

Spoken word boundary detection in ambiguous resyllabification contexts in Spanish

José María Lahoz-Bengoechea¹, Miguel Jiménez-Bravo²

¹ Department of Spanish Linguistics and Literary Theory, Complutense University of Madrid, Spain
jmlahoz@ucm.es

² Department of General Linguistics, Complutense University of Madrid, Spain
miguel.jimenez.bravo@ucm.es

Corresponding author: José María Lahoz-Bengoechea

Abstract

In Spanish, word-final consonants are resyllabified with the initial vowel of a following word, e.g. *mis amigos* ('my friends') is traditionally syllabified as *mi.sa.mi.gos*. Following previous acoustic descriptions, this study examines consonant duration as a perceptual cue to disambiguate the original lexical affiliation of such segments. Twelve minimal pairs of sentences were recorded by a native speaker differing only in the lexical affiliation of the consonant—four pairs for /s/, /n/, and /l/, respectively. Consonant duration was manipulated by computer in five equidistant steps, and the resulting stimuli were assessed by sixty-five native participants in a lexical decision experiment. Shorter consonant duration was found to bias participants' perception towards a word-final interpretation in the case of /s/ and /n/ but not for /l/. Drawing on a previous production study, we confirm that consonant duration plays a major role in the recognition of resyllabification contexts. Additionally, we argue that some segment-specific cues—possibly also affecting contextual segments—are originated by the gestural configuration of each consonant and might be responsible for the observed perceptual differences. Finally, evidence is provided for the simultaneous coda and onset status of resyllabified consonants, and accordingly they are proposed to be phonologically represented as ambisyllabic.

Keywords

Spanish; resyllabification; spoken word recognition; duration; gestural coordination; ambisyllabicity

1. Introduction

1.1. Spoken word recognition

Spoken word recognition is a hard task, especially since words are not usually separated by pauses in connected speech, and listeners have to identify word boundaries by means of different strategies. Many of these strategies rely on the phonological properties of the target language and make language processing more efficient for native listeners, while non-native speakers are typically slower and less effective, to the point of even missing some of these word boundaries (Cutler, 2012).

¹<https://orcid.org/0000-0002-4654-6776>

²<https://orcid.org/0000-0003-3579-1618>

33 From birth, speech segmentation is aided by the computation of the transitional probabilities that exist
34 between different kinds of units, such as segments and syllables. Thus, the high predictive power of
35 one element over the following is interpreted as both being part of the same higher-level unit, namely
36 a word. Conversely, a low transitional probability signals a word boundary. According to Saffran et al.
37 (1996), this procedure is available to both infants and adults when learning a new language, and it allows
38 segmentation even if no prior knowledge of vocabulary exists, finally helping in building up a lexicon.

39 As this learning process takes place, infants are biased to paying attention to elements occurring period-
40 ically (Cutler & Mehler, 1993)—i.e the foot, the syllable, or the mora, depending on the language. As a
41 result, infants learn to give more weight to lexical candidates that are aligned with the rhythmic unit of
42 their language. This habit persists in adulthood and increases the efficiency in deciding between com-
43 peting items. For example, in a word-spotting task, where subjects had to find real words embedded in
44 nonsense words, Cutler and Norris (1988) found that words that were split between two feet took longer
45 to be identified by English listeners, given that English is typically considered a stress-timed language,
46 i.e. its rhythm is based on the foot. In their turn, Mehler et al. (1981) had French subjects listen to
47 lists of words and press a button if any of the words started by a visually presented sequence of the form
48 /CV/ or /CVC/. The reaction times were significantly lower when the target sequence matched the syl-
49 labic structure of the words, showing that syllables play a major role in segmentation in syllable-timed
50 languages, like French. With a similar design, Otake et al. (1993) tested Japanese subjects, who had to
51 identify words beginning with a certain /CV/ or /CVN/ sequence. Their results did not replicate those
52 for French, but showed a preference for mora-based segmentation, i.e. for subjects who had to recognize
53 /CV/ sequences there was no significant difference in reaction times when exposed to words beginning
54 by /CV/ or /CVN/, whereas subjects spotting /CVN/ sequences missed items like *nanoka* ('the seventh
55 day of the month') because the mora structure of the word did not match that of the target sequence.

56 Spanish is generally considered a syllable-timed language (Pike, 1945), typically possessing a preponder-
57 ance of open syllables and lacking vowel reduction in the view of most authors. As such, similar to the
58 results obtained for French by Mehler et al. (1981), it is agreed that syllable structure is largely relied
59 upon in word recognition (Sebastián-Gallés et al., 1992). However, Spanish exhibits resyllabification, a
60 process that might affect the segmentation of words, since the structure of the syllable is altered, and
61 word-final consonants are realized across a word boundary together with the following vowel. Based on
62 previous acoustic descriptions (Hualde & Prieto, 2014; Jiménez-Bravo & Lahoz-Bengoechea, submitted;
63 Strycharczuk & Kohlberger, 2016), this paper presents a perceptual study that manipulates consonant
64 duration to assess its possible role in disambiguating the original lexical affiliation of segments in such
65 resyllabified cases.

66 1.2. Resyllabification

67 Resyllabification is a well-established phenomenon in many languages and involves the post-lexical re-
68 organization of syllable structure. In Spanish, for example, word-final consonants are realized as onsets
69 together with the initial vowel of the following word, e.g. *mis amigos* ('my friends') is traditionally syl-
70 labified as *mi.sa.mi.gos*.

71 Despite the fact that resyllabification is a widespread phenomenon across Spanish dialects, geographical
72 variation has also been reported. For example, in Castilian Spanish any consonant can undergo resyllab-
73 ification, whereas in certain dialects of Ecuador some consonants, such as /l/, are subject to this process
74 but others, namely /s/ and /n/, are not (Robinson, 2012). Additionally, in both Paraguayan Spanish
75 and Yucatan Spanish the insertion of a glottal stop before a word-initial vowel impedes resyllabification
76 in certain contexts due to the influence of the local indigenous languages (Michnowicz & Kagan, 2016;
77 Trawick & Michnowicz, 2019).

78 Another type of dialectal variation is related to the application order of resyllabification with respect
79 to other allophonic rules. Obstruents in coda position typically undergo aspiration in many Spanish
80 dialects and, thus, are realized as [h]. In some of these dialects, aspiration applies before resyllabification:
81 for example, in Caribbean Spanish and Rio Negro Argentinian Spanish, an utterance such as *dos alas*
82 ('two wings') surfaces as [do.ha.lah] (Kaisse, 1999). In contrast, the variety of Buenos Aires applies
83 resyllabification first and, therefore, bleeds aspiration, so that the aforementioned utterance *dos alas* rather
84 surfaces as [do.sa.lah] (Kaisse, 1999).

85 Examples such as the latter from dialects applying bleeding order reveal that word-final /s/ undoubtedly
86 becomes syllable onset after resyllabification. This interpretation is supported by additional evidence.
87 For example, the Spanish tap /r/ presents several allophones in coda, such as [r], [r̥], [ɾ], or [ɾ̥] (Blecua,
88 2001), but when resyllabified it surfaces as [r], the typical onset allophone. In addition, pre-nuclear pitch
89 accents of Castilian Spanish statements involve a pitch rise that is aligned to the beginning of the stressed
90 syllable. According to Jiménez-Bravo and Lahoz-Bengoechea (submitted), in resyllabification contexts
91 the beginning of the pitch rise is aligned not to the vowel—the first segment of the syllable in the lexical
92 representation—, but to the beginning of the resyllabified consonant, thus revealing that the consonant
93 has truly become the syllable onset. Other languages also seem to treat resyllabified consonants as onsets.
94 For example, articulographic measurements of resyllabified /m/ in Korean have made evident that the
95 labial gesture starts simultaneously with the articulation of the upcoming vowel, thus displaying the
96 coordination pattern that is characteristic of onsets (Cho et al., 2014).

97 The view traditionally held for Castilian Spanish is that resyllabified codas—henceforth referred to as de-
98 rived onsets—present no differences with word-initial onsets—referred to as canonical onsets—, and both
99 are considered to be phonetically identical, i.e. they are homophones (e.g. Hualde, 2005). However, the
100 view in which word-final consonants undergo complete resyllabification and are indistinguishable from
101 canonical onsets has been challenged. Several studies have found subphonemic differences for resyllab-
102 ified /s/ that set derived onsets apart as a separate category different from both canonical onsets and
103 canonical codas (Hualde & Prieto, 2014; Jiménez-Bravo & Lahoz-Bengoechea, submitted; Strycharczuk &
104 Kohlberger, 2016).

105 The study conducted by Hualde and Prieto (2014) analyzed duration of alveolar fricatives—/s/ for Castil-
106 ian Spanish and contrasting /s/~/z/ for Catalan—in word-initial, word-medial, and word-final positions.
107 Their results for both languages reported shorter duration of the target consonant when it was realized
108 as a derived onset /VC#V/ in comparison to both word-initial onsets /V#CV/ and word-medial onsets
109 /VCV/.

110 These results, obtained with spontaneous speech elicited through a map task, were later replicated and
111 complemented by a study conducted by Strycharczuk and Kohlberger (2016). In their case, they used
112 speech samples of Castilian Spanish obtained, rather, in more a controlled environment to avoid the
113 undesired effects potentially associated to prosody and speech rate. Similarly, the acoustic realization
114 of /s/ was compared among different syllable structures, and the results showed significant durational
115 differences along the following structures: Vs.C < Vs#C < Vs#V < V.sV < V#sV < Vs#sV. In the
116 resulting fine-grained contrast, derived onsets /Vs#V/ proved longer than canonical codas /Vs#C/ but
117 shorter than canonical onsets /V#sV/, thus yielding a three-way distinction in the phonetic realisation
118 of /s/, with derived onsets as an intermediate category. Differently, vowels preceding derived onsets
119 presented no durational differences with those preceding canonical onsets, although both were longer
120 than vowels in closed syllables, i.e. those with word-final and word-medial codas. Thus, derived onsets
121 did not condition the shortening of the preceding vowel, but behaved as canonical onsets despite their
122 being shorter in duration.

123 Like in the aforementioned studies, Jiménez-Bravo and Lahoz-Bengoechea (submitted) also observed that
124 /s/ is shorter in resyllabification contexts. Additionally, that study attested the same durational difference
125 for /n/ and /l/ when involved in resyllabification. On the other hand, reduced duration for vowels
126 preceding resyllabified /l/ was also observed when compared with those preceding canonical onsets.
127 However, similar to Strycharczuk and Kohlberger's results, this was not found for /s/, and neither was it
128 for /n/.

129 Another study on the acquisition of Spanish resyllabification by English speakers indirectly offered dura-
130 tional values for derived and canonical onsets (Scarpace, 2017), although the reported results are difficult
131 to interpret with respect to the studies previously mentioned. On the one hand, the different speech ma-
132 terials used in the experiments were recorded by a native speaker of Colombian Spanish and by a native
133 speaker of Mexican Spanish. On the other hand, no statistics were conducted in one of the experiments—
134 yet the reported durational means seem equivocal¹ (p. 48)—, and when they were conducted, no signif-
135 icant differences were reported (p. 72).

136 In other languages, evidence in favour of the different duration of derived onsets and canonical onsets
137 in resyllabification contexts are similar to those reported for Castilian Spanish, e.g. for English (Lehiste,
138 1960; Smith & Hawkins, 2012; Tao et al., 2018; Turk & Shattuck-Hufnagel, 2000), for French (Shoemaker,
139 2009; Spinelli et al., 2003), for Dutch (Quené, 1992; Shatzman & McQueen, 2006).

140 Nonetheless, a variety of other acoustic cues may also signal misaligned syllable and word boundaries. In
141 English, for example, word-final pre-vocalic stops seem to differ in the relative duration of its closure with
142 respect to its entire duration when compared to word-initial stops (Boucher, 1988). In addition, canonical
143 onsets seem to be also characterized by aspiration (Lee et al., 2019; Nakatani & Dukes, 1977), while glottal
144 stops often precede word-initial vowels (Hoard, 1966; Nakatani & Dukes, 1977). Also for English, studies
145 have observed differences in the articulatory gestures involved in derived onsets and canonical onsets
146 (Byrd et al., 2009; Gick, 2003) as well as a preponderant occurrence of full vowels word-initially in /CV/
147 sequences (Cutler & Carter, 1987).

148 1.3. Perceptual cues to resyllabification

149 Segmental duration is known to be affected by prosody at word level, i.e. pre-boundary lengthening and
150 domain-initial (post-boundary) strengthening (Byrd et al., 2000; Fougeron & Keating, 1997; Wightman
151 et al., 1992), whose role in facilitating speech segmentation and thereby lexical access is well-attested
152 (e.g. Lehiste, 1960; Quené, 1992; Spinelli et al., 2003; Shatzman & McQueen, 2006). More uncertainty
153 exists, however, whether lexical access is in fact hindered by the misalignment typically observed in
154 resyllabification contexts between the syllable boundary and the word boundary, as predicted by the
155 Possible Word Constraint model (Norris et al., 1997). The perception of prosodic misalignment and
156 the role that syllable structure plays on speech segmentation was addressed by Mehler et al. (1981), as
157 previously mentioned. Nevertheless, subsequent attempts to replicate their results proved unclear, and the
158 role of the syllable in the segmentation of continuous speech remains ambiguous (for a summary, see Côté,
159 2012). For example, Cutler et al. (1986) failed to observe the same effect for English, and consequently
160 they suggested that the cross-linguistic rhythmic differences between syllable-timed (French) and stressed-
161 timed (English) languages might be responsible for the disparate results. This same idea has been put
162 forward in later studies, where it has been argued that the syllable plays a role in the segmentation of
163 continuous speech to the extent that it can effectively provide exploitable cues to the listeners attuned to
164 the rhythmic properties of their native language (Bien et al., 2015; Feng et al., 2019).

