CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Emerging word segmentation abilities in European Portuguese-learning infants: new evidence for the rhythmic unit and the edge factor

Joseph BUTLER* and Sónia FROTA

Center of Linguistics, University of Lisbon, Portugal
*Corresponding author: Laboratorio de Fonetica & Lisbon Baby Lab (CLUL), Dept. Linguistica, Faculdade de Letras, Universidade de Lisboa, Alameda de Universidade, 1600-214 Lisboa, Portugal.
E-mail: drjoebutler@gmail.com; labfon@letras.ulisboa.pt

Abstract

Word segmentation plays a crucial role in language acquisition, particularly for word learning and syntax development, and possibly predicts later language abilities. Previous studies have suggested that this ability develops differently across languages, possibly affected by the languages' rhythmic properties (Rhythmic Segmentation Hypothesis) and target word location in the prosodic structure (Edge Hypothesis). The present study investigates early word segmentation in a language, European Portuguese, that exhibits both stress- and syllable-timed properties, as well as strong cues to both higher-level prosodic boundaries and the word level. Infants aged 4–10 months old were tested with target words located in utterance-medial and utterance-final positions. Evidence for word segmentation was found early in development but only for utterance-edge located target words, suggesting the more salient prosodic cues play a crucial role. There was some evidence for segmentation in utterance-medial position by 10 months, demonstrating that this ability is not yet fully developed, possibly due to mixed rhythmic properties.

Keywords: early word segmentation; mixed rhythm; prosodic edge; infant language acquisition; European Portuguese

Introduction

To learn a language, one of the abilities infants need to develop is to extract potential word-forms from continuous speech. This ability is known as 'word segmentation'. Infants must learn what cues are relevant to identify words to be able to segment speech, and to start building a lexicon. Early word segmentation plays a crucial role in language acquisition, particularly in relation to word learning and the development of syntax (Höhle & Weissenborn, 2003; Nazzi, Paterson, & Karmiloff-Smith, 2003; Shi, Cutler, Werker, & Cruickshank, 2006; Singh, Nestor, Parikh, & Yull, 2009; Singh, Reznick, & Xuehua, 2012), and may even be a predictor of later language abilities (Newman, Ratner, Jusczyk, & Jusczyk, 2006). It is thus

© Cambridge University Press 2018

crucial to understand when and how infants begin to succeed in the challenging task of segmenting word-like forms from the continuous speech stream. In the current study, we examine emerging word segmentation abilities by focusing on the role played by prosody in a language with a mixed prosodic profile: European Portuguese.

Previous studies have shown that the age at which segmentation abilities emerge varies across languages for monosyllabic words (e.g., Bosch, Figueras, Texidó, & Ramon-Casas, 2013; Gout, 2001; Höhle & Weissenborn, 2003; Jusczyk & Aslin, 1995; Nishibayashi, Goyet, & Nazzi, 2015) and bisyllabic words (e.g., Höhle & Weissenborn, 2005; Juszyck, Houston, & Newsome, 1999; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006; Nazzi, Mersad, Sundara, Iakimova, & Polka, 2014; Polka, Orena, Sundara, & Worrall, 2016; Polka & Sundara, 2012). Most of the previous work on emerging segmentation abilities has focused on languages that display stress-timed properties, such as English or German, or syllable-timed properties, such as French or Spanish. These studies led to the suggestion that prosody, namely rhythm, plays a role in early word segmentation.

The Rhythmic Segmentation Hypothesis (e.g., Mersad, Goyet, & Nazzi, 2010; for a review see Nazzi *et al.*, 2006) posits that infants may exploit the rhythmic structure of their native language to learn a language-specific (rhythmic) segmentation procedure, and use this information to identify word-forms. Learners of a stress-timed language, such as English, would first extract units starting with a stressed syllable. For example, American English-learning infants are able to successfully segment trochaic words at 7.5 months, at the same point in development where monosyllabic word segmentation is also shown, but fail with iambic words until 10.5 months (Jusczyk *et al.*, 1999). Learners of syllable-timed languages, such as French, would focus on the syllable as the rhythmic cue to segmentation. This is shown by European French-learning infants' successful segmentation of monosyllabic words as well as syllables embedded in bisyllabic words at 6 months, but their failure to segment bisyllabic (iambic) words until 8 months (Goyet, Nisibayashi, & Nazzi, 2013; Nazzi *et al.*, 2014; Nishibayashi *et al.*, 2015). Successful bisyllabic segmentation by French-learning infants before 12 months, however, seems to be constrained by passage-word order of presentation (i.e., whether the infant is familiarised with passages containing a target word and tested with isolated word lists, or the reverse) and consistent with distributional information, as suggested by the different findings reported in different studies (Goyet *et al.*, 2013; Nazzi *et al.*, 2006, 2014; Nishibayashi *et al.*, 2015). Early monosyllabic word segmentation has also been found in other syllable-timed languages, with Catalan-learning and Spanish-learning infants having been shown to successfully segment monosyllabic words at 6 months (Bosch *et al.*, 2013). In summary, these studies show an advantage in segmenting both monosyllabic and bisyllabic trochaic words in stress-timed languages (see also Höhle & Weissenborn, 2005, for German, and Kooijman, Hagoort, & Cutler, 2009, for Dutch), and an advantage in segmenting monosyllabic word-forms only in syllable-timed languages.

