Research Article

Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues?

Laurence White [a],*, Silvia Benavides-Varela [b], Katalin Mády [c]

[a] School of Education, Communication and Language Sciences, Newcastle University, United Kingdom
[b] Department of Developmental Psychology and Socialisation, University of Padua, Italy
[c] MTA Research Institute for Linguistics, Budapest, Hungary

ABSTRACT

Speech segments are lengthened at the onsets and offsets of linguistic constituents. Final-syllable vowel lengthening is proposed to be a language-universal cue to word segmentation, but cross-linguistic investigations of the perception of initial consonant lengthening are lacking. We compared the use of word-initial consonant lengthening and word-final vowel lengthening by native speakers of English, Hungarian and Italian, using an artificial language learning task and varying vowel and consonant durations between subjects within each language group. Word-final vowel lengthening was only exploited for segmentation by English speakers; we interpret its non-universality as potentially due, at least in part, to language-specific functional loads on vowel duration, used for indicating lexical stress in Italian and vowel identity in Hungarian. By contrast, all three language groups used word-initial consonant lengthening to locate word boundaries, but did not benefit from lengthening of vowels in word-initial syllables. If domain-initial consonant timing effects are universal, it may be because they promote two related but separable processing requirements for the listener: (a) word segmentation, boosted by lengthening across the boundary; (b) lexical access, boosted by articulatory strengthening and lengthening of word onsets.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

In many languages, speech sounds have been observed to be lengthened at the starts and/or ends of linguistic constituents (for reviews: Fletcher, 2010; Klatt, 1976; White, 2014). Here we explore how such timing effects in speech production may interact with listeners' perceptual requirements to give rise to word segmentation cues which may be widely used across languages and, potentially, tend towards perceptual universals. Specifically, in three artificial language learning experiments, we investigate lengthening of consonants in word onsets and lengthening of vowels in word-initial and word-final syllables. We test the interpretation of these localised lengthening effects by speakers of three languages – English, Hungarian, Italian – in which some linguistic functions of speech timing are otherwise rather distinct.

Such tendencies towards universal interpretations of linguistic sound patterns may have their origins in anatomical and physiological constraints on speech production (e.g., Gussenhoven, 2016) and in how such constraints interact with listeners' perceptual biases (e.g., Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015). Thus, although our experimental focus is on speech perception, we start by briefly reviewing the cross-linguistic occurrence of localised domain-edge lengthening effects in speech production and asking why such effects are so widespread, and then examining the evidence across languages for listeners' interpretation of these lengthening effects as cues to boundaries.

1.1. Domain-edge lengthening in speech production

It is well established, in prosodically diverse languages, that syllable onset consonants are lengthened word-initially (e.g., French, Korean, Taiwanese: Keating, Cho, Fougeron, & Hsu, 2004; English: Oller, 1973) and may be longer still phrase-initially (English: Fougeron & Keating, 1997). Additionally, syllable vowel nuclei are lengthened at the ends of phrases and

utterances, along with coda consonants (e.g., Hebrew: Berkovits, 1994; Dutch: Gussenhoven & Rietveld, 1992; English: Klatt, 1975; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992), a phenomenon assumed to be widespread in the world's languages (Vaissière, 1983) and possibly universal (Tyler & Cutler, 2009; see Fletcher, 2010, for an extensive review of production studies).

In the case of initial lengthening, segments are also hyper-articulated – produced with more forceful and exaggerated gestures – at the starts of utterances, with evidence of gradual "supralaryngeal declination" as utterances proceed, for both consonants (Hinton, 1996) and vowels (Vayra & Fowler, 1992). Thus, onset gestural strengthening and lengthening tend naturally to co-occur, with shorter segments later in constituents typically characterized by undershoot of target gestures, but lengthening and strengthening are not necessarily correlated (Moon & Lindblom, 1994). Indeed, *utterance*-initial consonants are strongly articulated (Fougeron & Keating, 1997) whilst often, but not always, being relatively short (Cho, Kim, & Kim, 2017; Fourakis & Monahan, 1988; White, 2002).

This overall tendency towards early hyperarticulation (or, concomitantly, hypoarticulation later in the utterance) is of course mediated by linguistic structure. Keating, Cho, Fougeron, & Hsu (2004) found that speakers of French, Korean and Taiwanese all distinguish at least two levels of linguistic structure through lengthening and strengthening of syllable onsets, with the number of distinct levels and the degree of association between lengthening and strengthening varying between languages. Importantly, such production trends appear to converge with perceptual demands: as discussed below, relative hyperarticulation early in linguistic units is consistent with the listener's need for maximally informative evidence about the identity of speech sounds to be available near the start of words and phrases (Hall, Hume, Jaeger, & Wedel, 2018). Firstly, however, we consider the specific speech production constraints that may contribute to the widespread occurrence of domain-edge lengthening both initially and finally.

