Beating the bounds: Localized timing cues to word segmentation

Laurence White, Sven L. Mattys, Linda Stefansdottir, and Victoria Jones

Citation: The Journal of the Acoustical Society of America **138**, 1214 (2015); doi: 10.1121/1.4927409
View online: https://doi.org/10.1121/1.4927409
View Table of Contents: https://asa.scitation.org/toc/jas/138/2
Published by the Acoustical Society of America

Beating the bounds: Localized timing cues to word segmentation

Laurence White,[1,a)] Sven L. Mattys,[2] Linda Stefansdottir,[3] and Victoria Jones[1]
[1]*School of Psychology, Plymouth University, Drake Circus, Plymouth PL4 8AA, United Kingdom*
[2]*Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom*
[3]*School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom*

(Received 25 February 2014; revised 10 June 2015; accepted 9 July 2015; published online 28 August 2015)

Prosody facilitates perceptual segmentation of the speech stream into a sequence of words and phrases. With regard to speech timing, vowel lengthening is well established as a cue to an upcoming boundary, but listeners' exploitation of consonant lengthening for segmentation has not been systematically tested in the absence of other boundary cues. In a series of artificial language learning experiments, the impact of durational variation in consonants and vowels on listeners' extraction of novel trisyllables was examined. Language streams with systematic lengthening of word-initial consonants were better recalled than both control streams without localized lengthening and streams where word-initial syllable lengthening was confined to the vocalic rhyme. Furthermore, where vowel-consonant sequences were lengthened word-medially, listeners failed to learn the languages effectively. Thus the structural interpretation of lengthening effects depends upon their localization, in this case, a distinction between lengthening of the onset consonant and the vocalic syllable rhyme. This functional division is considered in terms of speech-rate-sensitive predictive mechanisms and listeners' expectations regarding the occurrence of syllable perceptual centres.
[http://dx.doi.org/10.1121/1.4927409]

[CGC] Pages: 1214–1220

I. INTRODUCTION

Variations in suprasegmental dimensions—pitch, duration, loudness—are consistently associated with speech structure at prosodic heads and edges (e.g., Beckman, 1992). First the heads of prosodic domains—stressed syllables and accented words—are more prominent through a combination of greater duration, greater loudness, and pitch excursion, although the relative contribution of these dimensions is language-specific and some languages may lack head marking altogether (Beckman, 1992). Second, boundaries between words and between phrases are associated with intonational and durational variation. Boundary-adjacent intonational contours vary between languages (e.g., Ladd, 1996), but upcoming boundaries may be universally associated with segmental lengthening (Beckman, 1992). Indeed the slowing of articulation as boundaries approach has been associated with non-linguistic principles, such as deceleration at the end of motor sequences (e.g., Fowler, 1990; Tyler and Cutler, 2009), although the language-specific localization of final lengthening effects suggests that any underlying non-linguistic tendencies have become formalized into the phonology (e.g., Turk and Shattuck-Hufnagel, 2007; White, 2002, 2014).

It is well established that listeners use suprasegmental variation to segment speech into words and phrases (e.g., Christophe *et al.*, 2004; Price *et al.*, 1991). Considering specifically speech timing, lengthened vowels are interpreted as word-final in artificial language streams (e.g., Saffran *et al.*, 1996b). Similarly with natural language stimuli, longer stressed syllables are more likely to be interpreted as monosyllabic words rather than the start of disyllables (e.g., *ham* vs *hamster*; Salverda *et al.*, 2003; see also Davis *et al.*, 2002).

Such results are broadly in line with the iambic-trochaic law (Hayes, 1995), which proposes that the interpretation of prosodic salience depends on its phonetic realization: in particular, sounds made salient through greater loudness are perceived as sequence-initial, whereas sounds made salient through lengthening are perceived as sequence-final. Support for the iambic-trochaic law was found with both native English- and French-speaking listeners and with speech and non-speech sounds (Hay and Diehl, 2007). Similarly, Italian listeners' recall of disyllabic sequences was better when final syllables had longer vowels compared to when the vowels of both syllables had similar durations or when initial syllables had longer vowels (Bion *et al.*, 2011). Furthermore, for English listeners, lengthening of vocalic nuclei in word-final syllables promoted segmentation of artificial language streams but not in word-initial syllables (Saffran *et al.*, 1996b), a finding that also was obtained for Dutch and French listeners, despite cross-linguistic differences in the interpretation of word-initial and word-final pitch cues (Tyler and Cutler, 2009). The ineffectiveness of vowel lengthening in initial syllables as a boundary cue apparently runs counter to the trend in English for word-initial stress and consequent metrical segmentation preferences, but Cutler (1986) noted that English stress is most strongly cued by vowel quality; furthermore, Mattys, White, and Melhorn

[a)]Electronic mail: laurence.white@plymouth.ac.uk

(2005) showed that metrical segmentation is not exploited where more reliable cues—e.g., lexical, syntactic, segmental-acoustic boundary-related information—are available.