In recent literature there has been contradictory evidence of the perception, categorisation, and discrimination of language rhythms (by both adults and infants), based on a notion of rhythmic classes mostly grounded on timing distinctions (Arvaniti & Rodrigues, 2013; Rathcke & Smith, 2015; White, Delle Luche, & Floccia, 2016; White, Mattys, & Wiget, 2012). These studies suggest that the perception of rhythm arises from interactions between prosodic factors, such as timing, speech rate, final lengthening, or intonation, and different languages and language varieties may present different combinations of these factors impacting on rhythm perception.

Crucially, these findings do not undermine the hypothesis that there are rhythmic differences between languages (possibly due to several word-level and phrase-level prosodic properties), and these differences may play a role in early segmentation (Keij, 2017). Furthermore, the combinations of such prosodic properties may yield languages with mixed prosodic profiles (Frota & Vigário, 2001; Nespor, 1990; Nespor, Shukla, & Mehler, 2011).

The Rhythmic Segmentation Hypothesis provides a clear indication of how infants can exploit the rhythmic properties of a language for the segmentation of possible word-forms. It would therefore be interesting to investigate emerging segmentation abilities in infants learning a language that presents both stress- and syllable-timed properties. European Portuguese (EP) rhythm has been shown to have such a mixed nature, in contrast with languages like English or French (Frota & Vigário, 2001). In the case of this mixed-rhythm language, infants potentially have conflicting rhythmic properties to rely on to begin segmenting speech, and so may focus on one of these properties over the other. Perception studies with adults suggest that syllable-timed properties are the most salient ones, as adults are able to discriminate EP from Dutch on the basis of prosodic cues only (Frota, Vigário, & Martins, 2002). These findings suggest that children may identify EP as a syllable-timed language (Vigário, Frota, & Freitas, 2003). However, at the word level, penultimate stress is the most frequent pattern (although not as much as initial stress for English; Cutler & Carter, 1987), and vowel reduction in unstressed syllables is a prominent property of the language, similarly to English and unlike French or Spanish, which may suggest that stress-timed properties are also quite salient (Mateus & Andrade, 2000; Vigário, 2003).

The Rhythmic Segmentation Hypothesis does not allow a clear prediction as to how EP-learning infants may begin segmenting continuous speech. As learners of a syllable-timed language, EP infants are expected to demonstrate early segmentation abilities for monosyllabic words, similarly to French, Spanish, and Catalan infants, that is, by 6 months. As learners of a stress-timed language, EP infants would demonstrate emerging segmentation abilities for monosyllabic words at a later point in development, after 6 months (see Jusczyk & Aslin, 1995, for stressed word-forms in English; Höhle & Weissenborn, 2003, for unstressed function words in German). Another possibility is that the salience of the competing rhythmic properties may not be clear to the infant. In this conflicting environment, the language may not provide clear prosodic cues for speech segmentation, due to the mixed rhythmic properties, which may lead to a delay in the emergence of segmentation abilities.

Word position in the utterance may also play an important role in the development of early segmentation abilities, again due to prosodic factors. Utterance edges (i.e., words at the edge of an utterance) provide particularly salient cues, such as final lengthening, initial strengthening, and major pitch changes given the presence of edge tones and pitch accents (Byrd, Krivokapic, & Lee, 2006; Gussenhoven, 2015; Keating, Cho, Fougeron, & Hsu, 2003). Furthermore, phrasal prominence tends to be located at utterance-final position in many languages (Nespor & Vogel, 2007). Utterance edges are also a probable location for pauses, which may be yet another prosodic cue for utterance boundaries (Cole, 2015; Mo & Cole, 2010). As utterance edges align with word edges, words at the edge of utterances provide stronger cues to word boundaries, and these may trigger an early emergence of word segmentation skills. It has been shown that, for American English-learning infants, words located at utterance edges are easier to segment than those found in utterance-medial positions of the speech stream (Seidl & Johnson, 2006), and in fact American

English-learning infants as young as 6 months are able to successfully segment continuous speech, but only when the target monosyllabic word is located at an utterance edge (Johnson, Seidl, & Tyler, 2014). This role of utterance edges as facilitators of infants' initial segmentation attempts became known as the 'Edge Hypothesis' (Seidl & Johnson, 2006). Previous word segmentation studies did not control for position of the word within utterances, and so it is unclear whether the findings from those studies were due, or not, to some targets appearing at the edge of utterances and some in the middle. Edge effects have also been shown later in development, with children at 22 and 27 months demonstrating sensitivity to third person singular morpheme -*s* verbs in sentence-final, but not sentence-medial, positions (Sundara, Demuth, & Kuhl, 2011). Therefore, it would seem that prosodic cues provided by the sentence edge position offer a perceptual advantage for infants developing language. Cues to prosodic edges, however, may vary across languages (Frota, 2012). For example, American English, Dutch, and German, all stress-timed languages, weigh prosodic cues differently: for English, a pitch change seems to be a necessary boundary cue (Seidl, 2007); for Dutch, the presence of a pause is weighed higher (Johnson & Seidl, 2008; Kooijman *et al.*, 2009); for German, both a pitch change and pre-boundary lengthening are necessary (Wellmann, Holzgrefe, Truckenbrodt, Waterburger, & Höhle, 2012). Interestingly, infants' sensitivity to prosodic boundary cues seems to undergo a developmental change towards the specific pattern of the ambient language. English-learning infants at 4 months require the presence of a strong combination of prosodic cues (pitch change, pre-boundary lengthening, and pause) to detect clause boundaries (Seidl & Cristià, 2008). By 6 months, English-learning infants already weigh prosodic boundary cues differently, being more sensitive to the pitch change, in accordance with the adult pattern (Seidl, 2007). Although English-learning and Dutch-learning infants are both sensitive to major prosodic boundaries by 6 months, they have already tuned to the language-specific prosodic pattern, with Dutch infants requiring the presence of the pause cue, unlike English infants (Johnson & Seidl, 2008).