165 In this sense, an interesting study on the misalignment of word and syllable boundaries was conducted
166 for French by Dumay et al. (2002), who asked participants to press a button as soon as they detected
167 a word embedded in a pseudoword. Thus, using a word-spotting task they observed that French words
168 were more quickly identified when aligned with a syllable boundary, whether it appeared embedded at
169 the offset—e.g. such as *lac* ('lake') in *zun.lac* vs. *zu.glac*—or at the onset of pseudowords—e.g. *raid* ('stiff')
170 in *raid.vonce* vs. *rai.drupe*. In the offset condition, alignment not only yielded faster responses but also
171 fewer errors. Consequently, it was concluded that those cases where a misalignment existed between an
172 actual word boundary and a syllable boundary were more difficult to detect. This was not only observed
173 for French, generally held as a syllable-timed language, but also for a language considered as stress-timed
174 like Dutch (McQueen, 1998).

175 Interestingly, a naturally occurring case of misalignment between word and syllable boundaries is pre-
176 cisely that found in resyllabification contexts. Yet, contrary to the results obtained in the experiment
177 conducted by Dumay and his colleagues, derived onsets have not been reported to impede, but actu-
178 ally to help listeners recognize resyllabified words in French (Gaskell et al., 2002; Spinelli et al., 2003;
179 Tremblay & Spinelli, 2013). After conducting several priming experiments, Gaskell and his colleagues
180 (2002) found faster lexical activation of resyllabified words—whether through *enchaînement* or *liaison*—
181 when compared to words starting with canonical onsets. They suggest that the resyllabified consonant,
182 although at first ambiguous in its parsing, may be interpreted drawing on the presumed distinct phonetic
183 cues of derived onsets (and possibly with the aid of lexical knowledge). Once it is recognized as a derived
184 onset, it may help to predict that the following context is a vowel-initial word—since that is what licenses
185 resyllabification—, and this may give listeners an advantage in speech processing.

186 The results of Gaskell et al. (2002) just mentioned have been replicated by other studies (Spinelli et al.,

187 2003; Tremblay & Spinelli, 2014; but see Tremblay & Spinelli, 2013, for different results, which were
188 related to the different distribution of each segment). For example, Spinelli et al. (2003) used minimal
189 pairs of two-word sequences at the end of a carrier sentence to study what cues French native speakers
190 exploit in word segmentation. Thus, by means of *liaison*, sentences such as *C'est le dernier oignon* ('It's
191 the last onion') and *C'est le dernier rognon* ('It's the last kidney') were assumed to be homophonous and
192 have the same syllabification, i.e. *der.nie.ro.gnon*. In a series of experiments using a cross-modal priming
193 paradigm, subjects received auditorily the prime sentence—the one containing the derived onset, the
194 canonical onset, or either one of two unambiguous conditions—, upon which they had to make a lexical
195 decision when the visual target appeared on the screen short before the last word was completely uttered.
196 Spinelli and her colleagues reported that listeners were able to exploit the subphonemic cues present in
197 the stimulus and correctly identify the originally intended two-word sequence, whether a derived onset
198 or a canonical onset.

199 It must be borne in mind that *liaison* contexts such as those examined by the study just mentioned present
200 some important differences with respect to resyllabification through *enchaînement* (Fougeron, 2007) and
201 are thus not directly comparable to Spanish resyllabification. However, as previously reported, Gaskell et
202 al. (2002) found a similar processing advantage also for *enchaînement*, and Spinelli et al. (2003) concluded
203 that listeners' perception was driven by the significantly shorter duration measured for derived onsets with
204 respect to canonical onsets—an acoustic difference that was later confirmed also for the Spanish case, as
205 mentioned earlier (Hualde & Prieto, 2014; Jiménez-Bravo & Lahoz-Bengoechea, submitted; Strycharczuk
206 & Kohlberger, 2016).

207 Other studies on the perception of word and syllable boundaries were also conducted for French (Christophe
208 et al., 2004), Dutch (Shatzman & McQueen, 2006), and English (Davis et al., 2002), although they ana-
209 lyzed the role of duration as a cue signalling word boundaries in the context of consonant clusters. For
210 example, Christophe et al. (2004) conducted several experiments in which listeners were asked to de-
211 tect as soon as possible words such as *chat* ('cat') in a lexically ambiguous two-word sequence, e.g. *chat*
212 *grincheux* ('grumpy cat'), which also contains the word *chagrin* ('sorrow'). The results showed longer la-
213 tencies for such words in comparison to sentences lacking any lexical ambiguities. For Dutch, the study
214 conducted by Shatzman and McQueen (2006) using eye-tracking also found that durational differences
215 were exploited by listeners to locate word boundaries in ambiguous utterances, e.g. the word *pot* ('jar')
216 in *ze heeft wel eens pot gezegd* ('she said once jar') vs. the word *spot* ('spotlight') in *ze heeft wel een spot*
217 *gezegd* ('she did say a spotlight').

218 One of the earliest studies to point out the perceptual exploitation of temporal cues in the misalignment of
219 word and syllable boundaries for derived onsets was carried out for English by Lehiste (1960). She used
220 similar pairs of ambiguous utterances as those mentioned above, which were held to be homophonous,
221 e.g. *known ocean* vs. *no notion*—although she also included pairs with consonants in word-medial position
222 (*two lips* vs. *tulips*) and consonant clusters (*keep sticking* vs. *keeps ticking*). By means of a lexical decision
223 task, participants were generally able to correctly discriminate the member of the pair they had heard.
224 According to Lehiste, they did so by exploiting the subphonemic cues setting word-final consonants apart
225 from consonants in other positions—i.e. word-medial, word-initial—and from consonant clusters. From

226 all the analyzed word-final consonants, only /m/, /n/, /l/, /d/, and /p/ were reported in the results as
227 appearing both as derived onsets and as canonical onsets in the same pair—e.g. *seem able* vs. *see Mabel*.
228 The acoustic measurements conducted on the recordings from three different speakers revealed consistent
229 shorter duration of derived onsets, which was interpreted as the main acoustic cue exploited by listeners
230 in the perceptual experiment. However, each segment was also reported to possess other cues that were
231 salient enough to have also possibly driven the perception of listeners. For example, apart from the typical
232 shorter duration of derived onsets, Lehiste noted that laryngealization often followed word-final /m/ and
233 /n/; derived onsets /n/ showed a consistent lower F2 and a pattern of falling intensity; and in the case
234 of /l/, she also found different formant structures and different intensity patterns for each position.

235 Along similar lines, a perceptual study conducted for Dutch revealed that listeners were sensitive to
236 durational differences when segmenting sequences with ambiguous word boundaries, e.g. *diep in* ('deep
237 in') or *die pin* ('that pin') (Quené, 1992). In this study, further acoustic measurements for the ambiguous
238 sequences /CV#CVC/ and /CVC#VC/ also showed significant differences in amplitude rise time for the
239 second vowel—i.e. the duration from vowel onset to the maximum amplitude value—as well as significant
240 differences in amplitude decay time for the first vowel—i.e. duration from the maximal amplitude of the
241 vowel to vowel offset. However, after conducting a second experiment with manipulated stimuli, Quené
242 analyzed only consonant duration and V2 amplitude rise time, and he concluded that even if the latter had
243 partially contributed to affect listeners' perception, consonant duration was clearly the stronger perceptual
244 cue.

245 For Spanish, some acoustic measurements other than duration have also been suggested to differ between
246 derived and canonical onsets. For example, Jiménez-Bravo and Lahoz-Bengoechea (submitted) reported
247 lower values for the RMS of amplitude in both resyllabified /n/ and /l/ and in their surrounding vowels.
248 In addition, they also observed that when /l/ is a derived onset, V1 generally displays a shorter amplitude
249 decay time, and in the case of /i/ its transition to the consonant shows lower F3 values. In addition, F2
250 shows a slower transition between derived onset /l/ and a following /e/. Finally, the spectral slope,
251 expressed as H1–H2, is less steep during both V1 and /l/ in the derived condition.

252 In summary, duration stands out cross-linguistically from other acoustic-phonetic information as a crucial
253 cue in the segmentation of connected speech. Since its role in ambiguous word boundaries seems to be also
254 efficiently exploited by listeners in several languages, the question arises whether this is also the case for
255 Castilian Spanish. The acoustic data that set the background of the present study challenge the view that
256 derived onsets are indistinguishable from canonical onsets in Spanish—i.e. that complete resyllabification
257 occurs. Rather, it is suggested that derived onsets possess an intermediate phonological status different
258 from canonical codas and canonical onsets, as previously argued for Spanish /s/ (Hualde & Prieto, 2014;
259 Jiménez-Bravo & Lahoz-Bengoechea, submitted; Strycharczuk & Kohlberger, 2016). The motivation of
260 the present study is, therefore, to test whether the salient temporal cues showed for Spanish derived onset
261 /s/, /n/, and /l/ can actually be used by native speakers of Castilian Spanish to disambiguate the lexical
262 affiliation of the consonant. In order to test this we manipulate the duration of word-final and word-initial
263 consonants for minimal pairs of sentences whose meaning depends only on the lexical affiliation of the
264 consonant. In this sense, not many previous perceptual studies have manipulated duration to test the role

265 of temporal cues, and those which did analyzed *liaison* resyllabification in French (Gustafson & Bradlow,
266 2016; Shoemaker, 2009, 2014a) and resyllabified consonant clusters in Dutch (Shatzman & McQueen,
267 2006), that is, phenomena not entirely comparable to Spanish resyllabification.

268 For Castilian Spanish, as mentioned above, acoustic data have mostly focused on the temporal properties
269 of /s/ in resyllabification contexts, but it is well known that any consonant may also undergo resyllabi-
270 fication. Thus, we wish to test whether the durational differences observed for derived onset /s/ as well
271 as for /n/ and /l/ can be exploited by listeners². Therefore, according to the evidence presented, our
272 hypothesis is that a decrease in consonant duration will favour the interpretation of derived onsets by
273 listeners, showing the crucial role of duration in the segmentation of connected speech.

274 2. Methods

275 2.1. Stimuli

276 The test items used in our experiment were recorded in two recording sessions from a male native speaker
277 of Castilian Spanish. Four minimal pairs of sentences for each target consonant (/s/, /n/, and /l/) were
278 recorded showing ambiguous meaning depending on its lexical affiliation to either the first or the second
279 word, i.e. *Dijo que eras ancho* vs. *Dijo que era Sancho* ('He said you were wide-bodied' vs. 'He said he was
280 Sancho [a person's name]'), see Table A1 in Appendix for details). In addition, the prosodic structure
281 was similar between both members of each pair of sentences and care was taken to control for intonation.
282 The lexical frequency of the target words could be controlled for but only to a limited extent due to
283 the difficulty of finding minimal pairs of fully meaning sentences. Each of the sentences making up the
284 minimal pair was recorded separately in one of the two recording sessions together with 55 distractors
285 and the speaker was unaware of the purpose of the study.

286 The best repetition of each sentence was chosen—e.g. those pronounced with a brief pause between the
287 relevant words were discarded—and then used as a basis to resynthesize the stimuli for the perception
288 test. The three target consonants were then segmented in a TextGrid with Praat (Boersma & Weenink,
289 2018). The onset of /s/ was defined as the onset of frication appearing in the region of 3–5 kHz as done
290 by Strycharczuk and Kohlberger (2016), while the offset was made to coincide with the appearance of the
291 second formant of the following vowel. As for /n/ and /l/, the segmentation was based on the changes in
292 intensity and in the formant structure with respect to the surrounding vowels. Mean consonant duration
293 for /s/ was 0.082 ms ($SD = 0.018$) for derived onsets vs. 0.090 ms ($SD = 0.011$) for canonical onsets; for
294 /n/ duration was 0.050 ms ($SD = 0.012$) vs. 0.061 ms ($SD = 0.015$); and for /l/ duration was 0.051 ms
295 ($SD = 0.013$) vs. 0.056 ms ($SD = 0.014$). These data show a trend for mean duration of derived onsets
296 to be shorter than canonical onsets by 7 ms in the case of /s/, 12 ms in /n/, and 4 ms in /l/.

297 A script was then run to manipulate the duration tier and to resynthesize five versions of each of the
298 original sentences using the PSOLA method (Moulines & Charpentier, 1990) as implemented in Praat.
299 Thus, the duration of the target consonant increased progressively in five steps, ranging between 40-
300 120 ms for /s/, 30-90 ms for /n/, and 20-104 ms for /l/. The duration range for each consonant was
301 established so that the extremes were as distant as possible between each other while still sounding as an
302 acceptable production. The resulting range was later divided in five equidistant steps. The total number

303 of stimuli added up to 120, i.e. 12 minimal pairs of sentences (four per target consonant) x 2 lexical
304 affiliations x 5 steps.

305 2.2. *Participants and procedure*

306 Sixty-five native speakers of Spanish, all undergraduate students, participated in the study voluntarily.
307 Firstly, in a session supervised by one of the researchers they became familiar with the experimental
308 task, which was conducted as a forced-choice task in Praat. Then they were instructed to complete the
309 experiment independently in one sitting using headphones in a quiet environment within the following
310 week, and they submitted their data, which were anonymized.

311 Two tests were conducted. The first one included /s/ and /n/ as target consonants and was completed by
312 43 listeners, who rated a total of 80 stimuli (i.e. 1720 data points for each consonant). In the experimental
313 task, on a blank screen opened by Praat, they were able to listen only once to the sentence to be rated.
314 In order to make listeners rely on their first auditory perception, they were shown on the screen the
315 two possible interpretations of the sentence only immediately after cessation of the auditory stimulus.
316 This visual information just showed the two words for which the target consonant could be interpreted,
317 either as word-final (shown on the left-hand side of the screen) or word-initial (shown on the right) in
318 order to prevent listeners from processing the entire sentence a second time and to make them rather rely
319 on their first auditory impression. Listeners were then asked to disambiguate between the two possible
320 interpretations of the given target consonant, e.g. *eras ancho* vs. *era Sancho*, i.e. between the derived onset
321 interpretation and the canonical onset interpretation. Similarly, the second test was taken by 22 listeners,
322 who followed the same procedure as just explained and rated 40 stimuli for /l/ as target consonant (i.e.
323 880 data points). In both tests the order of the presentation of the stimuli was randomized, and none of
324 the listeners were aware of the differing characteristics of the manipulated stimuli.