Most prosodic timing studies have focused on the impact of vowel lengthening on segmentation. However, lengthening of consonants in word-initial position is consistently observed in several studied languages. For English, syllable onset consonants are substantially longer word-initially than word-medially (Oller, 1973), an effect also observed in French, Korean, and Taiwanese (Keating *et al.*, 2003). Whilst multiple consonants within the onset may be lengthened word-initially, the durational effect does not extend to the vowel nucleus of that syllable, at least in English (Oller, 1973; White, 2002).

Some studies of word segmentation have considered the impact of consonant lengthening in conjunction with other prosodic and segmental cues (e.g., Gout *et al.*, 2004; Gow and Gordon, 1995; Quené, 1992). For example, word-initial consonant lengthening, together with word-final vowel lengthening and other naturally occurring cues to prosodic boundaries, affects the interpretation of ambiguous sequences such as *pay per* vs *paper* in English-learning infants as young as ten months (Gout *et al.*, 2004) as well as French adults given parallel stimuli in their native language (Christophe *et al.*, 2004). In Dutch, listeners' interpretation of segmentally ambiguous sequences like *die pin* vs *diep in* is affected by the duration of the pivotal consonant, which tends to be interpreted as word-initial when long (Quené, 1992; see also Shatzman and McQueen, 2006). Consonant duration also affects Italian listeners' word segmentation in parallel with their identification of geminates vs singletons (Tagliapietra and McQueen, 2010), whilst French listeners interpret longer consonants as more likely to be word-initial than in liaison context (e.g., *dernier rognon* vs *dernier oignon,* Spinelli *et al.*, 2003).

The preceding studies suggest that lengthened consonants may be interpreted as word-initial by listeners. Importantly, however, all used natural speech—sometimes resynthesized to manipulate segment durations—with multiple potential cues to word boundaries. Thus segmental cues, such as boundary-related allophonic variations, and other prosodic cues, including lengthening of word-final vowels, were also available to listeners, precluding strict interpretation of segmentation as being driven by initial consonant lengthening. Furthermore, participants' awareness of implicit contrasts between two interpretations of near-homophonous sequences (e.g., *pay per* vs *paper*) may have modulated their use of segmentation cues relative to when there was only one lexical solution available.

We used an artificial language learning paradigm to focus specifically on lengthening of consonants and lengthening of vowels in the absence of any other cues. Listeners have consistently been shown to be able to learn and subsequently recall novel words from a nonsense speech stream when the syllable-to-syllable transitional probabilities within words are higher than those between words (e.g., Saffran *et al.*, 1996a; Saffran *et al.*, 1996b). Exploiting such a paradigm, we obviate the need to use near-homophonous sequences from natural languages and eliminate the presence of other potential cues to word boundaries. This allows us to focus precisely on the key question: does longer duration make consonants more likely to be interpreted as word-initial? After directly examining this question in Experiment 1, we compare the effects of vowel vs consonant lengthening in the word-initial syllable in Experiment 2, and test the effectiveness for segmentation of lengthened vowel + consonant sequences in Experiment 3.

II. EXPERIMENT 1

A. Method

Localized manipulations of segment duration in an artificial language were used to assess the impact of consonantal lengthening on segmentation, and thereby on subsequent recall, of words in an artificial language. We predicted that words should be better recalled when word-initial consonants were lengthened during language exposure relative to when all consonants had the same duration or when word-medial consonants were lengthened.

1. Participants

We tested 120 native British English speakers with no reported speech or hearing problems. All received a small honorarium or course credit for their participation. These participant characteristics were equivalent in all experiments. Participants were randomly allocated to the three duration conditions (40 in each condition).

2. Materials

To counterbalance any idiosyncrasies associated with the selected words, we prepared two artificial language streams (Table I), following those used in Saffran *et al.*, (1996a, their Experiment 2). Each stream comprised four trisyllabic words (C1V1-C2V2-C3V3).

The *en1* male British English voice in the diphone synthesizer MBROLA (Dutoit *et al.*, 1996) was used to generate 6-min streams containing these words in pseudo-random sequence, yielding streams 1 and 2. The same word never occurred twice in immediate succession. Due to an idiosyncrasy in the synthesis of /bu/ sequences, which had a marked nasal quality, we substituted these with /nu/, generating *tinudo* in stream 1 and *nudopa* in stream 2 (these were *tibudo* and *budopa* in Saffran *et al.*, 1996a). Otherwise the words were as for Experiment 2 in Saffran *et al.* In particular, because each syllable only occurred once within the four words of the language stream, the syllable sequence within words was entirely predictable, and so the within-word between-syllable transitional probability was always 1. In contrast, after the final syllable of a word, there were three possible syllables that could immediately follow, i.e., the

TABLE I. Words used in the two artificial language streams.