The Edge Hypothesis thus predicts that utterance-level prosody facilitates early word segmentation. However, this hypothesis has only been tested with infants acquiring American English (Johnson *et al.*, 2014; Seidl & Johnson, 2006). Given the language-specific nature of cues to prosodic edges, it is necessary to test this hypothesis with non-English-learning infants. EP is an interesting language for this purpose, again due to its unusual combination of prosodic properties. EP, unlike other Romance languages, provides strong cues to higher prosodic phrase boundaries, namely the utterance and the intonational phrase, as well as to prosodic word boundaries, but not to lower prosodic phrase boundaries (Frota, 2014; Frota & Prieto, 2015; Vigário, 2003). In other words, EP combines a phrase-based phonetics and phonology characteristic of other Romance languages with a word-based phonetics and phonology typical of Germanic languages (Vigário, 2003). Studies on adult speech have shown that both pitch range and lengthening are robust cues for higher-level prosodic edges in EP (Frota, 2000). The prosodic word is marked by an edge-specific phonotactics together with prominence-related cues. Word stress is particularly salient due to vowel reduction in unstressed position (i.e., the vowel system /i, e, ɛ, a, o, ɔ, u/ is reduced to [i, ɨ, ɐ, u] when vowels are unstressed), combined with the longer duration of stressed vowels (Vigário, 2003; Vigário, Frota, & Martins, 2011). These strong cues to the highest levels of prosodic structure and to the prosodic word level could provide reliable support for segmenting words found

at the edge of utterances, and so EP-learning infants should have an advantage in segmenting under those conditions.

To our knowledge, there have been no studies looking at emerging word segmentation abilities in EP-learning infants, or any other language with mixed prosodic properties. The present study is thus the first attempt to investigate emerging segmentation abilities in infants acquiring this type of language. There are two main goals for this paper. First, we aim to explore whether the mixed rhythmic properties of EP impact on the age of emergence of segmentation abilities. The second goal is to examine the effect of prosodic edges on early segmentation abilities, namely whether the position of the word in the utterance impacts on early segmentation abilities, providing the first study of the Edge Hypothesis in a language other than English.

Using a modified version of the visual familiarisation paradigm, we examined European Portuguese-learning infants' segmentation abilities from 4 to 10 months of age; that is, during the period where word segmentation has been shown to emerge in infants learning other languages (Bosch *et al.*, 2013; Gout, 2001; Höhle & Weissenborn, 2003; Johnson *et al.*, 2014; Jusczyk & Aslin, 1995; Nishibayashi *et al.*, 2015; Seidl & Johnson, 2006). The passages-first order was chosen, given that this testing order was found to promote segmentation in earlier studies (Goyet *et al.*, 2013; Nazzi *et al.*, 2006, 2014; Nishibayashi *et al.*, 2015), and was the order used in previous studies on the Edge Hypothesis. Moreover, it provides a more natural context for segmentation to arise. Monosyllabic target word-forms were presented in one of two prosodic conditions: utterance-edge-final and utterance-medial. After familiarisation with passages with target word-forms in each condition, infants were tested with sequences of isolated word-forms that either were or were not present in the familiarisation passages.

If evidence of monosyllabic segmentation is found at an early age (6 months, or even before), this will show that EP-learning infants are using the syllable as the rhythmic unit for segmentation, as expected from learners of syllable-timed languages and despite the mixed rhythm of their native language. If segmentation abilities, by contrast, are found to emerge at a later age, this would suggest the absence of a direct match between the target word-form and the basic rhythmic unit for segmentation, either because the stress-timing properties of EP are more salient to infants, or due to the conflicting cues found in a language with mixed rhythmic properties. Finally, if EP-learning infants are sensitive to the salient prosodic properties provided by utterance edges, they are expected to demonstrate earlier segmentation abilities at utterance-final than utterance-medial position. As in previous studies, recognition of familiarised targets would be reflected by different listening times to targets (word-forms present in the familiarisation) and distractors (word-forms not present in the familiarisation). Also in line with prior work, any consistent difference in looking times to targets (familiar word-forms) and distractors (unfamiliar word-forms), irrespective of direction of preference, is taken as an indication of segmentation abilities.

Method

Participants

Forty infants participated in this Experiment (17 female, mean age 7 months 19 days, range 4 months 19 days to 10 months 8 days). All were typically developing infants

raised in monolingual European Portuguese homes, recruited from the wider Lisbon area. Additionally, the data from five infants were rejected from the study, three for fussiness, one for living in a bilingual household, and one who had an older, autistic sibling.

Stimuli

Four monosyllable pseudo-words were used in this study: FUL ['fuɫ], QUEU ['kɛw], PIS ['piʃ], and SAU ['saw]. Pseudo-words were used to ensure that the infants were not familiar with any of the targets. All word-initial consonants used are frequent in EP, and monosyllabic words that end with a consonant or a glide are also frequent, both in child-directed speech and adult speech (Frota, Vigário, Martins, & Cruz, 2010; Vigário, Freitas, & Frota, 2006). [ʃ] is more frequent in word-final than word-initial position, and [ɫ] is velarised in word-final position (and syllable-finally, though it occurs mostly in word-final position, i.e., over 60%).