325 2.3. *Measures and statistical analyses*

326 For the statistical analyses, the discrete steps used to resynthesize the stimuli with an increasing consonant
327 duration were converted to a continuous variable. Thus, for both lexical affiliations a ratio between target
328 consonant duration and the duration of its preceding vowel was calculated, as summarized in Figure 1.

329 This was made, firstly, to offer a more global understanding of the effect of small increases or decreases in
330 duration, since a continuous variable is better suited to capture the general effect of manipulation beyond
331 the limited one-to-one comparison of steps allowed by the analysis of a categorical variable. Secondly, this
332 was intended to minimize the effects of the variability in speech tempo (e.g. [Lisker, 1974](#); for a summary,
333 see [Stilp, 2020](#)). Finally, it prevents the perceptual duration of word-initial or word-final consonants
334 from being altered by the respective longer duration that characterizes both stressed vowels and vowels
335 in non-coda rhymes at a word boundary position ([Beckman et al., 1992](#); [Wightman et al., 1992](#)).

336 The statistical analyses were conducted in R ([2018](#)), version 3.6.0, using linear mixed-effects models
337 through the package *lme4* ([Bates et al., 2015](#)). In each analysis, logistic models were built using maximum
338 likelihood (Laplace Approximation) and were optimised with BOBYQA ([Powell, 2009](#)).

339 The binary dependent variable *Derived Onset* (0 = canonical onset, 1 = derived onset) was modelled as a

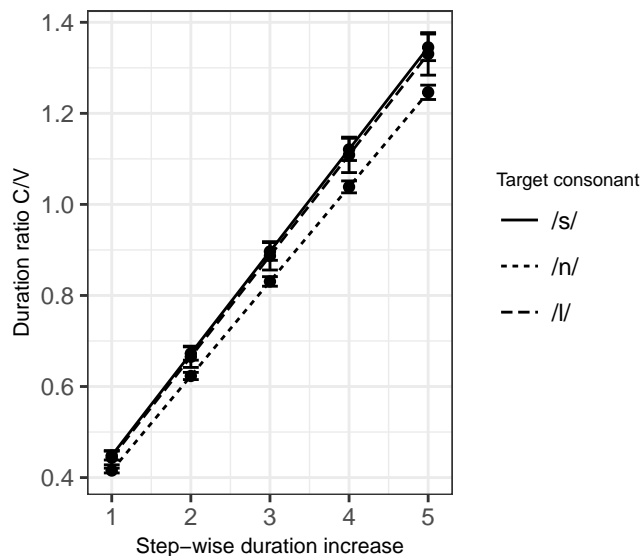


Figure 1. Relation of the initial step-wise manipulation of consonant duration and the equivalent increase of ratio between the duration of the target consonant and its preceding vowel.

340 function of duration increase of the target consonant, expressed as the standardized *z-Ratio* of consonant
 341 duration over vowel duration. A second predictor was included in the model with the original *Lexical*
 342 *Affiliation* of the target consonant (0 = canonical onset, 1 = derived onset), since both sentences in the
 343 minimal pair—i.e. one with the target consonant as derived onset and one as canonical onset—were
 344 equally manipulated, and differing cues other than duration possibly present in the original signal might
 345 have remained in the resynthesized stimuli. Furthermore, an interaction between *z-Ratio* and *Lexical*
 346 *Affiliation* was also modelled.

347 Random effects were declared for *subjects* and *items*, with the latter being specified for each pair of sen-
 348 tences. The best structure of random effects for the fitted models was determined by means of the Akaike
 349 Information Criterion (AIC). The AIC computes a correction based on the number of estimable parameters,
 350 a so-called ‘penalisation for complexity’, and its value has no meaning of its own except as a way to rank
 351 models (Akaike, 1973; Burnham & Anderson, 2002). Therefore, model selection proceeded by building
 352 a set of models differing in the progressive removal of predictors from the fully declared structure of by-
 353 items and by-subject random effects, and the AIC value of all candidate models was compared through
 354 the R package AICcmodavg (Mazerolle, 2017). This way of proceeding, i.e selecting the model having
 355 the lowest AIC value—the one that best accounts for the data—ensures the most reliable interpretation
 356 of the fixed effects (Jiménez-Bravo & Marrero-Aguilar, 2020).

357 3. Results

358 As can be seen in Figure 2, the responses given for the perception of derived onsets varied as a function
 359 of duration increase. There were consistently fewer responses for /s/ declaring perception of derived
 360 onset in the presence of a longer consonant at word boundary regardless of its original lexical affiliation.
 361 Similarly, but to a lesser extent, responses varied for /n/ according to consonant duration, with fewer

362 perceived derived onsets as a function of duration increase. Such trend proved opposite in the case of /l/,
 363 for which the number of responses perceiving a derived onset did not decrease accordingly with longer
 364 consonants at word boundary for neither lexical affiliation.

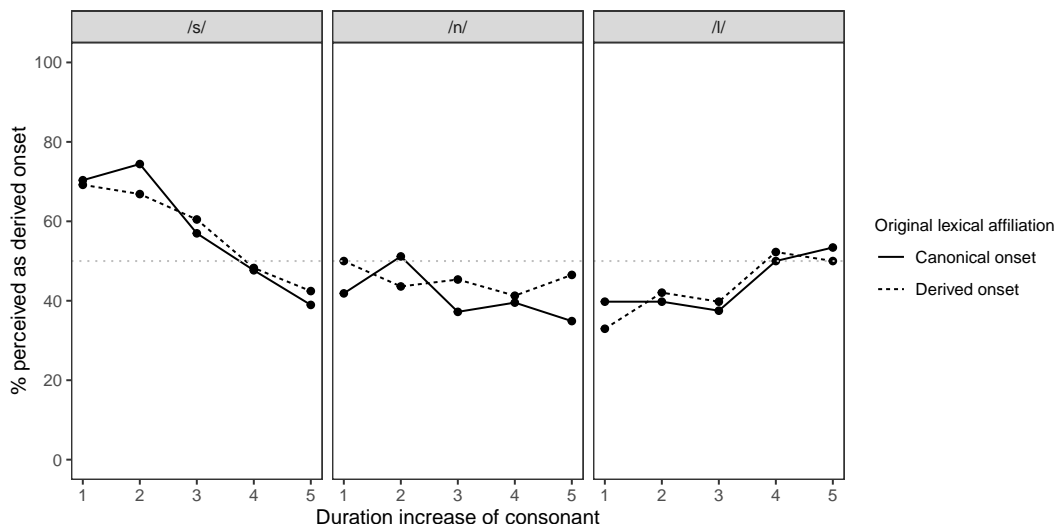


Figure 2. Relative number of responses perceiving the target consonant as a derived onset as a function of a step-wise increase in duration.

365 Three logistic regression models were independently fitted to predict the perception of derived onsets
 366 for each target consonant as a function of increase in consonant duration, its original lexical affiliation,
 367 and their interaction. The best-fitting model for /s/ (see Table A2 in Appendix for details of model selec-
 368 tion) showed a strong effect of consonant duration in the perception of the lexical boundary ($\beta = -0.67$,
 369 $SE = 0.24$, $z = -2.84$, $p = 0.004$, $OR = 0.51$). Besides, as interpreted from the intercept, the probability of
 370 participants to perceive a derived onset was 58.0% ($\beta = 0.32$, $SE = 0.16$, $z = 2.02$, $p = 0.04$, $OR = 1.38$; see
 371 Table 1) at the mean ratio of the duration of the consonant over that of the vowel ($M = 0.90$, $SD = 0.37$).
 372 This significance is in itself of little value because it all comes down to the specific point along the ma-
 373 nipulation continuum where the durational reference is situated. It is rather the increasing or decreasing
 374 of the consonant duration that is relevant inasmuch as they affect the listeners' perception.

375 For the effect of increase in consonant duration, the beta value, which corresponds to the slope of the
 376 logistic regression curve at its halfway point, can be interpreted as a decrease of 16.6% in the probability of
 377 the target consonant to be perceived as a derived onset for a ratio increase of 0.37, one standard deviation
 378 of the duration of the consonant over that of the preceding vowel. This effect was the same regardless of
 379 the original lexical affiliation of the consonant, as shown by the non-significant (ns) interaction. Besides,
 380 with C/V ratio at its mean ($M = 0.90$, $SD = 0.37$), the original lexical affiliation of the target consonant
 381 did not have any effect on how it was perceived by the listeners.

382 In the case of /n/ the responses for some duration steps were found to be near 50% (chance level), as can
 383 be seen in Figure 2. However, our statistical model showed that an increase in consonant duration—as
 384 also observed for /s/—significantly reduced the probability for /n/ to be perceived as a derived onset
 385 ($\beta = -0.20$, $SE = 0.08$, $z = -2.52$, $p = 0.01$, $OR = 0.82$). This effect corresponded to a modest decrease of

386 4.6% in the probability to perceive a derived onset when the C/V ratio mean ($M = 0.83$, $SD = 0.31$) was
 387 increased by 0.31, i.e. one standard deviation of the standardized variable. In addition, listeners were sen-
 388 sitive to the original lexical affiliation of the consonant ($\beta = 0.22$, $SE = 0.10$, $z = 2.15$, $p = 0.03$, $OR = 1.25$),
 389 so that they were 5.4% more likely to perceive a derived onset when the target consonant was originally
 390 also a derived onset (see Table A2 in Appendix for details of model selection).

391 As for /l/, instead, no main effect was observed for the increase of consonant duration, although as can be
 392 seen in Figure 2 a trend exists running in the direction opposite to that observed for /s/ and /n/. Besides,
 393 after controlling for the duration of the target consonant, the original affiliation of the consonant did not
 394 have any effect on the perception of word boundary by listeners. Table 1 and Figure 3 offer a summary
 395 of the results obtained for the three target consonants.

Table 1. Summary of fixed effects from a logistic mixed-effects regression model predicting perception of target consonant as a derived onset.

Predictor	β (SE)	p	95% CI for odds ratio		
			Lower limit	Odds ratio	Upper limit
Target consonant /s/					
Intercept	0.32 (0.16)	0.04 *	1.01	1.38	1.89
z-Ratio C/V	-0.67 (0.24)	0.004 **	0.32	0.51	0.81
Affiliation (1 = DerOnset)	0.03 (0.11)	ns	0.84	1.04	1.28
Affiliation x z-Ratio C/V	0.07 (0.11)	ns	0.86	1.07	1.34
Target consonant /n/					
Intercept	-0.42 (0.22)	ns	0.43	0.66	1.01
z-Ratio C/V	-0.20 (0.08)	0.01 *	0.70	0.82	0.96
Affiliation (1 = DerOnset)	0.22 (0.10)	0.03 *	1.02	1.25	1.52
Affiliation x z-Ratio C/V	0.16 (0.11)	ns	0.95	1.17	1.44
Target consonant /l/					
Intercept	-0.30 (0.40)	ns	0.34	0.74	1.64
z-Ratio C/V	0.30 (0.18)	ns	0.95	1.35	1.92
Affiliation (1 = DerOnset)	-0.16 (0.15)	ns	0.64	0.85	1.15
Affiliation x z-Ratio C/V	-0.08 (0.15)	ns	0.69	0.93	1.24

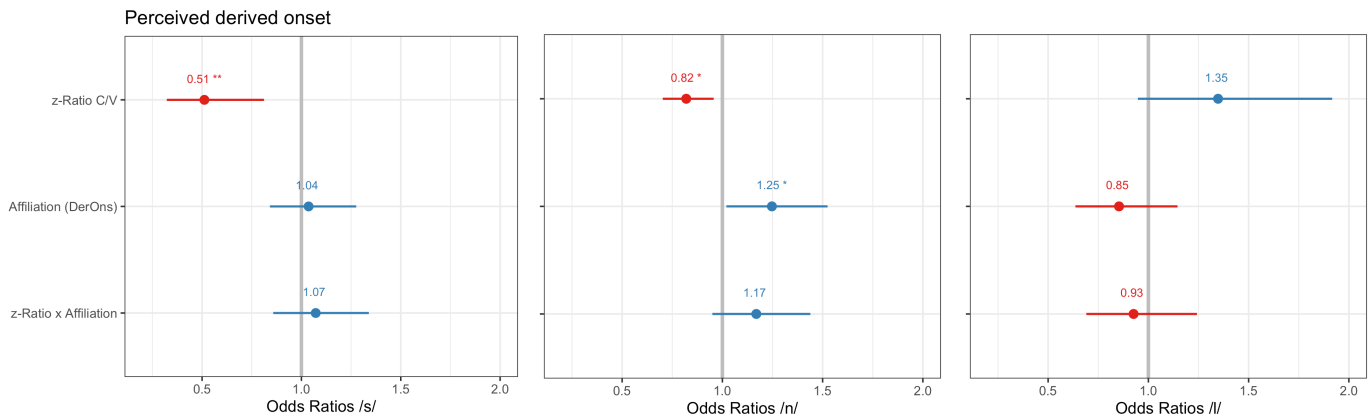


Figure 3. Forest plot showing odds ratios for main effects— z -Ratio C/V and Lexical Affiliation (1 = Derived Onset)—and their interaction to predict perception of target consonant as derived onset. For $OR < 1$, effect size equals $1/OR$. Error bars are 95% CI.