Stream 1	daropi	golatu	pabiku	tinudo
Stream 2	bikuti	nudopa	pigola	tudaro

initial syllables of the three other words, hence a between-word probability of 0.33. Some other artificial language experiments (e.g., Saffran *et al.*, 1996b) have varied the transitional probability within words, by repeating syllables, but here we focus on prosodic cues rather than statistical learning and hence preferred to maintain a consistent transitional probability contrast within vs between words.

Fundamental frequency was a constant 120 Hz, and the streams were faded in and out with five-second ramps.

To maintain a consistent overall speech rate (and hence information rate) between conditions, total trisyllabic word duration was kept constant at 720 ms, whilst the duration of individual segments was manipulated to generate three "lengthening" conditions.

Flat: All segments—vowels and consonants—were 120 ms

C1: The onset consonant of the first syllable of each word (*pabiku* etc.) was 170 vs 110 ms for all other segments.

C2: The onset consonant of the second syllable of each word (*pabiku* etc.) was 170 vs 110 ms for all other segments.

The magnitude of lengthening was a compromise between values used for vowel lengthening in previous studies (e.g., 100 ms in Saffran *et al.*, 1996b) and the smaller magnitude of typically observed phrase-medial word-initial lengthening. (NB: The lengthening manipulation, as implemented in the MBROLA synthesizer, affects both the closure and aspiration phases of voiceless stops.)

In the test phase, following exposure to the stream, isolated words and foils were played to participants. Foils were part-words derived from the end of one word and start of another (e.g., stream 1: *bikuti* from *pabiku tinudo*) and non-words, syllable sequences that never occurred in the language (e.g., *tipala*). Each word was paired with three different foils, two part-words and one non-word, with all pairs presented twice, once in each order (word-foil vs foil-word). The words in stream 1 were part-words in stream 2 and vice versa. Words and foils for the test phase were synthesized with all segments 120 ms in all three conditions. Within each duration condition, 20 participants were allocated to stream 1 and 20 to stream 2.

3. Procedure

Participants were told they would hear an artificial language through headphones for 6 min and that their task was to listen and try to discover the words in the language. After the exposure phase, they were given instructions for the test phase. In the test phase, they heard 24 pairs of trisyllabic strings, based on three word-foil pairs for each word and two orders of word-foil presentation (see preceding text). The two trisyllabic strings were separated by 500 ms. For each pair, participants were asked to press the left shift key on a computer keyboard if the artificial language word was the first string of the pair, and the right shift key if it was the second string. This two-alternative forced-choice test phase is in line with common practice for adult artificial language learning experiments (see Saffran *et al.*, 1996b and subsequent studies). We used a shorter test phase than in experiments where the focus is on statistical learning and all words are typically paired with all foils (e.g., Saffran *et al.*, 1996b). In our procedure, word-foil exposure was matched between timing conditions with participants presented with three different foils for each word.

B. Results and discussion

All analyses were carried out on the raw response data—"correct" or "incorrect"—using mixed-effects logistic regression models, including the random factors of subjects, streams, and items (*lmer* package in *R*, Baayen *et al.*, 2008). Items—nested under the two artificial language streams—were the 24 two-alternative-forced choice trials, taken separately for the two orders of presentation (word/foil; foil/word). The effect of the timing manipulations on performance was established by comparing models that included only the random structure to models that also included a fixed factor of lengthening condition(s), using log-likelihood χ^2 tests.

Mean correct responses by lengthening condition are shown in Fig. 1 (which also illustrates results for Experiments 2 and 3). Above-chance performance was found in all three timing conditions: flat: 67%, $\beta = 0.82$, SE = 0.22, $z = 3.79$, $p < 0.001$; C1: 73%, $\beta = 1.36$, SE = 0.27, $z = 5.05$, $p < 0.001$; C2: 63%, $\beta = 0.61$, SE = 0.23, $z = 2.61$, $p = 0.009$. With regard to our key question, comparison of logistic regression models with and without the fixed factor of lengthening (flat vs C1 vs C2), in addition to the common random structure, showed a main effect of lengthening, $\beta = 0.29$, SE = 0.09, $\chi^2(2) = 11.20$, $p < 0.001$. Lengthening of the consonant in the first syllable (C1) improved performance both compared to lengthening of the consonant in the second syllable (C2), $\beta = 0.61$, SE = 0.19, $\chi^2(1) = 10.02$, $p = 0.002$, and compared to the flat condition, $\beta = 0.42$, SE = 0.20, $\chi^2(1) = 4.20$, $p = 0.040$. There was no difference between C2 vs flat, $\beta = 0.20$, SE = 0.13, $\chi^2(1) = 2.14$, $p = 0.144$. These results indicate that segmentation of the artificial language was promoted by localized lengthening of the word-initial consonant. Thus consonantal lengthening appeared to cue listeners to the presence of an immediately preceding boundary. Given that result, it might also be expected to find deterioration in segmentation performance where the lengthened consonant was word-internal. However, the numerical drop in recognition from the flat to the C2 condition was not statistically robust, suggesting that this timing cue alone was not sufficient to

