incorrectly responded different on the “same” trials (i.e., X% of false alarms). A linear mixed-effect
684 regression model was run on d' data. This model included the type of cues (durational vs. pitch) as
685 fixed effect and participants as random intercepts. The model revealed a significant effect of the type
686 of cues ($X^2 = 179.66$, $p < .0001$) showing greater sensitivity when accentuation was cued by pitch cues
687 compared to durational cues. Note that because the hits of the d' are calculated on the different

688 pairs and the false alarms on the identical pairs, it is not possible to include the factor pair within this
689 analysis.

690 ⁹As in Experiment 3, d' values from the accuracy data were calculated. For both models, a
691 significant effect of the number of cues was found, showing greater sensitivity for the combined
692 cues than for one cue, whether it is for durational cues ($X^2= 128.78, p<.0001$) or pitch cues ($X^2=$
693 $5.9916, p<.05$).

694 **AUTHOR DECLARATIONS**

695 No conflicts to disclose.

696

697 **DATA AVAILABILITY**

698 All relevant data are available from OSF: <https://osf.io/7t6xz/>

699 **REFERENCES**

- 700 Abramson, A. & Lisker, L. (1985). "Relative power of cues: F0 shift versus voice timing", in *Phonetic*
701 *Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, New York), pp.
702 25-33.
- 703 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). "Mixed-effects modeling with crossed
704 random effects for subjects and items". *Journal of Memory and language*, **59(4)**, 390-412.
- 705 Bagou, O., Fougeron, C., & Frauenfelder, U. (2002). "Contribution of prosody to the segmentation
706 and storage of" words" in the acquisition of a new mini-language". *Proceedings of the speech*
707 *prosody 2002 conference* , 59-62.
- 708 Bagou, O., & Frauenfelder, U. H. (2006). "Stratégie de segmentation prosodique: rôle des
709 proéminences initiales et finales dans l'acquisition d'une langue artificielle". *Proceedings of the*
710 *XXVIèmes Journées d'Etude sur la Parole*, 571-574.

- 711 Bagou, O., & Frauenfelder, U. H. (2018). "Lexical segmentation in artificial word learning: The
712 effects of converging sublexical cues". *Language and speech*, **61(1)**, 3-30.
- 713 Banel, M. H., & Bacri, N. (1994). "On metrical patterns and lexical parsing in French". *Speech*
714 *Communication*, **15(1-2)**, 115-126.
- 715 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). "Random effects structure for confirmatory
716 hypothesis testing: Keep it maximal". *Journal of Memory and Language*, **68(3)**, 255-278.
- 717 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). "Fitting Linear Mixed-Effects Models
718 Using lme4". *Journal of Statistical Software*, **67(1)**, 1-48.
- 719 Beckman, M. E., & Pierrehumbert, J. (1986). "Intonational structure in English and Japanese,"
720 *Phonology*, **3**, 255–310.
- 721 Beckman, M. E., & Edwards, J. (1994). "Articulatory evidence for differentiating stress categories,"
722 in *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, edited by P. A.
723 Keating (Cambridge University, Cambridge, UK), pp. 7–33.
- 724 Boersma, Paul & Weenink, David (2024). "Praat: doing phonetics by computer" [Computer
725 program]. Version 6.4.20, retrieved 18 September 2024 from <http://www.praat.org/>
- 726 Brown, M. E., Salverda, A. P., Dilley, L. C., and Tanenhaus, M. K. (2015). "Metrical expectations
727 from preceding prosody influence perception of lexical stress," *J. Exp. Psychol. Hum. Percept.*
728 *Perform.* **41**, 306–323.
- 729 Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). "Phonological phrase
730 boundaries constrain lexical access I. Adult data". *Journal of Memory and Language*, **51(4)**, 523-
731 547.
- 732 Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, C., Feldhausen, I., ... Yoo, H.
733 (2015). "Developing a ToBI system for French" in *Intonational variation in Romance*, edited by S.
734 Frota & P. Prieto (Oxford: OUP).

735 Fry, D. B. (1958). "Experiments in the perception of stress". *Language and speech*, **1(2)**, 126-152.

736 Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models*, Third Edition (Sage).

737 Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. (2008). "Cue-specific effects of
738 categorization training on the relative weighting of acoustic cues to consonant voicing in
739 English," *The Journal of the Acoustical Society of America*. **124(2)**, 1234– 1251.