396 4. Discussion

397 The present study analyzed the role of consonant duration as a perceptual cue in contexts of ambigu-
398 ous resyllabification in Spanish. Previous evidence has pointed to the intermediate status of word-final
399 codas—specifically /s/—appearing in connected speech as word-initial resyllabified consonants (i.e. de-
400 rived onsets) in Castilian Spanish, which show a significantly different duration compared to both word-
401 final consonants (i.e. canonical codas) and word-initial consonants (i.e. canonical onsets) (Strycharczuk
402 & Kohlberger, 2016). Using a lexical decision task, listeners had to attend to the manipulated durational
403 differences to decide on the correct lexical affiliation of the ambiguous consonant—/s/, /n/, or /l/—
404 between one of two words included in contextual sentences. Our results clearly showed that listeners’
405 perception of ambiguous /s/ was driven by the duration of the consonant, so that a decrease in duration
406 made listeners bias their judgments in favour of /s/ as a derived onset regardless of its original lexical
407 affiliation. A similar, yet more modest effect of duration decrease, was observed for /n/, which addition-
408 ally revealed that listeners were sensitive to other cues present in the signal to disambiguate meaning,
409 since the original lexical affiliation of the consonant helped participants to correctly perceive /n/ as a
410 canonical onset. Besides, /n/ was the only target consonant for which this effect was observed. Probably
411 the most surprising finding of our study, and contrary to our initial hypothesis, is that the effect of dura-
412 tion was not observed for /l/, which even showed a trend opposite to that of /s/ and /n/ as a result of
413 the manipulation. All in all, we have confirmed that duration can certainly be exploited by listeners in
414 Castilian Spanish in the case of /s/ and /n/ in the segmentation of continuous speech, as predicted from
415 previous acoustic results. However, consonant identity seems to have a different effect on how listeners
416 make use of such a temporal cue, especially since such duration increase proved non-significant in the
417 case of /l/.

418 One possible interpretation might be that listeners’ perception is actually driven by segment-specific cues.
419 Alternatively, it would seem logical to think that one single acoustic cue actually exists that unifies the
420 behaviour of any segment—whatever its identity—for the sheer reason of being resyllabified. In this sense,
421 regardless of the structural interpretation that resyllabified segments may be given, they constitute one
422 phonological object and, as such, we could expect them to have some phonetic behaviour in common.
423 Considering the prosodic nature of this phonological object, such a common cue is very likely to be
424 suprasegmental. The study by Jiménez-Bravo and Lahoz-Bengoechea (submitted) described a variety of
425 acoustic features for /s/, /n/, and /l/ both as derived and as canonical onsets, as well as their surrounding
426 vowels, including measurements for different aspects of pitch, amplitude, spectral characteristics, and
427 temporal cues. Some of these parameters signalled differences in the lexical affiliation of the consonant,
428 but consonant duration was the only one that proved consistent across all three segments /s/, /n/, and
429 /l/, with shorter duration cueing derived onsets, as had previously been reported for /s/ (Hualde & Prieto,
430 2014; Strycharczuk & Kohlberger, 2016).

431 It is well known that duration is cross-linguistically affected by changes in prosodic structure and by the
432 presence of word boundaries (Byrd et al., 2000; Endress & Hauser, 2010; Monaghan et al., 2013; Tao
433 et al., 2018; Turk & Shattuck-Hufnagel, 2000). Furthermore, such changes not only affect segment dura-
434 tion but concomitantly also alter their articulatory and acoustic properties (Byrd et al., 2009; Jang et al.,
435 2018; Mooshammer et al., 2012; Scobbie & Pouplier, 2010). Durational variation in the proximity of

436 word boundaries has been observed for both vowels and consonants alike, although the perceptual effect
437 of segment lengthening is more consistent for consonants than for vowels (Monaghan et al., 2013; Ordin
438 et al., 2017; Tyler & Cutler, 2009). Consonant lengthening is well attested both as pre-boundary length-
439 ening and as domain-initial (post-boundary) strengthening (Byrd et al., 2000; Lehiste, 1960; Fougeron
440 & Keating, 1997; Smith & Hawkins, 2012; Tao et al., 2018; Turk & Shattuck-Hufnagel, 2000; Wightman
441 et al., 1992). Furthermore and differently from other cues, this longer segment duration associated to
442 word boundaries can be consistently relied on by listeners even in the segmentation of an unfamiliar
443 language (Endress & Hauser, 2010), which supports previous findings for the cross-linguistic relevance
444 of duration as a segmentation cue.

445 Similarly, in the present study we observed that listeners did use consonant duration as a cue to correctly
446 interpret the lexical affiliation of /s/ and /n/, even if such an effect was not found for /l/. As can be
447 seen in Figure 2 for /l/, answers to steps 1 to 3 hardly vary as the consonant duration increases, while
448 answers to steps 4 and 5 are also similar to each other and get close to 50%. Despite the lack of effect
449 in the case of /l/, we know that consonant duration stands out as the only cue that behaves consistently
450 across all three consonants, as seen in production data (Jiménez-Bravo & Lahoz-Bengoechea, submitted).
451 Consequently, consonant duration was the cue being manipulated in order to test its perceptual effect,
452 while any other possible cues differing among the target consonants did not undergo any manipulation
453 and remained in the signal. Of all these cues, perhaps it might be necessary to take into account the
454 duration of the surrounding vowels in order to better interpret our results, especially since listeners seem
455 to judge consonant duration relative to that of the contextual environment (Stilp, 2020).

456 In this sense, reducing the duration of the target consonant without also shortening V1 did not fully
457 parallel the natural production observed for derived onset /l/ (Jiménez-Bravo & Lahoz-Bengoechea, sub-
458 mitted). As a consequence, it is possible that the listeners were confounded by such contradictory—or
459 at least incomplete—cues and therefore did not perceive derived onsets. A further explanation in ges-
460 tural terms will be given later for the different implications of the contextual vowels in the case of /l/
461 with respect to /s/ and /n/. At any rate, it may be surprising that even those stimuli resynthesized from
462 original derived onsets (dotted line in Figure 2) were not clearly perceived as such. In our view, many
463 relevant distinctions are generally cued by tiny, albeit significant, differences in production terms, but
464 more extreme differences may be needed in order to obtain clear-cut perceptual judgements, especially—
465 like in the present study—when just one cue among several is manipulated and additionally appears in
466 an ambiguous context.

467 As for answers to steps 4 and 5, some participants mentioned after the experiment that they had perceived
468 some of the longest stimuli as a double consonant but, this not being a possible choice, they had hesitated
469 in their judgements. This might explain that many of the answers to the longest consonant tokens were
470 around chance level.

471 As previously said, resyllabified consonants constitute one phonological object and are expected to behave
472 similarly from a phonetic point of view. On the one hand, we have argued that consonant duration stands
473 out as the unifying cue, as observed in production. On the other hand, we have confirmed that duration
474 can certainly be exploited by listeners in Castilian Spanish in the segmentation of continuous speech,

475 even if the effect of consonant duration could be perceptually normalized with that of the surrounding
476 segments. Nonetheless, the more nuanced results obtained in this perception study could also be due
477 to the concurrence of other cues. In this sense, listeners' perception might additionally be driven by
478 segment-specific cues or even non-acoustic information interacting with the acoustic cues present in the
479 signal. This seems to be partially at odds with our data—i.e. our manipulation of segment duration
480 did affect interpretation—, although it might be the case that lexical and semantic factors such as word
481 frequency, for example, could also play a role in dispelling ambiguity. In the following, we discuss the
482 possible role played by lexical cues, suprasegmental features other than duration, and segment-specific
483 cues.

484 4.1. *Lexical cues*

485 The integration of different cues in speech segmentation has been suggested to follow a hierarchy in
486 which lexical, semantic, and syntactic cues interact with the acoustic-phonetic information present in the
487 speech signal (Mattys et al., 2005). However, it is likely that their dynamic interplay offers compensatory
488 mechanisms in the absence of one or several of these cues (Kim et al., 2012), especially since they have
489 been suggested to contribute to word segmentation differently depending on both language and listening
490 conditions (Gustafson & Bradlow, 2016; Mattys & Melhorn, 2007; Mattys et al., 2009; Ordin et al., 2017;
491 Toro-Soto et al., 2007; Tyler & Cutler, 2009; Weiss et al., 2010).

492 As for the semantic component, it has been claimed to play an important role in disambiguating meaning
493 (White et al., 2012), and the bare two-word sequences used as stimuli in many studies have been criticized
494 for actually leaving out the semantic context (Mattys et al., 2005). The argument is that an acoustic
495 phonetic realization such as /naitɛrt/—i.e. nitrate or night rate (Lehiste, 1960)—is likely to require the
496 listener to know the context and the situation—a pharmacy vs. a garage—in order to make a suitable
497 lexical decision. In addition, it is well-known that different lexical candidates may be activated in a
498 spoken word recognition process and that the more frequent competitor is given more weight (McQueen
499 et al., 1994). In this sense, we are aware that one of the two sentences making up our minimal pairs might
500 be disfavoured on certain occasions due to its lower lexical frequency or to an unlikely combination of
501 the two lexical items. In our study, it was our intention to present the ambiguous two-word sequences
502 within sentences able to provide a rich and coherent semantic context for each pair of words, while at the
503 same time maintaining the ambiguity between both possible lexical interpretations. However, finding
504 minimal pairs of sentences is a hard task, so the lexical frequency of the target words could be controlled
505 for only to a certain degree.

506 One may wonder to what extent all these cues are actually relevant if real life cases of ambiguity are scarce.
507 In our view, they are relevant not only because they allow listeners to make the right interpretation, but
508 also because they allow them to make it fast. This speed of comprehension may mean a large difference
509 between L1 and L2 processing (Altenberg, 2005; Ito & Strange, 2009; Shoemaker, 2014b), inasmuch as
510 some of the mentioned cues are definitely language-specific. Thus, understanding the interplay of more
511 universal and more language-dependent cues might help to develop appropriate auditory training plans
512 in the context of L2 teaching.

513 4.2. *Suprasegmental features other than duration*

514 Prosody facilitates speech processing, and word boundaries and prosodic structure are signalled through
515 variation in fundamental frequency (f_0), duration, and intensity (e.g. [Christophe et al., 2004](#)). It is known
516 that the use and combination of these are subject to language specificity, so that prosodic structure mod-
517 ulates the acoustic-phonetic realization in a variety of different ways (e.g. [Cho, 2016](#); [Tyler & Cutler,](#)
518 [2009](#)). In this sense, next to duration, changes in f_0 associated to intonation and stress have also been
519 found to affect word segmentation.

520 For example, Tyler and Cutler ([2009](#)) observed that the segmenting cue offered by an f_0 increase was
521 exploited by listeners to the extent that it matched the prosodic structure of their language. In the seg-
522 mentation of a nonsense sequence of syllables, listeners whose language showed trochaic stress benefitted
523 more from a pitch rise at the left-edge, while listeners relying on iambic stress did so if the increase was
524 located at the right-edge. In their turn, Prieto et al. ([2010](#)) found that the temporal alignment of tonal
525 peaks in Spanish—which depends on both lexical stress and syllable structure ([Prieto & Torreira, 2007](#))—
526 varied according to its location with respect to word boundary. In their study they used minimal pairs
527 made up of ambiguous two-word utterances differing in the lexical affiliation of one syllable, whether
528 word-final or word-initial—e.g. *compraremos tazas* ('We'll buy cups') vs. *compraré mostazas* ('I'll buy mus-
529 tards'). Their results showed that the delay in pitch peak was significantly longer when a word boundary
530 did not intervene (Figure 4). Interestingly, in a subsequent perceptual experiment in the same study, they
531 also confirmed that listeners were sensible to the manipulations conducted on the alignment of pitch peak
532 to correctly identify the lexical affiliation of the syllable.



Figure 4. Scheme of the different patterns of pitch peak delay with respect to word boundary, which is represented by double bars. Adapted from [Prieto et al., 2010](#), p. 696.

533 A slightly different perceptual effect of f_0 in word segmentation was reported for French ([Spinelli et al.,](#)
534 [2010](#); [Welby, 2007](#)). Welby, for example, observed that listeners used the presence of an early tone rise
535 and its precise temporal alignment—without the typical peak delay observed in Spanish—as a cue for
536 identifying the beginning of a content word in noise-masked speech. This effect was further explored
537 by Spinelli et al. ([2010](#)), who found that the alignment of f_0 rise with the start of content words can
538 be exploited by listeners to dispel the ambiguity of two-word utterances resyllabified through *elision*, i.e.
539 *l'affiche* ('the poster') vs. *la fiche* ('the sheet').

540 Thus, back to our results, we wonder whether pitch alignment could contribute as a perceptual cue

541 to disambiguate resyllabification cases in Castilian Spanish. In this sense, pre-nuclear pitch accents of
 542 statements consist of a pre-tonic low-pitch target and a rise over the stressed syllable with a delayed peak
 543 (Estebas-Vilaplana & Prieto, 2010). Therefore, this rise might be expected to start aligned to the vowel
 544 if the preceding consonant is a derived onset, i.e. lexically affiliated to the previous word. However, as
 545 observed by Jiménez-Bravo and Lahoz-Bengoechea (submitted), resyllabified consonants show a similar
 546 alignment to that of canonical onsets (see Figure 5). This suggests that, after syllable reorganization, the
 547 consonant becomes syllable-initial and renders the start of the f_0 rise unavailable as a potential cue to
 548 distinguish between derived and canonical onsets.