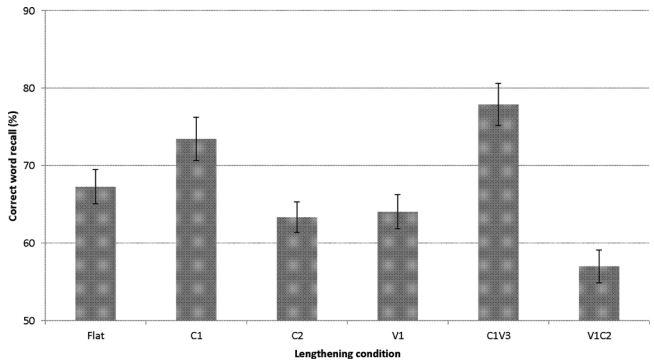

FIG. 1. Mean correct responses and standard errors. Experiment 1: Flat, C1, C2; Experiment 2: V1; Experiment 3: C1V3, V1C2. Chance level: 50%.

offset the effect of frequency of exposure to statistically-defined words. We explore the resolution of prosody vs statistics conflicts further in Experiment 3.

Lengthening of a vowel in a similar artificial language stream has been shown to act as a cue to a following boundary (Saffran *et al.*, 1996b). This suggests a functional difference in listeners' interpretation of lengthening in vocalic syllable rhymes and consonantal onsets, as we predicted initially. An alternative hypothesis is that syllables that are longer—whether through greater vowel or consonant duration—tend to be perceived as word-edges, either initial or final. This view is not supported by findings that vowel lengthening in word-initial syllables failed to facilitate segmentation relative to no lengthening (Saffran *et al.*, 1996b; Tyler and Cutler, 2009). However, in the Saffran *et al.* experiment, for which the design of the materials more closely resembles our own, vowels in initial syllables were only lengthened in half of the six artificial words. To confidently assert our interpretation that onset consonant lengthening, in contrast with vowel lengthening, is a cue to a preceding boundary, we test—with our materials and methodology—the segmentation effect of lengthening the first syllable vowel in every trisyllabic word.

III. EXPERIMENT 2

A. Method

There were 40 new participants, and the procedure was as for Experiment 1. However, in Experiment 2, participants heard the artificial language streams with the first vowel of each word lengthened: thus the underlined vowel in *pabiku* etc., was 170 vs 110 ms for all other segments, both consonants and vowels. This V1 condition was implemented for both artificial language streams with 20 participants randomly assigned to one of the two 6-min streams. The structure of the two-alternative forced-choice test phase was as for Experiment 1.

B. Results and discussion

Mean correct word recognition in the V1 condition was above chance: 64%, $\beta = 0.67$, SE = 0.17, $z = 3.87$, $p = 0.001$. To test the hypothesis regarding the localization of durational segmentation cues, the important comparisons were with the flat and C1 conditions in Experiment 1 (Fig. 1). There was no difference in recognition between the flat and V1 conditions, $\beta = 0.16$, SE = 0.15, $\chi^2(1) = 1.07$, $p = 0.30$, replicating previous findings that lengthening of the vowel in a word-initial syllable does not serve as a cue to a preceding boundary for English listeners despite the prevalence of word-initial stress in English (Saffran *et al.*, 1996b; Tyler and Cutler, 2009).

Performance on the C1 condition was reliably better than the V1 condition, $\beta = 0.59$, SE = 0.21, $\chi^2(1) = 7.52$, $p < 0.006$. This supports the hypothesis that localization of lengthening is important for segmentation: a lengthened consonant cues a preceding boundary; a lengthened vowel cues a following boundary.