740 Gay, T. (1978). "Physiological and acoustic correlates of perceived stress," *Language & Speech*, **21**,
741 347–353.

742 Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable
743 across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of*
744 *America*, **132(6)**, 3950-3964.

745 Jun, S.A., Fougeron, C. (2002) "Realizations of accentual phrase in French". *Probus*; **14**,147-172.

746 Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual
747 evidence". *The Journal of the Acoustical Society of America*, **59(5)**, 1208-1221.

748 Koffi, E. (2018). "A Just Noticeable Difference (JND) reanalysis of Fry's original acoustic correlates
749 of stress in American English". *Linguistic Portfolios*, **7(1)**, 2.

750 Lindblom, B. (1963). "Spectrographic study of vowel reduction, *The Journal of the Acoustical Society of*
751 *America*, **35**, 1173–1181.

752 Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/versus/p/in
753 trochees. *Language and speech*, **29(1)**, 3-11.

754 Michelas, A., & d'Imperio, M. (2012). "When syntax meets prosody: Tonal and duration variability
755 in French Accentual Phrases". *Journal of Phonetics*, **40(6)**, 816-829.

756 Michelas, A., & D'Imperio, M. (2015). "Prosodic boundary strength guides syntactic parsing of
757 French utterances". *Laboratory Phonology*, **6(1)**, 119-146.

758 Moulines, E., & Charpentier, F. (1990). "Pitch-synchronous waveform processing techniques for
759 text-to-speech synthesis using diphones". *Speech communication*, **9(5-6)**, 453-467.

760 Pallier, C. (2002). "Computing discriminability and bias with the R software". URL
761 <http://www.pallier.org/ressources/aprime/aprime>.

762 Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., & Höhle, B.
763 (2017). "Prosodic boundary cues in German: Evidence from the production and perception of
764 bracketed lists". *Journal of phonetics*, **61**, 71-92.

765 R Core Team (2022). R: A language and environment for statistical computing. R Foundation for
766 Statistical Computing", Vienna, Austria. URL <https://www.R-project.org/>.

767 Rietveld, A. C. M. (1980). "Word boundaries in the French language". *Language and Speech*, **23(3)**,
768 289-296.

769 Schertz, J., & Clare, E. J. (2020). "Phonetic cue weighting in perception and production." *Cognitive*
770 *Science*, **11(2)**, e1521.

771 Sluijter, A. M. C., & van Heuven, V. J. (1996). "Acoustic correlates of linguistic stress and accent in
772 Dutch and American English". In *Proceedings of Fourth International Congress on Spoken Language*
773 *Processing*, October 3–6, Newark, DE.

774 Smith, C. L., Erickson, D., & Savariaux, C. (2019). "Articulatory and acoustic correlates of
775 prominence in French: Comparing L1 and L2 speakers". *Journal of Phonetics*, **77**, 100938.

776 Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). "Differential contribution of
777 prosodic cues in the native and non-native segmentation of French speech". *Laboratory*
778 *Phonology*, **3(2)**, 385-423.

779 Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). "Effects of the native language on
780 the learning of fundamental frequency in second-language speech segmentation". *Frontiers in*
781 *psychology*, **7**, 985.

782 Tremblay, A., Namjoshi, J., Spinelli, E., Broersma, M., Cho, T., Kim, S., Martinez-Garcia, M. T., &
783 Connell, K. (2017). “Experience with a second language affects the use of fundamental
784 frequency in speech segmentation”. *PloS one*, **12(7)**, e0181709.

785 Tyler, M. D., & Cutler, A. (2009). “Cross-language differences in cue use for speech segmentation”.
786 *The Journal of the Acoustical Society of America*, **126(1)**, 367-376.

787 Tyler, M. & Fenwick, S. (2012). “Perceptual Assimilation of Arabic Voiceless Fricatives by English
788 Monolinguals”. In *Proceedings of Interspeech*, September 9-13, Portland, Orega, USA, 911-914.

789 van Heuven, V. J., and de Jonge, M. (2011). “Spectral and temporal reduction as stress cues in
790 Dutch,” *Phonetica* **68(3)**, 120–132.

791 Welby, P. (2006). “French intonational structure: Evidence from tonal alignment”. *Journal of Phonetics*,
792 **34(3)**, 343-371.

793 Zhang, Y., & Francis, A. (2010). “The weighting of vowel quality in native and non-native listeners’
794 perception of English lexical stress,” *Journal of Phonetics*. **38**, 260–271.

795