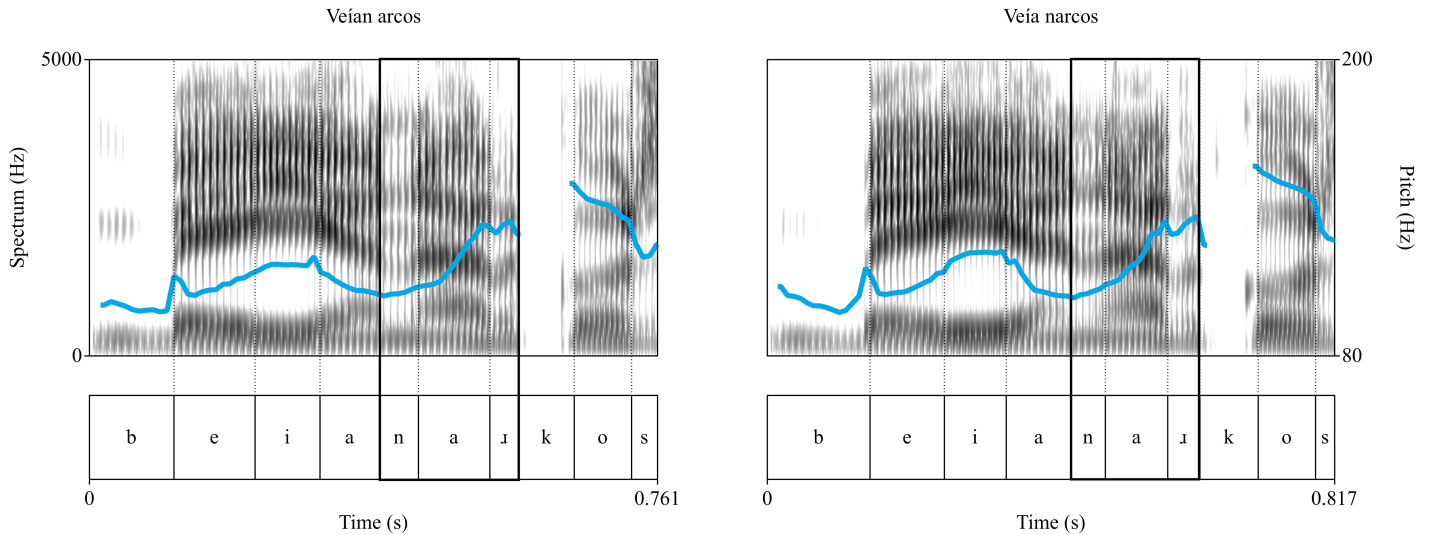


Figure 5. Example of a minimal pair showing similar pitch rise alignment between derived onset /n/ (left, ‘They saw arches’) and canonical onset /n/ (right, ‘He saw drug dealers’) (Jiménez-Bravo & Lahoz-Bengoechea, submitted).

549 4.3. Segment-specific cues

550 Studies on word segmentation reporting other salient cues different from duration often do so for a wide
 551 range of consonants (e.g. Spinelli et al., 2003), and therefore segment-specific cues are rarely described.
 552 Nonetheless, Lehiste (1960) offers details of specific consonants in her acoustic analysis of word bound-
 553 ary misalignment from recordings of several speakers. In her study she observed that, next to shorter
 554 duration, English derived onset /m/ and /n/ were characterized by laryngealization, while derived onset
 555 /n/ showed different starting frequencies of F2 and an intensity decrease. In the case of /l/, derived and
 556 canonical onsets differed both in their formant structures when preceded and followed by a vowel such
 557 as /i/ and in their intensity pattern, which consistently fell in resyllabification contexts.

558 In the case of Spanish, Jiménez-Bravo and Lahoz-Bengoechea (submitted) similarly observed a decreased
 559 intensity as an effect of the derived onset condition on both /n/ and /l/ as well as on their surrounding
 560 vowels. Also in that paper, and in line with Lehiste’s results, it was reported that derived onset /l/
 561 following /i/ always had a higher F1, while the vowel tended to show a lower F2. Additionally, the
 562 velocity of the F3 transition between V1 and /l/ tended to be less steep in the derived condition, while
 563 the velocity of the F2 transition was less steep between derived onset /l/ and a following /e/. Also,

564 the spectral slope, expressed as H1–H2, was less steep during both V1 and /l/ in the derived condition.
565 Finally, the amplitude decay time (ADT)—i.e. the lapse between the time point of maximum amplitude of
566 V1 and the offset of the vowel (Quené, 1992)—was significantly shorter for derived onset /l/, which may
567 be interpreted as a higher degree of coarticulation between both segments. As a matter of fact, /l/ was
568 the analyzed consonant that showed the greatest number of particularities varying between conditions,
569 which might be related to the fact that in the present study listeners behaved for /l/ differently than for
570 the other two target consonants.

571 In this sense, we would like to propose that such segment-specific cues differing among the three analyzed
572 consonants might be due to the different articulatory gestures involved in their production, which in their
573 turn could help to explain the differences observed in perception. Gestures have been shown to differ
574 in their temporal coordination according to syllable structure (e.g. Byrd et al., 2009; Sproat & Fujimura,
575 1993). More precisely, they are coupled in anti-phase mode when in coda, and it has been argued that
576 the relative instability resulting from this coupling mode makes segments in coda shorter than in onset
577 position (Goldstein et al., 2006).

578 Acoustically, it has been pointed out that derived onsets present a duration intermediate between canon-
579 ical codas and canonical onsets (Strycharczuk & Kohlberger, 2016), which suggests that they constitute
580 an intermediate phonological category. Furthermore, articulatory evidence of intermediate realizations
581 of derived onset /l/ has been presented for American English using EMMA (Gick, 2003). In that study ar-
582 ticulatory gestures of derived onsets were observed to be simultaneously phased to both flanking vowels,
583 thus showing articulatory properties not only of onsets but also of codas.

584 That may explain why segments in derived onset position show a shorter duration than their canon-
585 ical counterparts. This finds support in our perceptual results inasmuch as reducing segmental duration
586 induced a higher number of derived onset answers for both /s/ and /n/ although significance was not
587 reached for /l/. In this sense, our data show that listeners are indeed able to perceptually exploit the
588 acoustic differences in duration described for /s/ by Strycharczuk and Kohlberger (2016), which in our
589 study has also been confirmed for /n/.

590 Additionally, an effect of the original lexical affiliation of the consonant was found for /n/, which could be
591 interpreted based on its specific gestural configuration. Data for English obtained through MRI (Magnetic
592 Resonance Imaging) revealed that the articulatory gestures of /n/ are synchronous as canonical onsets,
593 but follow a sequential order when as derived onsets—i.e. the velum lowering gesture and the contact
594 of tongue tip at the alveolar ridge occur simultaneously in sequences such as *toe node*, but sequentially
595 in resyllabification contexts such as *tone ode* (Byrd et al., 2009). This is in line with the typical timing
596 of gestures described for /n/ in coda position, where the velum gesture is triggered in advance with
597 respect to the tongue-tip gesture (Goldstein et al., 2006; Krakow, 1993). Thus, the earlier timing of the
598 velum gesture should induce a relatively higher degree of nasalization on V1 in a derived onset context.
599 Although this seems a reasonable expectation from a gestural point of view, acoustic data, however,
600 have not yielded significant differences in the nasalization of V1 between derived and canonical onsets in
601 Spanish (Jiménez-Bravo & Lahoz-Bengoechea, submitted). Therefore, the cause underlying the perceptual
602 effect of the original lexical affiliation must be further studied attending also to the possible role of other

603 segment-specific cues.

604 Finally, the specific gestural makeup of /l/ has been described as having a tongue-body and a tongue-tip
605 gesture in Spanish (Proctor, 2009). When in coda position, the tongue-tip gesture is considered to reach
606 its target later than the tongue-body gesture (Sproat & Fujimura, 1993). In their turn, the articulatory
607 gestures of derived onset /l/ have also been observed to be anti-phase coupled, and it has been argued
608 that derived onsets possess intermediate properties between canonical codas and canonical onsets (Gick,
609 2003). In Spanish there exists evidence that could point to a tongue-tip delay of /l/ with respect to
610 the tongue-body gesture in derived onset contexts, as argued by Jiménez-Bravo and Lahoz-Bengoechea
611 (submitted) drawing on various acoustic parameters: a lower F2 of the preceding vowel; slower transitions
612 of F3 and F2 at the beginning and the end of the consonant, respectively; and a less steep H1–H2 along
613 the consonant. Consequently, in that study it is suggested that the relative anticipation of the tongue-
614 body gesture could be responsible for the aforementioned shortening of V1 in the context of derived onset
615 /l/. Therefore, the differences observed in our perceptual results might be related to the different gestural
616 configuration of /l/ with respect to both /n/ and /s/. Given the presence of two lingual gestures, it would
617 reasonable to think that a phase shift between them could interact with the contextual vowels in a more
618 crucial way than in the case of the other analyzed target consonants, possibly making the role of those
619 vowels more relevant in the perception of derived onset /l/.

620 One might think that such a tip delay should make derived onset /l/ in Spanish acquire some of the proper-
621 ties described for dark [ɫ], which is against most acoustic descriptions of Spanish /l/ (e.g. Recasens, 2012).
622 Traditionally, [ɫ] has been understood as the consequence of an earlier timing of the tongue-body retrac-
623 tion relative to the tongue-tip gesture (Sproat & Fujimura, 1993). However, further evidence suggests
624 that darkness is not actually rendered by an anticipation of the tongue-body gesture but rather depends
625 on the degree of gestural constriction achieved between the tongue-body and the pharynx (Narayanan
626 et al., 1997; Proctor, 2009). Thus, our proposal of an anti-phase coupling between the tongue-body and
627 the tongue-tip gestures does not contradict what is generally known about the acoustic properties of coda
628 /l/ in Spanish.

629 4.4. Phonological representation

630 In the face of these results the question arises as to what phonological representation derived onsets
631 should have. In other words, how can derived onsets be distinguished from both canonical onsets and
632 canonical codas if the observed phonetic differences are to be taken into account?

633 Strycharczuk and Kohlberger (2016) in the discussion of their acoustic results considered the possibility
634 of an ambisyllabic representation for derived onsets, although they objected that two different phonolog-
635 ical objects such as derived onsets and geminate consonants cannot have the same representation, and
636 reserved ambisyllabicity for the latter, especially since geminates have traditionally been analyzed in au-
637 tosegmental terms as occupying two prosodic slots, i.e. coda and onset. On the contrary, we would like to
638 suggest that geminates must be given a different phonological representation, while derived onsets may
639 actually be analyzed as ambisyllabic. Therefore, in the following we offer evidence against an ambisyl-
640 labic representation of geminates, which in our view cannot constitute one single segment simultaneously

641 associated to both the coda and the onset positions.

642 The first piece of evidence pointing to a reinterpretation of geminate consonants can be found in Cypriot
643 Greek, where it has been observed that f_0 minima are approximately aligned with an intermediate point
644 of the geminate consonant, so that the observed pitch rise takes place in its second half (Tserdanelis &
645 Arvaniti, 2001). A similar tonal alignment has also been reported for concatenated geminates in English
646 (Gao & Xu, 2010). As a result, it is possible to infer that geminates are not one single but longer consonant,
647 since otherwise tonal events would be expected to align rather with its beginning, as they do in the case
648 of singletons.

649 In support of this view, we draw on the framework of articulatory phonology, which initially posited the
650 notion of the articulatory gesture as the basic unit of phonological events (Browman & Goldstein, 1992),
651 as previously mentioned. The model of articulatory phonology was later enriched with two more types
652 of gestures: on the one hand, gestures aligned with the boundaries of prosodic constituents affecting the
653 dynamics of articulation—i.e. π -gestures, (Byrd & Saltzman, 2003)—, and on the other hand, gestures
654 targeting a H(igh) tone and L(ow) tone—i.e. tonal gestures, (Gao, 2008). The dynamics and overlap of
655 these three different types of gestures are determined by their coordination—whether in-phase or anti-
656 phase—in the gestural score of articulation, where the beginning of each gesture serves as the anchor
657 point in the control of coordination (for a review, see Hall, 2017). Specifically, the beginning of tonal
658 gestures is coordinated in-phase with the beginning of the gestures belonging to the nuclear vowel. If
659 Cypriot Greek shows a pitch turning point at about the middle of the geminate consonant, as it does,
660 this should be interpreted as the articulatory beginning of the vowel. Onset consonants are also coupled
661 in-phase with the nuclear vowel, so the onset is estimated to begin roughly in alignment with the pitch
662 turning point. Crucially, the geminate cannot be considered one single but longer articulatory gesture—if
663 that were the case, the vowel would start halfway through the consonant, in what could be seen as a sort
664 of anti-phase relation. As just said, anchor points for coordination are necessarily located at the beginning
665 of gestures. Since this is the case, and the onset is required to be in-phase with the vowel, the f_0 turning
666 point reveals the start of a new gesture, leading to the conclusion that geminates must be represented as
667 the sequence of two consonants.

668 Secondly, our interpretation of geminates as sequences of two identical segments is supported by Tilsen's
669 (2016) description of the acquisition of phonological patterns in speech development. In his account,
670 Tilsen points out that geminate consonants arise by reselecting the same motor plan as that of the previ-
671 ously realized segment. In the first stage of development, children have not learned yet how to reselect
672 action units and therefore produce geminates as short as singletons. On a subsequent stage, they can
673 reselect the motor plan, but rely solely on external feedback to know that the first instance of that plan is
674 completed prior to reselecting it. In other words, not until they hear the acoustic consequences of the first
675 unit or feel the appropriate degree of muscular tension do they proceed to the reselection. This gives rise
676 to an abnormally long geminate production. Finally at a later stage of the development, children learn
677 to anticipate the consequences of their articulation, and this so-called internalized feedback allows them
678 to reselect units in advance, so that the greater overlap between consecutive actions yields geminates as
679 long as expected.

680 Consequently, this view allows to posit a similar account for lexical geminates and for geminates con-
 681 catenated across word boundaries, both consisting of a sequence of two consonants. The difference is
 682 that lexical geminates are produced by reselecting one unit of motor planning, whereas concatenated
 683 geminates correspond to two such units, albeit identical to each other. Such articulatory difference may
 684 account for other well known phonetic and phonological differences between lexical and concatenated
 685 geminates.

686 A third piece of evidence in favour of the rearticulation of geminates is provided by Miller (1987) for
 687 Levantine Arabic, where spikes of energy corresponding to the subtle release bursts have been observed
 688 in the middle of geminate consonants on certain occasions. In line with Tilsen’s view (2016), this is
 689 suggestive of the reselection of a previously executed motor plan, which results in a new constriction
 690 gesture to prolong the consonant. Following Tilsen’s account, it is possible to interpret the appearance
 691 of such subtle bursts as arising from an inefficient internal feedback. Thus, the speaker must rely rather
 692 on the external feedback—i.e. the acoustic effect provided by the release—to detect when the gestural
 693 target has been reached prior to reselecting the motor commands that result in a geminate consonant.