In Experiment 3, to explore the power of such cues further, we tested the efficacy of vowel and consonant lengthening in combination. In particular, we examined whether a lengthened vowel immediately followed by a lengthened consonant was a strong cue to an intervening boundary. In one condition (C1V3, see following text), the juncture between lengthened vowels and lengthened consonants was congruent with word boundaries as defined by syllable transitional probabilities, and so the two sources of segmentation information—statistics and prosody—were mutually reinforcing. In the other condition (V1C2, see following text), the lengthened vowel/lengthened consonant sequences occurred in the middle of words defined by transitional probabilities, and so statistics and prosody were in conflict. We expected the latter condition to be detrimental to segmentation in line with previous studies exploring interactions between prosody and statistical word boundary information (e.g., Johnson and Seidl, 2009; Shukla *et al.*, 2007).

IV. EXPERIMENT 3

A. Method

The procedure was equivalent to Experiment 1, with 40 new participants in each of two conditions. In condition C1V3, the first consonant and the final vowel of each word (e.g., *pabiku*) were each 160 ms, vs 100 ms for all other segments. In condition V1C2, the vowel of the first syllable and the consonant of the second syllable (e.g., *pabiku*) were each 160 ms, vs 100 ms for all other segments. This was effectively a composite of the V1 and C2 conditions. Note that the lengthened segments were 160 ms and the others 100 ms in contrast with 170 and 110 ms in the other experiments: this was to preserve total word duration at 720 ms in all conditions across all three experiments.

B. Results and discussion

As shown in Fig. 1, performance was reliably above chance in the C1V3 condition, 78%, $\beta = 1.54$, SE = 0.20, $z = 7.83$, $p < 0.001$, where lengthening in the onset consonant and the vocalic rhyme were both congruent with the statistical word boundaries. However, it was not above chance in the V1C2 condition, 57%, $\beta = 0.31$, SE = 0.21, $z = 1.44$, $p = 0.15$, where lengthened vocalic rhyme and onset consonant sequences implied boundaries *within* statistically defined words. Accordingly, performance was significantly better in the C1V3 condition than the V1C2 condition, $\beta = 1.16$, SE = 0.19, $\chi^2(1) = 32.18$, $p < 0.0001$.

Comparison with the earlier experiments showed that performance on C1V3 was no better than on C1, $\beta = 0.29$, SE = 0.27, $\chi^2(1) = 1.11$, $p = 0.29$. This may be due to intrinsic performance limitations on the language learning task given the memory component combined with the repeated exposure to words and foils during the 24 two-alternative forced-choice test trials. However, performance on C1V3 was better than on all other conditions ($p < 0.001$ for all comparisons).

Performance in the V1C2 condition was worse than either the V1 condition alone, $\beta = 0.29$, SE = 0.12,

$\chi^2(1) = 0.58$, $p = 0.02$, or the C2 condition alone $\beta = 0.34$, SE = 0.14, $\chi^2(1) = 5{,}23$, $p = 0.02$. Thus a combination of timing cues suggesting prosodic boundaries in the middle of words was more effective than either word-medial vocalic rhyme or onset consonant lengthening alone at inhibiting encoding of statistically defined words. This accords with the finding that statistically defined trisyllables that straddle intonationally defined boundaries in artificial language streams are not well recognized (Shukla *et al.*, 2007). Participants' recall on V1C2 was also worse than on the flat or C1 conditions ($p < 0.0001$ for both comparisons).

V. CONCLUSION

The three experiments show, in combination with previous findings, that segmental lengthening can serve as a cue to both preceding and following prosodic boundaries according to its localization. As shown in Fig. 1, word recognition performance was best in the two conditions (C1 and C1V3) where the onset consonant of the first syllable in each word was lengthened. Thus even in the absence of other segmental and prosodic cues, listeners interpret lengthened onset consonants to indicate the start of a new word, providing a segmentation boost when the timing cues were congruent with statistically defined boundaries. The sequence of lengthened vocalic rhyme and lengthened onset consonant was, in contrast, effective at inhibiting segmentation when it occurred *within* a statistically defined word and was indeed more effective than either word-medial vowel or consonant lengthening alone. The latter results show the power of combined vocalic rhyme plus onset consonant lengthening cues for defining an intervening boundary, reinforcing previous findings regarding the perceptual significance of preboundary vowel lengthening (e.g., Price *et al.*, 1991; Saffran *et al.*, 1996b).