The pseudo-words were embedded in carrier sentences either in utterance-medial or utterance-edge-final position. Utterance-medial position was always aligned with a lower phrase boundary in half of the cases, and corresponded to a lower phrase-internal word in the other half. For each pseudo-word, two passages were constructed, one for pseudo-words in utterance-medial position and one for pseudo-words at the edge of utterances. Each passage consisted of six sentences, and each sentence was around ten syllables in length (range 9–11 syllables). The pseudo-words were paired for passage presentation, FUL/QUEU and PIS/SAU, and lexical repetition was avoided in passages for each pair. Within each sentence there were no internal intonational phrase boundaries. The target pseudo-word was preceded by a functional word half of the time and a lexical word the other half.

The passages and word lists were recorded by a female, native EP-speaker. The speaker was instructed to read each sentence aloud as if she were talking to an infant. A sound file was created using Audacity for each passage, made up of the six individually recorded sentences, with a 500 ms pause between each sentence. The speaker recorded several different spoken exemplars of each target word, all with differing intonation. A sound file was created for each target word-form, with 15 different exemplars used for each, with a 500 ms pause between each exemplar. Target words recorded in isolation, with varying intonation patterns, were used, rather than splicing examples from the stimuli, to ensure that the infants' task was not being able to simply match the isolated exemplars to previously heard word-forms encountered in passages. Instead, infants would need to extract the word-form and recognise new exemplars of this form. All the sound stimuli used, as well as lists of the passages, are available at <http://labfon.letras.ulisboa.pt/babylab/infant_word_segmentation/word_segmentation_supporting_materials.htm>.

The acoustic measurements of the stimuli can be found in Table 1. The main differences in the acoustic features of the stimuli are particularly evident due to the salient prosodic cues found in the edge condition. The word targets in the edge condition are characterised by pre-boundary lengthening in comparison with the medial condition, as well as by a greater pitch range, manifested by a pitch fall. The pitch fall in utterance-final position is due to the presence of a low edge tone (annotated as L% in Table 1, following common labelling conventions within the intonational phonology framework; Ladd, 2008) that signals the utterance-final boundary, whereas no tonal boundaries are found in utterance-medial position

Table 1. Acoustic Properties of the Stimuli; a Negative Value in Pitch Range Indicates a Pitch Fall

	Medial		Edge		
	Mean	SD	Mean	SD	*t*-test
Sentence length (ms)	2000.63	143.36	1952.88	154.91	1.11, *p* = .27
Word duration (ms)	308.79	52.49	494.5	53.6	12.13, *p* < .001
Pitch range (Hz)	−24.52	32.32	−59.58	21.83	4.4, *p* < .001
Tonal event	−		L%		−

(Frota, 2014). Overall sentence length between the two conditions was not significantly different.

Procedure

A modified version of the visual familiarisation paradigm was used (Altvater-Mackensen & Mani, 2013). Infants were seated on a caregiver's lap in front of a computer monitor, with speakers hidden behind the monitor. The experiment began with an attractive, attention-getting image on the monitor. Once the infant fixated the image for 2 consecutive seconds, a trial began with a red and black checkerboard display paired with a sound file. The sound file continued playing until the infant looked away from the screen for more than 2 seconds, or the sound file came to the end. At this point, the attractive image was once more presented on the screen until the infant fixated it for 2 consecutive seconds, when the next trial commenced. All trials followed the same pattern.

The experiment was divided into two sections, familiarisation and test. During familiarisation, infants were presented with two passages, one made up of utterances with a target word-form in utterance-medial position, and the other with a different target word-form located at the utterance edge. The two passages were presented alternatively until the infant had accumulated 25 seconds of looking time to each passage. Once this occurred, the experiment moved to the test phase, where they were presented with four sound files of isolated word-form occurrences. Two of the sound files consisted of the target word-forms heard during familiarisation, and two were new word-forms that were unfamiliar to the infants. Presentation of the test sound trials was randomised, and each word-form list was presented three times, split into blocks so that each word-form list was heard once before a word-form list was presented for the second time, and all word-form lists were presented twice before any word-form list was presented for a third time. Once all 12 test trials (4 x 3) had been presented, the experiment was over.

Stimuli presentation was controlled by the LOOK software (Meints and Woodford, 2005) and the experimenter, who was hidden from the infants' view and monitored the experiment on a monitor connected to a camera placed discreetly above the monitor in front of the infant. Infants' orientation to the screen was recorded by the experimenter pressing and holding a key on the keyboard when the infant fixated the screen, and releasing the key when the infant looked away. Both the experimenter and caregiver wore headphones playing masking music during the experiment, and the experimenter was blind to the condition the infant was assigned to and the stimuli being presented.

The four target word-forms were counterbalanced so that half of the infants heard FUL and QUEU as targets, and the other half heard PIS and SAU as targets. Additionally, within these groups the position of the target word within the utterances was also counterbalanced (i.e., for half the infants FUL was presented in the medial position, and for the other half FUL was presented at the edge, and so on).

Results

The number of familiarisation trials and the total looking time to edge and medial familiarisation passages were analysed to look for differences in the familiarisation phase. Paired *t*-tests revealed no differences in the number of trials required in the familiarisation phase (mean edge = 3.45, mean medial = 3.63; t(39) = 0.77, p = .44) or the total looking time to each of the prosodic conditions (mean edge = 28.31s, mean medial = 29.89s; t(19) = 1.68, p = .1).