694 Strycharczuk and Kohlberger (2016) discarded an ambisyllabic representation for derived onsets because
 695 they argued that such representation corresponds to geminates. If, as just explained, there are reasons
 696 against this interpretation of geminates, it is worth considering the arguments in favour of ambisyllabicity
 697 as a possible interpretation of derived onsets.

698 As previously mentioned in the introduction (§ 1.2), there is sufficient evidence to interpret resyllabified
 699 consonants as onsets after the post-lexical reorganization of syllable structure occurring in resyllabifica-
 700 tion. Additionally, it is possible to consider that derived onsets still show coda properties and have an
 701 intermediate status that points to an ambisyllabic representation (Figure 6), as will be shortly discussed.

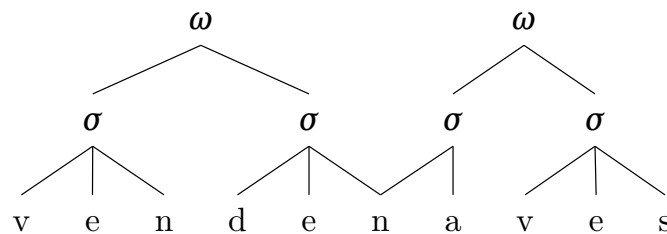


Figure 6. *Venden aves* ('they sell birds'), ambisyllabic representation of derived onset /n/.

702 The onset status of resyllabified consonants is supported by a study conducted by Cho et al. (2014) on
 703 the temporal realization of /CV/ articulatory gestures in Korean. The researchers observed that the labial
 704 gesture of derived onset /m/ presented the same pattern of articulation when compared to canonical onset
 705 /m/, i.e. the gesture was in-phase coordinated with the following vowel in both cases. A similar account
 706 can be given for Spanish tap /ɾ/, which is made up of two tongue gestures—as is the lateral /l/—, one
 707 with the body of the tongue and one with the tip (Proctor, 2009). In contrast to canonical codas, which
 708 manifest an array of allophones, this phoneme surfaces as the tap allophone in resyllabification contexts,
 709 as is typically the case of onsets. This may be interpreted as the tongue-tip gesture being coordinated

710 in-phase with the following vowel, similarly to what happens with Korean /m/. Additionally, in certain
711 dialects of Spanish, aspiration of obstruents in coda is ordered after resyllabification and therefore does
712 not target derived onsets due to the bleeding order of the rules. Thus, in Buenos Aires Argentinian Spanish,
713 for example, an utterance such as *dos alas* ('two wings') is realized as [do.sa.lah] (Kaisse, 1999), which
714 makes evident that syllable structure has reorganized, and aspiration does not affect onset /s/ due to the
715 greater stability that in-phase coupling provides. Finally, results by Jiménez-Bravo and Lahoz-Bengoechea
716 (submitted) point to the fact that resyllabified consonants receiving a pre-nuclear pitch accent show, as
717 mentioned earlier, a pitch rise alignment similar to that of typical onsets (see Figure 5).

718 Next to this, however, there is evidence suggesting that derived onsets still show coda properties. As
719 previously pointed out, gestures in coda position are in anti-phase relation—which is inherently more
720 unstable—, and this fact renders coda segments shorter than their onset counterparts. In this sense, our
721 results show that shortening the duration of the target consonant favoured the listeners' interpretation of
722 the segment as a derived onset. In addition, the segment-specific cues previously described for derived
723 onset /l/ have been argued to be indicative of the anti-phase coupling between its two lingual gestures
724 (Jiménez-Bravo & Lahoz-Bengoechea, submitted), which is characteristic of the coda position. Another
725 piece of evidence is offered by Byrd et al. (2009), who observed that the articulatory gestures of /n/ are
726 synchronous as canonical onsets, but follow a sequential order when as derived onsets, as it occurs in
727 coda position.

728 As a matter of fact, derived onsets can be interpreted as constituting an intermediate category between
729 canonical codas and canonical onsets. This is suggested by the intermediate temporal duration observed
730 for derived onsets (Strycharczuk & Kohlberger, 2016) and by their articulatory gestures, which are simul-
731 taneously phased to its two flanking vowels (Gick, 2003).

732 In our proposal, the onset status of derived onsets /s/, /n/, and /l/ is instantiated through the in-phase
733 coordination of their tongue-tip gesture with the following vowel (V2), as shown by the solid lines in
734 Figure 7. The coda status of derived onsets manifests itself through the anti-phase coordination between
735 the tongue-tip and some other possible gestures (dotted line). In the case of /s/ and /n/, the tip gesture is
736 anti-phased to the previous vowel (V1), and the tension existing between the anti-phase coordination with
737 V1 and the in-phase coordination with V2—i.e. the hallmark of what makes derived onsets ambisyllabic—
738 is in accordance with their intermediate duration. This durational effect is consistent with the findings of
739 our study, inasmuch as shorter /s/ and /n/ elicited a higher number of answers as derived onsets rather
740 than canonical onsets³. Additionally, in the case of /n/, the velum is anti-phased with the tongue tip, as
741 typically occurs in coda position. However, the velum gesture does not seem to be directly coupled with
742 V1, as data on the nasalization for Spanish suggest. According to the studies conducted by Solé (1992,
743 1995), the temporal extent of nasalization remains fairly constant and does not vary proportionally with
744 changes in vowel duration. Solé argues that nasalization in Spanish is simply a consequence of phonetic
745 mechanisms and not of a phonologically specified relation between the vowel and the velum gesture. This
746 representation, without any direct coupling between the velum and the vowel, is also the one preferred
747 by Byrd et al. (2009), even if with some reservations. As for /l/, its coda status is reflected by the anti-
748 phase coordination between the tongue-tip and tongue-body gestures, while its tongue-body gesture is

749 also anti-phased with the preceding vowel (V1). As shown in Figure 7, this results in a different pattern
 750 with respect to /s/ and /n/, which might underlie the different results obtained for /l/ in this study.

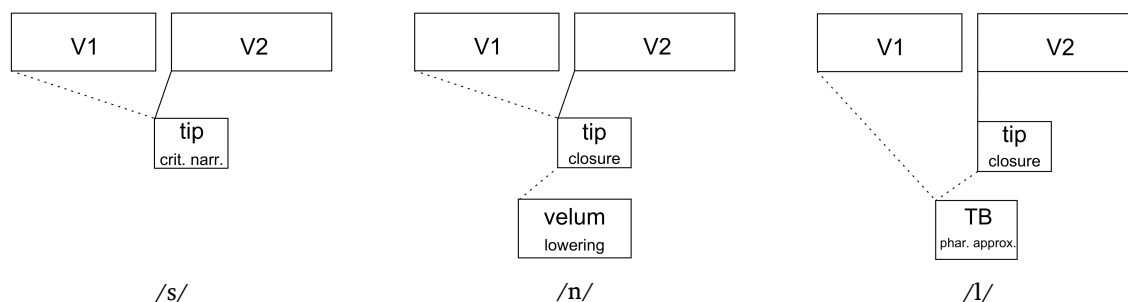


Figure 7. Graphs showing the coordination of the articulatory gestures proposed for the derived onsets analyzed in this study. Box sizes are not intended to iconically reflect acoustic duration. Solid lines represent in-phase coordination, dashed lines correspond to anti-phase coordination. Abbreviations: crit. narr. (critical narrowing), TB (tongue-body), phar. approx. (pharyngeal approximation).

751 5. Conclusions

752 This study shows that duration can be exploited in the segmentation of continuous speech in Castilian
 753 Spanish to distinguish derived onsets from canonical onsets and, therefore, to locate word boundaries in
 754 ambiguous resyllabification contexts. Our results indicate that the interpretation of ambiguous /s/ and
 755 /n/ as a derived onset by listeners was determined by the shorter duration of the consonant, while the
 756 effect of consonant duration proved non-significant for /l/. Additionally, it was observed that the identity
 757 of the segment may make extra cues available, as it was the case for /n/, where the original lexical
 758 affiliation had an effect on listeners' perception. It remains to be further investigated which segment-
 759 specific cues may be responsible for this effect.

760 There is sufficient evidence to consider resyllabified consonants as onsets after the post-lexical reorga-
 761 nization of syllable structure typical of resyllabification. Next to this, derived onsets still show coda
 762 properties—showing their intermediate status between canonical codas and canonical onsets—, which
 763 supports the idea that derived onsets should be given an ambisyllabic phonological representation. This
 764 ambisyllabic character accounts for the shorter duration of the resyllabified consonant, which in turn
 765 justifies that reducing duration favours a derived onset interpretation by listeners.

766 The differences found in the perception of the three studied consonants may be explained by the different
 767 coordination patterns of gestures that are involved in the articulation of each segment. On the one hand,
 768 /s/ and /n/ have a tongue-tip gesture that is anti-phase coordinated to the preceding vowel, while at the
 769 same time being in-phase with respect to the following vowel. On the other hand, the tongue-tip gesture
 770 of /l/ is in-phase with V2 but anti-phased with respect to the tongue-body gesture, this being in its turn
 771 anti-phased with respect to V1. This different pattern in /l/ may account for the different perceptual
 772 results inasmuch as it causes a shortening of V1—as observed for production data—, which may make it
 773 necessary to also consider the duration of this vowel to fully account for the perceptual results.

774 In our view, future research ought to address the extent to which duration can prove a cross-consonant

775 acoustic cue for word segmentation, especially in languages varying in their rhythmic typology. Addi-
776 tionally, it should be assessed whether segment-specific cues also contribute together with duration to
777 the perception of the correct parsing of the consonant.

778 Finally, it will be of great interest to confirm that the results obtained here for duration can be extended to
779 L2 listeners, as the literature suggests. Analyses should also be extended in order to assess the sensitivity
780 of listeners to segment-specific cues according to their L1, with the aim of developing appropriate auditory
781 training plans in the context of L2 teaching.

782 **Competing interests**

783 The authors declare that they have no competing interests.

784 **Acknowledgements**

785 We are most grateful to Germán Palero for recording the stimuli and to the listeners who volunteered
786 to participate in our study. We also appreciate the feedback received from the audience at presenta-
787 tions of preliminary versions of this work: at LabPhon 16, at Australian National University, and at the
788 MARCS Institute for Brain, Behaviour and Development of Western Sydney University. The presentation
789 of this work at LabPhon 16 was supported by a fund from the Department of Spanish Linguistics and
790 Literary Theory (Complutense University of Madrid) granted to the first author. Finally, we also thank
791 two anonymous reviewers for their insightful comments.

792 **Notes**

793 ¹ Measurements did not seem to be significantly different for /s/ (derived onsets: $M = 119.81$ ms, $SD = 16.01$; canonical
794 onsets: $M = 122.76$ ms, $SD = 16.14$), although it is not clear whether the same applies to /n/ (derived onsets: $M = 67.59$ ms,
795 $SD = 23.97$; canonical onsets: $M = 85.94$ ms, $SD = 10.60$).

797 ² We excluded /r/—another common coda consonant—due to the phonotactic restriction existing in Spanish for this phoneme.
798 On the one hand, it presents several allophones in coda position, but [r] is the only possibility in the case of derived onsets. On
799 the other hand, /r/ cannot appear in word-initial position. Therefore, the allophone surfacing in resyllabification cases makes
800 its lexical affiliation totally unambiguous.

801 ³ An ambisyllabic account of derived onset /s/ may be at odds with the evidence mentioned earlier (§ 1.2) for the bleeding
802 order of rules applied in certain dialects. In this sense, if /s/ were arguably still a coda, it should surface after resyllabification
803 as the aspirated allophone typically occurring in coda position, even if the aspiration rule is applied at a later stage. The only
804 reason one might provide for it not to surface as the aspirated allophone would be that the aspiration rule actually does not
805 apply to segments in coda position but rather to segments occurring *not* in onset position. Such a rule formulation is certainly
806 unusual, but although both formulations are generally equivalent, this difference would actually explain the absence of the
807 aspirated allophone in this case. Given that aspiration is an instance of weakening, its occurrence might be prevented by the
808 greater stability provided by the onset in-phase coordination, even if the segment is concurrently engaged in a coda relation.

Table A1. Speech material

Minimal pair sentences for /s/

*Dijo que **eras ancho**.*
(He said you were wide-bodied)

*Me imagino que no **querrás obras**.*
(I guess you don't want construction works)

*¿No **buscabas ocio**?*
(Weren't you looking for entertainment?)

***Tienes odio acumulado**.*
(You have accumulated hatred)

*Dijo que **era Sancho**.*
(He said he was Sancho [a person's name])

*Me imagino que no **querrá sobras**.*
(I guess he doesn't want leftovers)

*¿No **buscaba socio**?*
(Wasn't he looking for a partner?)

***Tiene sodio acumulado**.*
(It has accumulated sodium)

Minimal pair sentences for /n/

***Ven arcos** desde su ventana.*
(They see arches out their window)

***Venden aves**.*
(They sell birds)

***Proponían hombres** nuevos cuando hacía falta.*
(They proposed new men when necessary)

***Tienen hormas** nuevas.*
(They have new shoe lasts)

***Ve narcos** desde su ventana.*
(He sees drug dealers out his window)

***Vende naves**.*
(He sells ships)

***Proponía nombres** nuevos cuando hacía falta.*
(He proposed new names when necessary)

***Tiene normas** nuevas.*
(He has new rules)

Minimal pair sentences for /l/

*Recuerdo que la **vil oca** y las gallinas salieron despavoridas.*
(I remember that the vile goose and the hens ran away terrified)

*Los caprichos **del ego** son incomprensibles.*
(The whims of the ego are unfathomable)

*Todavía no tenemos ningún indicio **del ogro** que nos parezca convincente.*
(We still don't have any convincing trace of the ogre)

*Esa imagen **del oro** no se me borrará de la cabeza.*
(That image of the gold won't go out of my mind)

*Recuerdo que la **vi loca** y las gallinas salieron despavoridas.*
(I remember that I saw her mad, and the hens ran away terrified)

*Los caprichos **de lego** son incomprensibles.*
(The whims of a layman are unfathomable)

*Todavía no tenemos ningún indicio **de logro** que nos parezca convincente.*
(We still don't have any convincing sign of success)

*Esa imagen **de loro** no se me borrará de la cabeza.*
(That image of the parrot won't go out of my mind)

Table A2. Summary of AIC results for the random effect structures in GLMMs modelling perception of target consonant as a derived onset as a function of fixed effects *z-Ratio* and original *Lexical affiliation*. The model with the lower AIC value appears boldfaced. *K* indicates the number of estimated parameters in the model. The statistics associated to each model are Akaike Information Criterion (AIC), increase Δ_i of each model respect to the minimum AIC value, the Akaike weight w_i for each candidate model to be the minimal adequate model, and the log-likelihood of the model.