The trade-off between diverse sources of segmentation information was examined by Mattys *et al.* (2005), who found that acoustic-phonetic and segmental cues were not as heavily weighted as cues derived from lexical, semantic, and syntactic knowledge. It is, of course, a non-trivial task to fully characterize the range of relevant sources of knowledge that listeners bring to bear in word segmentation even in tightly controlled artificial language learning experiments. Although only syllable transitional probabilities and prosodic cues (specifically timing) were explicitly manipulated in the experiments reported here, listeners are very likely to be additionally influenced by preconceptions about what constitutes a well-formed word, together with incidental partial resemblances to existing vocabulary items. Nonetheless the results of these experiments are instructive on the nature of the specific interaction between prosody and statistics for word segmentation. Shukla *et al.* (2007) suggested that lexical candidates were removed from consideration where prosodic cues—in their case, final lengthening combined with a phrase-final intonational contour—disagreed with the segmentation suggested by statistics. The need for statistics to concur with prosody was reinforced in a study of 11-month-old infants (Johnson and Seidl, 2009), which found that statistical learning was disrupted by non-word-initial lexical stress placement. Similarly, in the current experiments, the worst performance was in condition V1C2, where a lengthened vocalic rhyme was followed by a lengthened onset consonant within the same word. Here the combined lengthening cues indicated a boundary that was incongruent with transitional probabilities; furthermore, these timing cues were sufficiently salient to prevent listeners from effectively learning the language with word recognition no better than chance. In contrast, within the word, either vocalic rhyme or onset consonant lengthening alone was not sufficient to cause a reliable deterioration in performance, suggesting that there is a threshold level of salience for prosodic cues to overturn the perception of statistically defined boundaries.

The power of multiple lengthening cues may partially derive from their mutual congruency: in natural speech, preboundary vowel lengthening and post-boundary consonant lengthening typically co-occur. The current results clear show the importance of *localization* of lengthening within the syllable, but further work is needed to ascertain what *magnitude* of timing effects are readily interpretable by listeners. The typical magnitude of word-initial consonant lengthening in phrase-medial context in English may be around 20% in words without phrasal accent and over 30% in accented words (White and Turk, 2010). Even in the latter case, this is less than the durational contrast between baseline and consonant lengthening conditions utilized here. However, as with lengthening of vowels in word-final syllables, the magnitude of word-initial consonant lengthening increases following phrase boundaries (Byrd *et al.*, 2005; Fougeron and Keating, 1997), where it is more in line with our experimental durational contrasts. Phrase-initial lengthening and articulatory strengthening of consonants have been shown to affect listeners' interpretation of the structure of ambiguous phrases (Cho *et al.*, 2007). It may be that variation in the magnitude of onset consonant lengthening is particularly associated by listeners with higher prosodic levels of preceding boundaries rather than being ubiquitous at word boundaries throughout the utterance. Further work is required to examine whether degrees of consonant lengthening are associated with different levels of prosodic structure (e.g., word vs phrase) as has been found for vowel lengthening and subsequent boundaries (e.g., Price *et al.*, 1991).

Integrating the current results with previous findings suggests a possible modification of the iambic-trochaic law as interpreted for spoken language to reflect the perceptual importance of the *locus* of prosodic lengthening effects (see White, 2002, 2014, regarding the domain vs locus distinction). A strong version of this claim would be that lengthened vowels cue a following boundary, whilst lengthened consonants cue a preceding boundary. Such a proposal for a functional division between vowels and consonants could be seen to align with proposals that the two types of segments carry distinct informational loads in speech processing (Benavides-Varela *et al.*, 2012; Bonatti *et al.*, 2005). It should be noted, however, that as well as the vocalic nucleus, coda consonants may also be lengthened preceding prosodic boundaries (e.g., Wightman *et al.*, 1992). In line with previous artificial language learning experiments, the

materials used here only featured consonants in syllable onset position, and so the impact on juncture perception of the interaction of onset, nucleus, and coda lengthening remains to be fully characterized. Listeners' structural interpretation of consonant timing processes may also be influenced by articulatory weakening of certain consonants in coda position, although this is offset by strengthening utterance-finally (e.g., Keating *et al.*, 1999). It should also be noted that many consonants do not manifest word-onset lengthening in absolute utterance-initial position (e.g., Fougeron and Keating, 1997). As discussed in White (2002, 2014), this is congruent with the functional interpretation of onset lengthening as a cue to a preceding boundary, perceptually redundant where the transition from silence to speech is itself a wholly reliable cue.