Average looking times in the test phase were calculated at half-month periods (see Table 2), and curve fitting (Burchinal & Appelbaum, 1991; Fenson, Bates, Dale, Goodman, Reznick, & Thal, 2000) was applied and plotted to demonstrate the changes in looking times across the ages of participants for the three experimental conditions (edge, utterance-medial, and unfamiliar word-forms; Figure 1). As can be seen, in all three experimental conditions looking times increase with age, with a more pronounced increase for looking times to target word-forms located in utterance-medial position in comparison with utterance-edge position and unfamiliar word-forms. Looking times to targets in utterance-edge position were longer across the whole age range compared with the other two conditions (Figure 2). Looking times to targets in utterance-medial position and unfamiliar word-forms were similar at the earlier ages, with a difference becoming evident for older infants, showing longer looking times to the targets in utterance-medial position, but still less than the looking times to targets in utterance-edge position.

A repeated-measures ANOVA, with the within-participant factor of target word-form (edge, medial, unfamiliar) and age as a covariate, revealed no main effect of age (F(1,38) = 1.38, p = .25, η^2 = .04), but a significant effect of target word-form (F(2,76) = 6.83, p < .01, η^2 = .15), and an almost significant interaction between target and age (F(2,76) = 2.94, p = .059, η^2 = .07). To ascertain what was driving these effects, each experimental condition (edge, medial, unfamiliar) was analysed separately with age as a covariate. A significant effect of age was found for the medial condition (F(1,38) = 4.64, p < .05, η^2 = .11), but not for edge (F(1,38) < 1), or unfamiliar (F(1,38) = 1.94, p = .17, η^2 = .05). Finally, paired *t*-tests were carried out comparing the three experimental conditions to each other, independently of age, and the corrected p value of .02 (rather than .05) was used due to multiple *t*-tests being carried out. There were significant differences between edge and medial (t(39) = 4.72, p < .001, Cohen's *d* = 0.698), and edge and unfamiliar (t(39) = 7.81, p < .001, Cohen's *d* = 1.142), but not between medial and unfamiliar (t(39) = 2.23, p = .03, Cohen's *d* = 0.359).

These results show that looking times to target words in utterance-edge-final position and unfamiliar words were significantly different to each other, while also neither significantly differed as a function of age. This demonstrates evidence for early segmentation for word forms in utterance-edge-final position. For words in utterance-medial position, however, looking times increased significantly with age, and were significantly different to words in utterance-final position only, suggesting that segmentation abilities for word-forms in medial position are only beginning to

Table 2. Number and Average Age of the Infants in Each Half-month Interval in the Analysis

	Number	Average age (months)
4.5–5 m	2	4.77
5–5.5 m	3	5.17
5.5–6 m	2	5.84
6–6.5 m	7	6.26
6.5–7 m	4	6.76
7–7.5 m	2	7.31
7.5–8 m	1	7.90
8–8.5 m	5	8.39
8.5–9 m	3	8.89
9–9.5 m	6	9.18
9.5–10 m	2	9.84
10–10.5 m	3	10.20
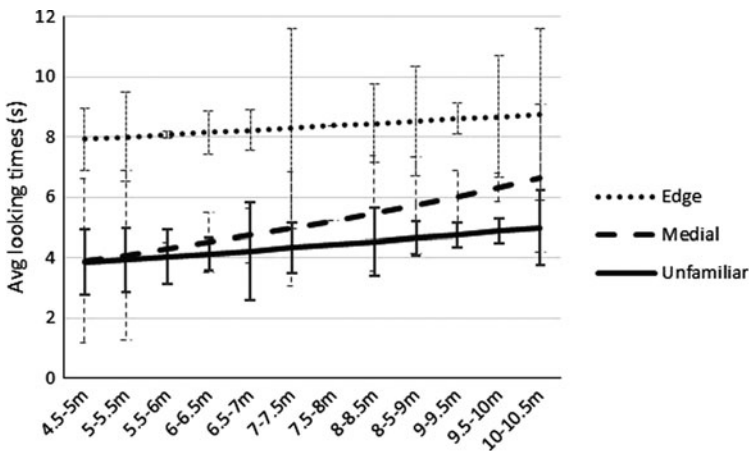
Figure 1. Average looking times (s) and standard error at half-month intervals for the three experimental conditions.

emerge, and are not as well developed as segmentation abilities in utterance-edge-final position by 10 months.

Discussion

In this study, we have begun investigating European Portuguese-learning infants' developing segmentation abilities. Using a variation of the visual familiarisation paradigm, we have shown that evidence of monosyllabic segmentation is found early
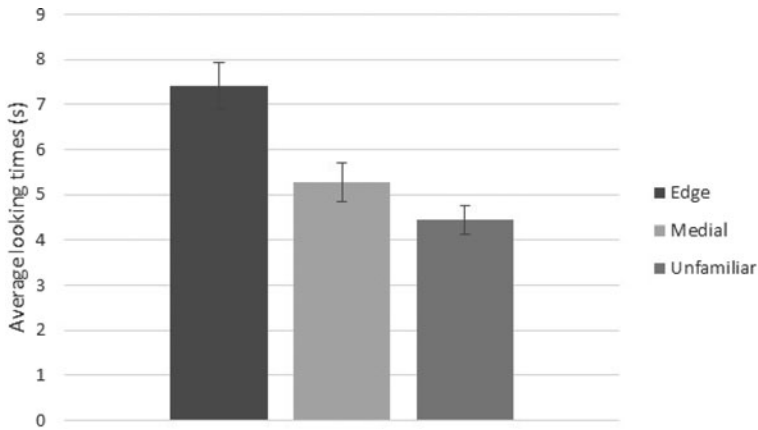
Figure 2. Average looking times (s) for the three experimental conditions, collapsed across age.