Model	By-items and by-subjects random effects	<i>K</i>	AIC	Δ_i	w_i	LogLikelihood
Target consonant /s/						
1	(1 sentence) + (1 subject)	6	2229.67	26.89	0.00	-1108.83
2	(1 sentence) + (z-Ratio subject)	8	2228.09	25.31	0.00	-1106.04
3	(1 sentence) + (Lexical Affiliation subject)	8	2230.16	27.38	0.00	-1107.08
4	(1 sentence) + (z-Ratio + Lexical Affiliation subject)	11	2236.98	34.20	0.00	-1107.49
5	(z-Ratio sentence) + (1 subject)	8	2206.06	3.28	0.10	-1095.03
6	(z-Ratio sentence) + (z-Ratio subject)	10	2202.78	0.00	0.53	-1091.39
7	(z-Ratio sentence) + (Lexical Affiliation subject)	10	2206.35	3.57	0.09	-1093.18
8	(z-Ratio sentence) + (z-Ratio + Lexical Affiliation subject)	13	2204.63	1.85	0.21	-1089.31
9	(Lexical Affiliation sentence) + (1 subject)	8	2233.53	30.75	0.00	-1108.77
10	(Lexical Affiliation sentence) + (z-Ratio subject)	10	2231.91	29.13	0.00	-1105.96
11	(Lexical Affiliation sentence) + (Lexical Affiliation subject)	10	2234.01	31.23	0.00	-1107.00
12	(Lexical Affiliation sentence) + (z-Ratio + Lexical Affiliation subject)	13	2233.93	31.14	0.00	-1103.96
13	(z-Ratio + Lexical Affiliation sentence) + (1 subject)	11	2211.37	8.59	0.01	-1094.68
14	(z-Ratio + Lexical Affiliation sentence) + (z-Ratio subject)	13	2208.05	5.27	0.04	-1091.03
15	(z-Ratio + Lexical Affiliation sentence) + (Lexical Affiliation subject)	13	2211.77	8.99	0.01	-1092.89
16	(z-Ratio + Lexical Affiliation sentence) + (z-Ratio + Lexical Affiliation subject)	16	2210.03	7.25	0.01	-1089.02
Target consonant /n/						
1	(1 sentence) + (1 subject)	6	2265.77	0.00	0.58	-1126.88
2	(1 sentence) + (z-Ratio subject)	8	2269.46	3.69	0.09	-1126.73
3	(1 sentence) + (Lexical Affiliation subject)	8	2269.75	3.98	0.08	-1126.88
4	(1 sentence) + (z-Ratio + Lexical Affiliation subject)	11	2277.45	9.69	0.00	-1126.73
5	(z-Ratio sentence) + (1 subject)	8	2269.22	3.46	0.10	-1126.61
6	(z-Ratio sentence) + (z-Ratio subject)	10	2272.81	7.05	0.02	-1126.41
7	(z-Ratio sentence) + (Lexical Affiliation subject)	10	2273.21	7.44	0.01	-1126.60
8	(z-Ratio sentence) + (z-Ratio + Lexical Affiliation subject)	13	2278.81	13.04	0.00	-1126.40
9	(Lexical Affiliation sentence) + (1 subject)	8	2269.77	4.00	0.08	-1126.88
10	(Lexical Affiliation sentence) + (z-Ratio subject)	10	2273.46	7.69	0.01	-1126.73
11	(Lexical Affiliation sentence) + (Lexical Affiliation subject)	10	2273.75	7.98	0.01	-1126.88
12	(Lexical Affiliation sentence) + (z-Ratio + Lexical Affiliation subject)	13	2279.45	13.69	0.00	-1126.73
13	(z-Ratio + Lexical Affiliation sentence) + (1 subject)	11	2274.17	8.40	0.01	-1126.08
14	(z-Ratio + Lexical Affiliation sentence) + (z-Ratio subject)	13	2277.77	12.00	0.00	-1125.88
15	(z-Ratio + Lexical Affiliation sentence) + (Lexical Affiliation subject)	13	2278.15	12.39	0.00	-1126.08
16	(z-Ratio + Lexical Affiliation sentence) + (z-Ratio + Lexical Affiliation subject)	16	2283.76	18.00	0.00	-1125.88
Target consonant /l/						
1	(1 sentence) + (1 subject)	6	1122.36	7.99	0.01	-555.18
2	(1 sentence) + (z-Ratio subject)	8	1124.19	9.81	0.00	-554.09
3	(1 sentence) + (Lexical Affiliation subject)	8	1125.92	11.55	0.00	-554.96
4	(1 sentence) + (z-Ratio + Lexical Affiliation subject)	11	1129.94	15.56	0.00	-553.97
5	(z-Ratio sentence) + (1 subject)	8	1114.38	0.00	0.38	-549.19
6	(z-Ratio sentence) + (z-Ratio subject)	10	1114.64	0.26	0.33	-547.32
7	(z-Ratio sentence) + (Lexical Affiliation subject)	10	1117.91	3.53	0.06	-548.95
8	(z-Ratio sentence) + (z-Ratio + Lexical Affiliation subject)	13	1120.46	6.08	0.02	-547.23
9	(Lexical Affiliation sentence) + (1 subject)	8	1123.38	9.01	0.00	-553.69
10	(Lexical Affiliation sentence) + (z-Ratio subject)	10	1125.24	10.87	0.00	-552.62
11	(Lexical Affiliation sentence) + (Lexical Affiliation subject)	10	1127.28	12.90	0.00	-553.64
12	(Lexical Affiliation sentence) + (z-Ratio + Lexical Affiliation subject)	13	1131.18	16.80	0.00	-552.59
13	(z-Ratio + Lexical Affiliation sentence) + (1 subject)	11	1117.22	2.85	0.09	-547.61
14	(z-Ratio + Lexical Affiliation sentence) + (z-Ratio subject)	13	1117.57	3.19	0.08	-545.78
15	(z-Ratio + Lexical Affiliation sentence) + (Lexical Affiliation subject)	13	1121.09	6.71	0.01	-547.54
16	(z-Ratio + Lexical Affiliation sentence) + (z-Ratio + Lexical Affiliation subject)	16	1123.51	9.14	0.00	-545.76

810 **References**

- 811 Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N.
812 Petrov, & F. Csaki (Eds.) *Proceedings of the 2nd International Symposium on Information Theory*, (pp. 267–
813 281). Budapest: Akadémiai Kiadó.
- 814 Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language*
815 *Research*, 21(4), 325–358. <http://dx.doi.org/10.1191/0267658305sr250oa>
- 816 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*.
817 *Journal of Statistical Software*, 67(1), 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- 818 Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of
819 articulatory dynamics. In G. J. Docherty, & D. R. Ladd (Eds.) *Gesture, Segment, Prosody*, (pp. 68–89).
820 Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511519918.004>
- 821 Bien, H., Bölte, J., & Zwitserlood, P. (2015). Do syllables play a role in German speech perception?
822 Behavioral and electrophysiological data from primed lexical decision. *Frontiers in Psychology*, 5, 1544.
823 <http://dx.doi.org/10.3389/fpsyg.2014.01544>
- 824 Blecua, B. (2001). *Las vibrantes del español: Manifestaciones acústicas y procesos fonéticos*. PhD dissertation,
825 Universitat Autònoma de Barcelona.
- 826 Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer. Computer program. Version
827 6.0.43, retrieved 8 September 2018 from <http://www.praat.org/>.
- 828 Boucher, V. (1988). A parameter of syllabification for *VstopV* and relative-timing invariance. *Journal of*
829 *Phonetics*, 16(3), 299–326. [http://dx.doi.org/10.1016/S0095-4470\(19\)30505-4](http://dx.doi.org/10.1016/S0095-4470(19)30505-4)
- 830 Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4), 155–
831 180. <http://dx.doi.org/10.1159/000261913>
- 832 Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: a Practice Information-*
833 *Theoretic Approach*. Berlin/Heidelberg/New York: Springer-Verlag.
- 834 Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. B.
835 Broe, & J. B. Pierrehumbert (Eds.) *Papers in Laboratory Phonology 5*, (pp. 70–87). Cambridge University
836 Press: Cambridge.
- 837 Byrd, D., & Saltzman, E. (2003). The elastic phrase: Dynamics of boundary-adjacent lengthening. *Journal*
838 *of Phonetics*, 31(2), 149–180. [http://dx.doi.org/10.1016/S0095-4470\(02\)00085-2](http://dx.doi.org/10.1016/S0095-4470(02)00085-2)
- 839 Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). Timing effects of syllable struc-
840 ture and stress on nasals: A real-time MRI examination. *Journal of Phonetics*, 37(1), 97–110.
841 <http://dx.doi.org/10.1016/j.wocn.2008>.<http://dx.doi.org/10.002>
- 842 Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and*
843 *Linguistic Compass*, 10(3), 120–141. <http://dx.doi.org/10.1111/lnc3.12178>
- 844 Cho, T., Yoon, Y., & Kim, S. (2014). Effects of prosodic boundary and syllable struc-
845 ture on the temporal realization of CV gestures. *Journal of Phonetics*, 44, 96–109.
846 <http://dx.doi.org/10.1016/j.wocn.2014.02.007>
- 847 Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase bound-
848 aries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51(4), 523–547.
849 <http://dx.doi.org/10.1016/j.jml.2004.07.001>

- 850 Côté, M.-H. (2012). The role of the syllable in the organization and realization of sound systems. In
851 A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.) *The Oxford Handbook of Laboratory Phonology*, (pp.
852 232–242). New York: Oxford University Press.
- 853 Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge,
854 MA: The MIT Press.
- 855 Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary.
856 *Computer Speech and Language*, 2(3-4), 133–142. [http://dx.doi.org/10.1016/0885-2308\(87\)90004-0](http://dx.doi.org/10.1016/0885-2308(87)90004-0)
- 857 Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, 21, 103–108.
858 [http://dx.doi.org/10.1016/S0095-4470\(19\)31323-3](http://dx.doi.org/10.1016/S0095-4470(19)31323-3)
- 859 Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1986). The syllable's differing role in
860 the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385–400.
861 [http://dx.doi.org/10.1016/0749-596X\(86\)90033-1](http://dx.doi.org/10.1016/0749-596X(86)90033-1)
- 862 Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical ac-
863 cess. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113–121.
864 <http://dx.doi.org/10.1037/0096-1523.14.1.113>
- 865 Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden-path: segmen-
866 tation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception*
867 *and Performance*, 28(1). <http://dx.doi.org/10.1037//0096-1523.28.1.218>
- 868 Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lex-
869 ical segmentation in French: Word-spotting data. *Brain and Language*, 81(1-3), 144–161.
870 <http://dx.doi.org/10.1006/brln.2001.2513>
- 871 Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive*
872 *Psychology*, 61(2), 177–199. <http://dx.doi.org/10.1016/j.cogpsych.2010.05.001>
- 873 Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto, & P. Roseano (Eds.)
874 *Transcription of intonation of the Spanish language*, (pp. 17–48). München: Lincom Europa.
- 875 Feng, C., Yue, Y., & Zhang, Q. (2019). Syllables are retrieved before segments in the
876 spoken production of Mandarin Chinese: An ERP study. *Scientific Reports*, 9, 11773.
877 <http://dx.doi.org/10.1038/s41598-019-48033-3>
- 878 Fougeron, C. (2007). Word boundaries and contrast neutralization in the case of enchaînement in French.
879 In J. Cole, & H. J.I. (Eds.) *Papers in Laboratory Phonology IX: Change in Phonology*, (pp. 609–642). Berlin:
880 Mouton de Gruyter.
- 881 Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening in prosodic domain-initial position. *The*
882 *Journal of the Acoustical Society of America*, 101, 3728–40. <http://dx.doi.org/10.1121/1.418332>
- 883 Gao, H., & Xu, Y. (2010). Ambisyllabicity in English: How real is it? In *Proceedings of The 9th Phonetics*
884 *Conference of China (PCC2010)*. Tianjin, China.
- 885 Gao, M. (2008). *Mandarin tones: An articulatory phonology account*. Ph.D. thesis, Yale University, USA.
- 886 Gaskell, M. G., Spinelli, E., & Meunier, F. (2002). Perception of resyllabification in French. *Memory &*
887 *Cognition*, 30, 798–810. <http://dx.doi.org/10.3758/BF03196435>
- 888 Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local,
889 R. Ogden, & R. Temple (Eds.) *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*,
890 (pp. 222–236). Cambridge: Cambridge University Press.