A promising alternative to a straightforward functional distinction between vowel and consonantal lengthening is suggested by consideration of the role of prediction in the interpretation of timing effects. Variation in foregoing speech rate affects both judgments of presence of phonetic material (Dilley and Pitt, 2010) and of the location of word boundaries (Reinisch *et al.*, 2011), but the means by which rate mediates such judgments is still being explored. It seems likely that listeners use speech rate to generate expectations about the duration of upcoming units (White, 2014), and the apparent functional division between the vocalic rhyme and the consonantal onset found here suggests that the location of the perceptual centre (P-centre) of the syllable may be important. The P-centre is where the syllable is perceived to occur in time and is approximately located around the start of the vowel nucleus but varies with syllable structure (Morton *et al.*, 1976). In particular, longer onsets shift the P-centre later in the syllable (Cooper *et al.*, 1986). As lengthening of the rhyme (nucleus and any coda) of the preceding syllable would also delay the upcoming P-centre, this suggests a rationale for the observed relationship between lengthening localization and segmentation behaviour. Specifically, any lengthening from (approximately) the onset of one vowel nucleus to the onset of the next will have the effect of delaying the latter syllable's P-centre relative to listeners' expectations. The salience thereby conferred on that particular syllable-to-syllable juncture may thus lead it to be interpreted as a prosodic boundary. If validated, this account would lend support to the perception-based argument for the primacy of lengthening rather than shortening effects as structural cues (White, 2014). There are obvious differences in experimental predictions prompted by the P-centre account compared with the strong vowel-consonant functional division that suggest directions for future research.

Developmental studies may also be useful in determining the nature of the mechanism through which timing cues are interpreted. It might be thought that language experience is required before the development of differential sensitivity to localized durational effects in vowels and consonants. Considering the conceptual framework of the iambic-trochaic law, for example, a preference for initial pitch-salience comparable to that of adults has been shown with 7-month-old infants, but no distinction between initial and final length-salience was found at the same age (Bion *et al.*, 2011). However, some studies suggest that the interpretation of timing cues to boundaries may be relatively independent of language-specific experience. For example, Kim, Broersma, and Cho (2012a) showed that final syllable lengthening was interpreted by both Dutch and Korean listeners as a boundary cue, whilst a concomitant F0 rise was initially only useful for Korean listeners. Furthermore, Kim, Cho, and McQueen (2012b), also testing Dutch and Korean listeners, found that both groups used VOT lengthening of word-initial voiceless stops as a segmentation cue despite the pattern being contrary to that actually observed in Dutch speech. Thus timing cues may indeed have a universal robustness that transcends language-specific phonetic details. It remains to be seen whether the functional distinction—vocalic rhyme lengthening is final, onset consonant lengthening is initial—holds in languages other than English. Such cross-linguistic data would be invaluable for determining the nature of the "perceptual" processes through which timing patterns are interpreted to linguistic ends.

ACKNOWLEDGMENTS

This work was supported by British Academy Grant No. SG090947 to the first author. We would like to thank the editor and three anonymous reviewers for their constructive comments on earlier drafts that have greatly benefited the paper. We gratefully acknowledge the assistance of Elizabeth Gabe-Thomas, Laura König, and Jean Roper in running experiments.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modeling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Beckman, M. E. (**1992**). "Evidence for speech rhythms across languages," in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (IOS Press, Oxford, UK), pp. 457–463.

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., and Mehler, J. (**2012**). "Newborn's brain activity signals the origin of word memories," Proc. Natl. Acad. Sci. U.S.A. **109**, 17908–17913.

Bion, R. A., Benavides-Varela, S., and Nespor, M. (**2011**). "Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences," Lang. Speech **54**, 123–140.

Bonatti, L. L., Pena, M., Nespor, M., and Mehler, J. (**2005**). "Linguistic constraints on statistical computations. The role of consonants and vowels in continuous speech processing," Psychol. Sci. **16**, 451–459.

Byrd, D., Lee, S., Riggs, D., and Adams, J. (**2005**). "Interacting effects of syllable and phrase position on consonant articulation," J. Acoust. Soc. Am. **118**, 3860–3873.

Cho, T., McQueen, J., and Cox, E. (**2007**). "Prosodically driven detail in speech processing: The case of domain-initial strengthening in English," J. Phonet. **35**, 210–243.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (**2004**). "Phonological phrase boundaries constrain lexical access. I. Adult data," J. Mem. Lang. **51**, 523–547.

Cooper, A. M., Whalen, D. H., and Fowler, C. A. (**1986**). "P-centers are unaffected by phonetic categorization," Percept. Psychophys. **39**, 187–196.

Cutler, A. (**1986**). "Forbear is a homophone: Lexical prosody does not constrain lexical access," Lang. Speech **29**, 201–220.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (**2002**). "Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition," J. Exp. Psychol. Hum. Percept. Perform. **28**, 218–244.