in development, but only for word-forms located at the utterance edge. We have also found evidence for the beginning of the emergence and development of segmentation abilities for word-forms found in utterance-medial position. However, the ability to segment in utterance-medial position is not as well developed as segmentation for words at the utterance edge, the prosodically salient position, even by 10 months of age, demonstrating that segmentation abilities for monosyllabic word-forms are crucially modulated by utterance-level prosody in EP.

The results are in line with previous findings for American English-learning infants showing word segmentation only at the edge at 6 months (Johnson *et al.*, 2014), but contrast with the findings reported for Spanish-, Catalan-, and French-learning infants (Bosch *et al.*, 2013; Nazzi *et al.*, 2014; Nishibayashi *et al.*, 2015), who demonstrated evidence of segmentation at 6 months regardless of utterance-level prosody. Previous findings in other stress-timed languages have found no evidence of segmentation at 6 months, using stimuli that include target words in utterance-medial position (Floccia *et al.*, 2016; Höhle & Weissenborn, 2003; Jusczyk & Aslin, 1995; Schreiner & Mani, 2017), and our results are in line with these findings, as EP-learning infants failed to segment word-forms located in utterance-medial position. Although it is possible that Spanish-, Catalan-, and French-learning infants were able to segment based solely on words located at the utterance edge (Bosch *et al.*'s 2013 and Nazzi *et al.*'s 2014 stimuli contained words both within and at the edge of sentences within a passage), it is also possible that, as learners of syllable-timed languages, these infants are focusing on the syllable as the major rhythmic unit for segmentation, as suggested by Bosch and colleagues (2013) to explain the success of Spanish/Catalan infants at 6 months, in contrast with findings from other languages, such as English (Jusczyk & Aslin, 1995). In Spanish, Catalan, or French, monosyllabic words match the rhythmic unit for segmentation, i.e., the syllable. Infants learning syllable-timed languages may thus have an advantage over infants learning languages with other rhythmic properties. EP-learning infants, unlike Spanish, Catalan, or French infants, may not be able to rely on this rhythmic segmentation strategy, given the mixed rhythm of the language that combines syllable- and stress-timed properties. The failure of EP-learning infants, as well as

infants learning stress-timed languages (Floccia *et al.*, 2016; Höhle & Weissenborn, 2003; Jusczyk & Aslin, 1995, Schreiner & Mani, 2017), to segment target monosyllabic words at early ages regardless of utterance-level prosody further suggests that there are developmental differences in the emergence of segmentation abilities between languages, depending on their rhythmic properties. Therefore, the current findings provide support for the Rhythmic Segmentation Hypothesis, while suggesting that more detailed attention needs to be given to the rhythmic properties of the language(s) infants are exposed to, even in the case of languages traditionally placed within the same rhythmic groupings. Importantly, early segmentation abilities in other languages that do not easily conform to the traditional stress-timed and syllable-timed groupings due to the mixed nature of their rhythmic properties (like Bulgarian or Turkish; Dimitrova, 1997; Nespor *et al.*, 2011), need to be studied.

By 10 months, EP-learning infants' segmentation abilities for words found in utterance-medial positions are beginning to develop. This is in line with studies showing that segmentation abilities generally emerge from around 7.5 months of age (e.g., Höhle & Weissenborn, 2003; Jusczyk & Aslin, 1995), or later in the first year (Floccia *et al.*, 2016; Schreiner & Mani, 2017). However, EP-learning infants' ability to segment word forms in utterance-medial position is not fully developed by 10 months, as an advantage was evident for segmenting word-forms at the edge of utterances. Words found at the edge of utterances are easier to segment than those within utterances, which are not aligned with a higher-level prosodic boundary. The prosodic cues found at utterance-edges appear to offer the most salient cues for infants to utilise when beginning to learn to process continuous speech, as early as 4–5 months. This advantage, that continues to be demonstrated by 10-month-olds segmenting word-forms found at utterance edges, may be due to the presence of robust prosodic cues, which are the result of the combination found in EP of salient cues to the highest level of the prosodic structure, and cues to the prosodic word level. These findings give clear support to the Edge Hypothesis, in a language other than American English.

The materials used in the current study provided the strongest combination of prosodic boundary cues offered by the language, namely pitch change, pre-boundary lengthening, and pause. It is not known whether EP-learning infants need all the three cues present for segmentation, or whether their sensitivity to prosodic boundary cues involves a language-particular weighting of these cues that undergoes a developmental change, as reported for American English, Dutch, and German (Johnson & Seidl, 2008; Seidl, 2007; Seidl & Cristià, 2008; Wellmann *et al.*, 2012). Adults' performance suggests that pitch change and pre-boundary lengthening are the main prosodic boundary markings in EP, with the pause being an optional cue (Frota, 2000; Severino, 2017). However, we do not know whether EP-learning infants were able to exploit non-pause prosodic cues, and how the different cues were weighed. Yet, when pitch change, pre-boundary lengthening and pause were all present, evidence for segmentation was found. Further work is needed to identify what prosodic cues infants are, or are not, utilising to begin segmenting continuous speech, and whether and when a language-specific prosodic cue weighting develops.