- 891 Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding
892 the evolution of phonology. In M. Arbib (Ed.) *From Action to Language: The Mirror Neuron System*, (pp.
893 215–249). Cambridge: Cambridge University Press.
- 894 Gustafson, E., & Bradlow, A. R. (2016). French speech segmentation in liaison contexts by L1 and
895 L2 listeners. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1–16.
896 <http://dx.doi.org/10.5334/labphon.59>
- 897 Hall, N. (2017). Articulatory Phonology. In S. J. Hannahs, & A. R. K. Bosch (Eds.) *The Routledge Handbook*
898 *of Phonological Theory*, (pp. 530–552). New York: Routledge.
- 899 Hoard, J. (1966). Juncture and syllable structure in English. *Phonetica*, 15(2), 96–109.
900 <http://dx.doi.org/10.1159/000258541>
- 901 Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge, MA: Cambridge University Press.
- 902 Hualde, J. I., & Prieto, P. (2014). Lenition of intervocalic alveolar fricatives in Catalan and Spanish.
903 *Phonetica*, 71, 109–127. <http://dx.doi.org/10.1159/000368197>
- 904 Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese
905 second language learners of English. *The Journal of the Acoustical Society of America*, 125(4), 2348–
906 2360. <http://dx.doi.org/10.1121/1.3082103>
- 907 Jang, J., Kim, S., & Cho, T. (2018). Focus and boundary effects on coarticulatory vowel nasalization in
908 Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical*
909 *Society of America*, 144(1), EL33–EL39. <http://dx.doi.org/10.1121/1.5044641>
- 910 Jiménez-Bravo, M., & Lahoz-Bengoechea, J. M. (submitted). Acoustic cues to resyllabification in Spanish.
- 911 Jiménez-Bravo, M., & Marrero-Aguiar, V. (2020). Multimodal Perception of Prominence in Spontaneous
912 Speech: A Methodological Proposal using Mixed Models and AIC. *Speech Communication*, 124, 28–45.
- 913 Kaisse, E. M. (1999). Resyllabification precedes all segmental rules. In J.-M. Authier, B. E. Bul-
914 lock, & L. A. Reed (Eds.) *Formal perspectives on Romance linguistics: Selected papers from the 28th Lin-*
915 *guistic Symposium on Romance Languages (LSRL XVIII)*, (pp. 197–210). Amsterdam: John Benjamins.
916 <http://dx.doi.org/10.1075/cilt.185.15kai>
- 917 Kim, D., Stephens, J. D., & Pitt, M. A. (2012). How does context play a part in splitting words apart?
918 Production and perception of word boundaries in casual speech. *Journal of Memory and Language*, 66(4),
919 509–529. <http://dx.doi.org/10.1016/j.jml.2011.12.007>
- 920 Krakow, R. A. (1993). Nonsegmental influences on velum movement patterns: syllables, sentences, stress,
921 and speaking rate. In M. K. Huffman, & R. A. Krakow (Eds.) *Nasals, nasalization, and the velum*, (pp.
922 87–116). New York: Academic Press.
- 923 Lee, Y., Kaiser, E., & Goldstein, L. (2019). I scream for ice cream: Resolving lexical ambiguity
924 with sub-phonemic information. [*published online ahead of print, 2019 Aug 6*]. *Language and Speech*.
925 <http://dx.doi.org/10.1177/0023830919866870>
- 926 Lehiste, I. (1960). An acoustic–phonetic study of internal open juncture. *Phonetica*, 5(supplementum),
927 5–54. <http://dx.doi.org/10.1159/000258062>
- 928 Lisker, L. (1974). On time and timing in speech. In T. A. Sebeok (Ed.) *Current Trends in Linguistics, Vol.*
929 *12*, (pp. 2387–2418). The Hague: Mouton.
- 930 Mattys, S. L., Brooks, J., & M., C. (2009). Recognizing speech under a processing load:

- 931 Dissociating energetic from informational factors. *Cognitive Psychology*, 59(3), 203–243.
932 <http://dx.doi.org/10.1016/j.cogpsych.2009.04.001>
- 933 Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the percep-
934 tion of word boundaries. *The Journal of the Acoustical Society of America*, 122(1), 554–567.
935 <http://dx.doi.org/10.1121/1.2735105>
- 936 Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation
937 cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477–500.
938 <http://dx.doi.org/10.1037/0096-3445.134.4.477>
- 939 Mazerolle, M. J. (2017). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R
940 package version 2.1-1, URL: <https://cran.r-project.org/package=AICcmodavg>.
- 941 McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and*
942 *Language*, 39(1), 21–46. <http://dx.doi.org/10.1006/jmla.1998.2568>
- 943 McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words
944 in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 621–638.
- 945 Mehler, D., Dommerges, J., Frauenfelder, U., & Seguí, J. (1981). The syllable's role
946 in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20(3), 298–305.
947 [http://dx.doi.org/10.1016/S0022-5371\(81\)90450-3](http://dx.doi.org/10.1016/S0022-5371(81)90450-3)
- 948 Michnowicz, J., & Kagan, L. (2016). On glottal stops in Yucatan Spanish. In S. Sessarego, & F. Tejedo-
949 Herrero (Eds.) *Spanish Language and Sociolinguistic Analysis. Hispanic and Lusophone Linguistics*, 8, (pp.
950 217–240). Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/ihll.8.09mic>
- 951 Miller, A. M. (1987). Phonetic characteristics of Levantine Arabic geminates with differing morpheme
952 and syllable structures. *Working Papers in Linguistics (Ohio State University)*, 36, 120–141.
- 953 Monaghan, P., White, L., & Merkx, M. M. (2013). Disambiguating durational cues for
954 speech segmentation. *The Journal of the Acoustical Society of America*, 134(1), EL45–EL51.
955 <http://dx.doi.org/10.1121/1.4809775>
- 956 Mooshammer, C., Goldstein, L., Nam, H., McClure, S., Saltzman, E., & Tiede, M. (2012). Bridg-
957 ing planning and execution: Temporal planning of syllables. *Journal of Phonetics*, 40(3), 374–389.
958 <https://doi.org/http://dx.doi.org/10.1016/j.wocn.2012.02.002>
- 959 Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques
960 for text-to-speech synthesis using diphones. *Speech Communication*, 9(5-6), 453–467.
961 [https://doi.org/10.1016/0167-6393\(90\)90021-Z](https://doi.org/10.1016/0167-6393(90)90021-Z)
- 962 Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *The Journal of the*
963 *Acoustical Society of America*, 62(3), 714–719. <http://dx.doi.org/10.1121/1.381583>
- 964 Narayanan, S. S., Alwan, A. A., & Haker, K. (1997). Toward articulatory-acoustic models for liquid
965 approximants based on MRI and EPG data. Part I. The laterals. *The Journal of the Acoustical Society of*
966 *America*, 101(2), 1064–1077. <http://dx.doi.org/10.1121/1.418030>
- 967 Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word con-
968 straint in the segmentation of continuous speech. *Cognitive Psychology*, 34, 191–243.
969 <http://dx.doi.org/10.1006/cogp.1997.0671>
- 970 Ordin, M., Polyanskaya, L., Laka, I., & Nespors, M. (2017). Cross-linguistic differences in the use of

971 durational cues for the segmentation of a novel language. *Memory & Cognition*, 45(5), 863–876.
972 <http://dx.doi.org/10.3758/s13421-017-0700-9>

973 Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese.
974 *Journal of Memory and Language*, 32(2), 258–278. <https://doi.org/10.1006/jmla.1993.1014>

975 Pike, L. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

976 Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without deriva-
977 tives. *Department of Applied Mathematics and Theoretical Physics, Cambridge England, technical report*
978 *NA2009/06*.

979 Prieto, P., Estebas-Vilaplana, E., & Vanrell, M. M. (2010). The relevance of prosodic structure in tonal
980 articulation Edge effects at the prosodic word level in Catalan and Spanish. *Journal of Phonetics*, 38(4),
981 687–705. <https://doi.org/10.1016/j.wocn.2010.10.004>

982 Prieto, P., & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable struc-
983 ture and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, 35(4), 473–500.
984 <http://dx.doi.org/10.1016/j.wocn.2007.01.001>

985 Proctor, M. I. (2009). *Gestural characterization of a phonological class: The liquids*. PhD dissertation, Yale
986 University.

987 Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20, 331–350.
988 [http://dx.doi.org/10.1016/S0095-4470\(19\)30638-2](http://dx.doi.org/10.1016/S0095-4470(19)30638-2)

989 R Development Core Team (2018). *R: A language and environment for statistical computing*. R Foundation
990 for Statistical Computing, Vienna, Austria.

991 Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Com-*
992 *munication*, 54, 368–383.

993 Robinson, K. (2012). The dialectology of syllabification: A review of variation in the Ecuadorian High-
994 lands. *Romance Philology*, 66, 115–145. <http://dx.doi.org/10.1484/j.rph.5.100801>

995 Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal*
996 *of Memory and Language*, 35(4), 606–621. <http://dx.doi.org/10.1006/jmla.1996.0032>

997 Scarpace, D. (2017). *The acquisition of resyllabification in Spanish by English speakers*. Ph.D. thesis, Univer-
998 sity of Illinois at Urbana-Champaign, USA.

999 Scobbie, J. M., & Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG
1000 study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics*, 38, 240–259.
1001 <https://doi.org/10.1016/j.wocn.2009.10.005>

1002 Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and
1003 Spanish. *Journal of Memory & Language*, 31, 18–32. [https://doi.org/10.1016/0749-596X\(92\)90003-G](https://doi.org/10.1016/0749-596X(92)90003-G)

1004 Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word
1005 recognition. *Perception & Psychophysics*, 68, 1–16. <http://dx.doi.org/10.3758/BF03193651>

1006 Shoemaker, E. (2014a). Durational cues to word recognition in spoken French. *Applied Psycholinguistics*,
1007 35(2), 243–273. <http://dx.doi.org/10.1017/S0142716412000380>

1008 Shoemaker, E. (2014b). The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Stud-*
1009 *ies in Second Language Acquisition*, 36(4), 709–731. <http://dx.doi.org/10.1017/S027226311400014X>

1010 Shoemaker, E. M. (2009). *Acoustic cues to speech segmentation in spoken French: Native and non-native*
1011 *strategies*. Ph.D. thesis, University of Texas at Austin, USA.

- 1012 Smith, R., & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word
1013 boundaries. *Journal of Phonetics*, 40(2), 213–233. <http://dx.doi.org/10.1016/j.wocn.2011.11.003>
- 1014 Solé, M. (1992). Phonetic and phonological processes: the case of nasalization. *Language and Speech*,
1015 35(1-2), 29–43.
- 1016 Solé, M. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasal-
1017 ization. *Language and Speech*, 38(1), 1–23.
- 1018 Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmen-
1019 tation in phonemically identical sequences. *Attention, Perception, & Psychophysics*, 72(3), 775–787.
1020 <http://dx.doi.org/10.3758/APP.72.3.775>
- 1021 Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of*
1022 *Memory and Language*, 48(2), 233–254. [http://dx.doi.org/10.1016/S0749-596X\(02\)00513-2](http://dx.doi.org/10.1016/S0749-596X(02)00513-2)
- 1023 Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its im-
1024 plications for phonetic implementation. *Journal of Phonetics*, 21(3), 291–311.
1025 [http://dx.doi.org/10.1016/S0095-4470\(19\)31340-3](http://dx.doi.org/10.1016/S0095-4470(19)31340-3)
- 1026 Stilp, C. (2020). Acoustic context effects in speech perception. *WIREs Cognitive Science*, 11, e1517.
1027 <http://dx.doi.org/10.1002/wcs.1517>
- 1028 Strycharczuk, P., & Kohlberger, M. (2016). Resyllabification reconsidered: On the durational properties
1029 of word-final /s/ in Spanish. *Laboratory Phonology*, 7(1), 1–24. <http://dx.doi.org/10.5334/labphon.5>
- 1030 Tao, J., Torreira, F., & Clayards, M. (2018). Durational cues to word boundaries in spontaneous speech.
1031 In *Proceedings of 9th International Conference on Speech Prosody 2018*, (pp. 240–244). Poznan, Poland.
1032 <http://dx.doi.org/10.21437/SpeechProsody.2018-49>
- 1033 Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological
1034 structure. *Journal of Phonetics*, 55, 53–77. <http://dx.doi.org/10.1016/j.wocn.2015.11.005>
- 1035 Toro-Soto, J. M., Rodríguez-Fornells, A., & Sebastián-Gallés, N. (2007). Stress placement and word seg-
1036 mentation by Spanish speakers. *Psicológica*, 28, 167–176.
- 1037 Trawick, S., & Michnowicz, J. (2019). Glottal insertion before vowel-initial words in the Spanish of
1038 Asunción, Paraguay. In G. L. Thompson, & S. M. Alvord (Eds.) *Contact, community and connections:*
1039 *Current approaches to Spanish in multilingual populations*, (pp. 147–171). Wilmington, DE: Vernon Press.
- 1040 Tremblay, A., & Spinelli, E. (2013). Segmenting liaison-initial words: The role
1041 of predictive dependencies. *Language and Cognitive Processes*, 28(8), 1093–1113.
1042 <http://dx.doi.org/10.1080/01690965.2012.689306>
- 1043 Tremblay, A., & Spinelli, E. (2014). Utilisation d'indices acoustico-phonétiques dans la reconnaissance
1044 des mots en contexte de liaison. In C. Soum-Favaro, A. Coquillon, & C. J.-P. (Eds.) *La liaison : approches*
1045 *contemporaines*, (pp. 93–113). Bern: Peter Lang.
- 1046 Tserdanelis, G., & Arvaniti, A. (2001). The acoustic characteristics of geminate consonants in Cypriot
1047 Greek. In *Proceedings of the 4th International Conference on Greek Linguistics - 1999*, (pp. 29–36). Nicosia,
1048 Greece.
- 1049 Turk, A., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal*
1050 *of Phonetics*, 28(4), 397–440. <http://dx.doi.org/10.1006/jpho.2000.0123>
- 1051 Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The*
1052 *Journal of the Acoustical Society of America*, 126(1), 367–376. <http://dx.doi.org/10.1121/1.3129127>

- 1053 Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2010). Colliding cues in word segmentation: The role
1054 of cue strength and general cognitive processes. *Language and Cognitive Processes*, 25(3), 402–422.
1055 <http://dx.doi.org/10.1080/01690960903212254>
- 1056 Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation.
1057 *Speech Communication*, 49(1), 28–48. <https://doi.org/10.1016/j.specom.2006.10.005>
- 1058 White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation cues in conversational speech: Robust semantics
1059 and fragile phonotactics. *Frontiers in Psychology*, 3, 375. <http://dx.doi.org/10.3389/fpsyg.2012.00375>
- 1060 Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the
1061 vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91, 1707–1717.
1062 <http://dx.doi.org/10.1121/1.402450>