Dilley, L. C., and Pitt, M. A. (**2010**). "Altering context speech rate can cause words to appear or disappear," Psychol. Sci. **21**, 1664–1670.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Van Der Vreken, O. (**1996**). "The MBROLA project: Towards a set of high-quality speech

synthesizers free of use for non-commercial purposes," in *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia, pp. 1393–1396.
Fougeron, C., and Keating, P. A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.
Fowler, C. A. (**1990**). "Lengthenings and the nature of prosodic constituency: Comments on Beckman and Edwards's paper," in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge, UK), pp. 201–207.
Gout, A., Christophe, A., and Morgan, J. L. (**2004**). "Phonological phrase boundaries constrain lexical access. II. Infant data," J. Mem. Lang. **51**, 548–567.
Gow, D. W., and Gordon, P. C. (**1995**). "Lexical and prelexical influences on word segmentation: Evidence from priming," J. Exp. Psychol. Hum. Percept. Perform. **21**, 344–359.
Hay, J. S., and Diehl, R. L. (**2007**). "Perception of rhythmic grouping: Testing the iambic/trochaic law," Percept. Psychophys. **69**, 113–122.
Hayes, B. (**1995**). *Metrical Stress Theory: Principles and Case Studies* (University of Chicago Press, Chicago), pp. 1–455.
Johnson, E. K., and Seidl, A. H. (**2009**). "At 11 months, prosody still outranks statistics," Dev. Sci. **12**, 131–141.
Keating, P. A., Cho, T., Fougeron, C., and Hsu, C. (**2003**). "Domain-initial strengthening in four languages," in *Papers in Laboratory Phonology 6*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge, UK), pp. 145–163.
Keating, P., Wright, R., and Zhang, J. (**1999**). "Word-level asymmetries in consonant articulation," UCLA Work. Pap. Phonet. **97**, 157–173.
Kim, S., Broersma, M., and Cho, T. (**2012a**). "The use of prosodic cues in learning new words in an unfamiliar language," Stud. Second Lang. Acquisit. **34**, 415–444.
Kim, S., Cho, T., and McQueen, J. M. (**2012b**). "Phonetic richness can outweigh prosodically driven phonological knowledge when learning words in an artificial language," J. Phonet. **40**, 443–452.
Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge, UK), pp. 1–334.
Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.
Morton, J., Marcus, S., and Frankish, C. (**1976**). "Perceptual centers (P-centers)," Psychol. Rev. **83**, 405–408.
Oller, D. K. (**1973**). "The effect of position in utterance on speech segment duration in English," J. Acoust. Soc. Am. **54**, 1235–1247.
Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (**1991**). "The use of prosody in syntactic disambiguation," J. Acoust. Soc. Am. **90**, 2956–2970.
Quené, H. (**1992**). "Durational cues for word segmentation in Dutch," J. Phonet. **20**, 331–350.
Reinisch, E., Jesse, A., and McQueen, J. M. (**2011**). "Speaking rate from proximal and distal contexts is used during word segmentation," J. Exp. Psychol. Hum. Percept. Perform. **37**, 978–996.
Saffran, J., Aslin, R., and Newport, E. (**1996a**). "Statistical learning by 8-month old infants," Science **274**, 1926–1928.
Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996b**). "Word segmentation: The role of distributional cues," J. Mem. Lang. **35**, 606–621.
Salverda, A. P., Dahan, D., and McQueen, J. M. (**2003**). "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," Cognition **90**, 51–89.
Shatzman, K. B., and McQueen, J. M. (**2006**). "Segment duration as a cue to word boundaries in spoken-word recognition," Percept. Psychophys. **68**, 1–16.
Shukla, M., Nespor, M., and Mehler, J. (**2007**). "Interaction between prosody and statistics in the segmentation of fluent speech," Cognit. Psychol. **54**, 1–32.
Spinelli, E., McQueen, J. M., and Cutler, A. (**2003**). "Processing resyllabified words in French," J. Mem. Lang. **48**, 233–254.
Tagliapietra, L., and McQueen, J. M. (**2010**). "What and where in speech recognition: Geminates and singletons in spoken Italian," J. Mem. Lang. **63**, 306–323.
Turk, A. E., and Shattuck-Hufnagel, S. (**2007**). "Multiple targets of phrase-final lengthening in American English words," J. Phonet. **35**(4), 445–472.
Tyler, M. D., and Cutler, A. (**2009**). "Cross-language differences in cue use for speech segmentation," J. Acoust. Soc. Am. **126**, 367–376.
White, L. (**2002**). "English speech timing: A domain and locus approach," Ph.D. dissertation, University of Edinburgh (http://www.cstr.ed.ac.uk/projects/eustace/dissertation.html).
White, L. (**2014**). "Communicative function and prosodic form in speech timing," Speech Commun. **63**, 38–54.
White, L., and Turk, A. E. (**2010**). "English words on the Procrustean bed: Polysyllabic shortening reconsidered," J. Phonet. **38**, 459–471.
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (**1992**). "Segmental durations in the vicinity of prosodic phrase boundaries," J. Acoust. Soc. Am. **91**, 1707–1717.