This advantage for the extraction of word-forms at the utterance-edge versus utterance-medial position may affect how infants begin learning to segment bisyllabic, and larger, word-forms (assuming that clear infant-directed speech properties that facilitate segmentation are present; Floccia *et al.*, 2016; Schreiner & Mani, 2017). Prosodic boundary cues at utterance edges allow infants to extract

monosyllabic word-forms; however, an additional problem is presented to the infants with bisyllabic words, as the infant needs to identify what constitutes a complete word, rather than part of a word (Nazzi *et al.*, 2006; Nishibayashi *et al.*, 2015). The Rhythmic Segmentation Hypothesis does not allow a clear prediction as to how EP learning-infants may begin segmenting bisyllabic word-forms, given the mixed rhythmic properties of the language. It is possible that infants' early segmentation abilities for monosyllabic word-forms at the edge of utterances are guided by recognising only the syllable that is aligned with the edge of the utterance. This may suggest that infants may begin to develop segmentation abilities by focusing on, for example, only the syllable that is aligned with the edge of utterances, or of utterance-internal higher-level prosodic boundaries, such as utterance-internal intonational phrases. If this is the case, bisyllabic segmentation may emerge later in development, as infants would initially segment part of the bisyllabic word, namely the syllable at the prosodic edge, rather than the whole word. Another possibility would be that whole word-form segmentation is first found at the prosodic edge, rather than in utterance-medial position, triggered by the combination of salient cues to both higher-level prosodic edges and the prosodic word level found in EP. The present study thus sets the stage for future research examining bisyllabic word segmentation in a language with mixed prosodic properties, such as European Portuguese.

In conclusion, the present findings demonstrated that EP-learning infants are sensitive to the salient prosodic cues found at the edge of utterances when beginning to segment continuous speech, providing new evidence for the edge factor in emerging word segmentation abilities at an early age and in a language with a mixed prosodic profile, such as European Portuguese, not previously described in the literature. Results from this research have also shown that monosyllabic word segmentation not at utterance-edge position develops later and is not yet robust by 10 months of age, giving new evidence for the rhythmic segmentation hypothesis from a language with mixed rhythm. Taken together, these findings show that utterance-level prosody constrains the emergence and development of early segmentation in EP-learning infants throughout the first year. They offer relevant information for future cross-linguistic studies on early segmentation abilities in languages with different rhythmic properties, including mixed languages, which also take into account higher-level prosodic structure.

References

Altvater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science*, *16*(6), 980–90.

Arvaniti, A., & Rodriquez, T. (2013). The role of rhythm class, speaking rate, and F0 in language discrimination. *Laboratory Phonology*, *4*(1), 7–38.

Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology*, *4*, 106.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: methods and their assumptions. *Child Development*, *62*, 23–43.

Byrd, D., Krivokapic, J., & Lee, S. (2006). How far, how long: on the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, *120*, 1589–99.

Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience*, *30*(1/2), 1–31.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133–42.

Dimitrova, S. (1997). Bulgarian speech rhythm: stress-timed or syllable-timed? *Journal of the International Phonetic Association*, *27*(1/2), 27–33.

Fenson, L., Bates, E., Dale, P., Goodman, J.Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: don't shoot the messenger. *Child Development*, *71*(2), 323–8.

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9.

Frota, S. (2000). *Prosody and focus in European Portuguese: phonological phrasing and intonation.* New York: Garland Publishing.

Frota, S. (2012). Prosodic structure, constituents and their implementation. In A. C. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 255–65). Oxford University Press.

Frota, S. (2014). The intonational phonology of European Portuguese. In Sun-Ah Jun (Ed.), *Prosodic typology II: the phonology of intonation and phrasing* (pp. 6–42). Oxford University Press.

Frota, S., & Prieto, P. (2015). Intonation in Romance: systemic similarities and differences. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 392–418). Oxford University Press.

Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus*, *13*, 247–72.

Frota, S., Vigário, M., & Martins, F. (2002). Language discrimination and rhythm class: evidence from Portuguese. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody* 2002 (pp. 315–18), Aix en Provence: Laboratoire Parole et Langage.

Frota, S., Vigário, M., Martins, F., & Cruz, C. (2010). *Frequency patterns of phonological objects in Portuguese*. Lisbon: Laboratório de Fonética (CLUL), Faculdade de Letras da Universidade de Lisboa. Retrieved from: <http://frepop.letras.ulisboa.pt/> ISLRN: 064-984-771-090-2.

Gout, A. (2001). *Etapes precoces de l'acquisition du lexique* [Early steps in lexical acquisition] (Dissertation). Ecole des Hautes Etudes en Sciences Sociales, Paris.

Goyet, L., Nisibayashi, L.-L., & Nazzi, T. (2013). Early syllable segmentation of fluent speech by infants acquiring French. *PLoS ONE*, *8*(11), e79646.

Gussenhoven, C. (2015). Suprasegmentals. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences, 2nd ed., vol. 23* (pp. 714–21). Oxford: Elsevier.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, *6*(2), 122–7.

Höhle, B., & Weissenborn, J. (2005). *Word segmentation in German-learning infants*. Paper presented at the international workshop Early Word Segmentation: a Crosslinguistic Approach Taking Advantage of Europe's Linguistic Diversity, 25–26 February, Paris, France.

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: a crosslinguistic perspective. *Infancy*, *13*(5), 440–55.

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PLoS ONE*, *9*(1), e83546.

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

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

Keating, P., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation* (Papers in Laboratory Phonology 6) (pp. 143–61). Cambridge University Press.

Keij, B. (2017). *Rhythm & cues: rhythmic structure and segmentation in early language acquisition* (Doctoral dissertation). Utrecht: LOT.

Kooijman, V., Hagoort, P., & Cutler, A. (2009). Prosodic structure in early word segmentation: ERP evidence from Dutch ten-month-olds. *Infancy*, *14*(6), 591–612.

Ladd, D. R. (2008). *Intonational phonology*, 2nd ed. Cambridge University Press.

Mateus, M. H. M., & Andrade, E. (2000). *The phonology of Portuguese*. Oxford University Press.

Meints, K., & Woodford, A. (2005). Lincoln infant lab package 1.0: a new programme package for IPL preferential listening, habituation and eyetracking. Retrieved from <https://www.lincoln.ac.uk/home/psychology/research/lincolnbabylab/>.

Mersad, K., Goyet, L., & Nazzi, T. (2010). Cross-linguistic differences in early word form segmentation: a rhythmic-based account. *Journal of Portuguese Linguistics*, 9/10, 37–65.

Mo, Y., & Cole, J. (2010). Perception of prosodic boundaries in spontaneous speech with and without silent pause. *Journal of the Acoustical Society of America*, 127, 1956.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, 54, 283–99.

Nazzi, T., Mersad, K., Sundara, M., Iakimova, G., & Polka, L. (2014). Early word segmentation in infants acquiring Parisian French: task-dependent and dialect-specific aspects. *Journal of Child Language*, 41(3), 600–33.

Nazzi, T., Paterson, S., & Karmiloff.Smith, A. (2003). Early word segmentation by infants and toddlers with Williams Syndrome. *Infancy*, 4(2), 251–71.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. Roca (Ed.), *Logical issues in language acquisition* (pp. 157–75). Dordrecht: Foris.

Nespor, M., Shukla, M., & Mehler, J. (2011) Stress-timed vs. syllable timed languages. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1147–59). Malden, MA: Wiley-Blackwell.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology*, 2nd ed. Berlin and New York: Walter de Gruyter.

Newman, R., Ratner, N. B., Jusczyk, A. M., & Jusczyk, P. W. (2006). Infants' ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, 42(4), 643–55.

Nishibayashi, L.-L., Goyet, L., & Nazzi, T. (2015). Early speech segmentation in French-learning infants: monosyllabic words versus embedded syllables. *Language and Speech*, 58 (3), 334–50.

Polka, L., Orena, A. J., Sundara, M., & Worrall, J. (2016). Segmenting words from fluent speech in infancy: challenges and opportunities in the bilingual context. *Developmental Science*, 20(1), e12419.

Polka, L., & Sundara, M. (2012). Word segmentation in monolingual infants acquiring Canadian English and Canadian French: native language, cross-dialect and cross-language comparisons. *Infancy*, 17(2), 198–232.

Rathcke, T., & Smith, R. (2015). Rhythm class perception by expert phoneticians. In The Scottish Consortium for ICPhS 2015 (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow.

Schreiner, M. S., & Mani, N. (2017). Listen up! Developmental differences in the impact of IDS on speech segmentation. *Cognition*, 160, 98–102.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57, 24–48.

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, 11(4), 596–606.

Seidl, A., & Johnson, E. K. (2006). Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*, 9(6), 565–73.

Severino, C. (2017). *Perception of phrasal prosody in the acquisition of European Portuguese* (Doctoral dissertation). Faculdade de Letras da Universidade de Lisboa.

Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *JASA Express Letters*, 6, EL61–EL67.

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, 14(6), 654–66.

Singh, L., Reznick, S., & Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Developmental Science*, 15(4), 482–95.

Sundara, M., Demuth, K., & Kuhl, P. K. (2011). Sentence-position effects on children's perception and production of English third person singular -s. *Journal of Speech, Language, and Hearing Research*, 54, 55–71.

Vigário, M. (2003). *The prosodic word in European Portuguese* (Interface Explorations Series, 6). Berlin & New York: Mouton de Gruyter.

Vigário, M., Freitas, M. J., & Frota, S. (2006). Grammar and frequency effects in the acquisition of the Prosodic Word in European Portuguese. *Language and Speech 49*(2), 175–203.

Vigário, M., Frota, S., & Freitas, M. J. (2003). From signal to grammar: rhythm and the acquisition of syllable structure. In *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 809–21). Sommerville, MA: Cascadila Press.

Vigário, M., Frota, S., & Martins, F. (2011). Para uma caracterização da distinção entre palavras prosódicas e clíticos com base em dados de frequência. In *XXVI Encontro Nacional da Associação Portuguesa de Linguística: Textos Seleccionados* (pp. 589–603). Lisbon: Associação Portuguesa de Linguística.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Warterburger, I., & Höhle, B. (2012). How each prosodic cue matters: evidence from German infants. *Frontiers in Psychology*, 3, 580.

White, L., Delle Luche, C., & Floccia, C. (2016). Five-month-old infants' discrimination of unfamiliar languages does not accord with 'rhythm class'. *Speech Prosody*, *2016–116*, 567–71.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, 66(4), 665–79.