1.1.1. Production-based accounts of domain-edge lengthening

Cooper & Paccia-Cooper (1980) proposed that phrase-/utterance-final lengthening arises from speech planning, allowing speakers more time to construct the post-boundary phrase. This explanation obviously does not account for the lengthening observed at the end of a speaker's *turn*, however, and so a subsidiary account invoked the notion of deceleration at the end of major units due to processing fatigue (Cooper & Paccia-Cooper, 1980; Gussenhoven & Rietveld, 1992). An alternative, possibly complementary, interpretation of final lengthening – within a framework of biologically-based prosodic codes – associates longer vowels with pitch peak delay, another means of signalling pitch height (Gussenhoven, 2016); this is unlikely to be the only motivation for final lengthening, however, given its ubiquity in the face of the variability of final pitch movements.

We offer the additional suggestion that this apparent ubiquity – of both initial and final lengthening cross-linguistically – may be linked to a non-linguistic phenomenon evident in the organisation of limb movement in primates, formalised as the "minimum-jerk" principle (Flash & Hogan, 1985; Hogan, 1984). This principle, which has also been related to jaw trajectories during speech (Nelson, 1983), states that movements are planned so as to minimise the first derivative of acceleration (i.e., jerk: the rate of change of acceleration). The implications for speech are that the onset of articulation, other constraints aside, should be marked by smooth, gradual acceleration and thus a lengthening of initial gestures, whilst articulation offset should be marked by parallel smooth deceleration, with final gestures also lengthened. Of course, this general principle of motor control is potentially subject in speech planning to linguistically or perceptually motivated amendment: indeed, as noted above, segments are often actually shorter utterance-initially, potentially because listeners do not require a word-boundary cue at the utterance onset (White, 2002, 2014). Nevertheless, the minimum-jerk principle provides what Dingemanse et al. (2015) call a perceptuomotor analogy, encouraging listeners to interpret lengthening – where it does occur – as boundary-adjacent (further discussed below).

These production-based explanations of the cross-linguistic preponderance of domain-edge lengthening effects are not mutually exclusive: indeed, prosodic universals are more likely to emerge and endure if there are multiple convergent pressures towards a particular form-function mapping. Moreover, whilst Cutler (1990) contrasted production-based explanations of final lengthening with its interpretation as an explicit cue to structure for listeners, a prosodic effect may have its origins in anatomical or physiological constraints, but also be a signal for listeners under the speaker's control (as discussed by Gussenhoven, 2016). Communicative pressures – considered in the following section – may modify or reinforce the realisation, distribution and interpretation of articulatorily-derived cues.

1.2. Domain-edge lengthening in speech perception

Word onsets are critical for both segmentation and lexical recognition by listeners. As described by Content, Kearns, and Frauenfelder (2001), there is no necessary processing symmetry between the starts and the ends of constituents; indeed, due to recursion, some syntactic constituents may remain open whilst new constituents of the same type are established. More generally, decisions about when to end a constituent may be delayed – for example, to allow the resolution of temporary ambiguity – whilst the activation of new lexical candidates is much more urgent. This asymmetry gives rise to the hypothesized Syllable Onset Segmentation Heuristic, which prioritizes identification of potential word onsets in speech recognition (Content et al., 2001).

Lengthening of word onsets has been shown to provide Dutch and English listeners with a cue to structure (Quené, 1993; White, Mattys, Stefansdottir, & Jones, 2015), in particular, indicating the presence of a foregoing word boundary. This is in line with a functional approach to speech timing which proposes that lengthening effects are widely used as cues to constituent heads and edges (White, 2014). That functional perspective also provides an account of the relative lack of lengthening sometimes observed utterance-initially, because the timing cue to the foregoing word boundary is redundant when the word is immediately preceded by a inter-utterance pause (White, 2002, 2014).

Congruent with this functional view, articulatory strengthening early in linguistic constituents (or, from the converse perspective, "supralaryngeal declination" later on) is reinforced through its facilitation of word recognition. The word onset is the point of maximum uncertainty about lexical identity and thus the listener's needs are supported by hyperarticulation of initial consonants. By contrast, increasing availability of foregoing context as the constituent – word, phrase, utterance – unfolds will tend to allow listeners better to predict and hence disambiguate segmental identity, licensing the speaker towards relative hypoarticulation (Houlihan, 1975).

It has long been demonstrated that final lengthening promotes perception of an upcoming linguistic boundary (for English: Beach, 1991; Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991). Indeed, final lengthening has been widely suggested as a universal cue to upcoming linguistic boundaries: for example, Tyler and Cutler (2009) found that Dutch, English and French learners of an artificial language (AL) all benefitted from lengthening of the word-final vowel, but not from the lengthening of the vowel in the word-initial syllable, despite the association between vowel lengthening and predominant word-initial stress for Dutch and English. Similarly, both Kim, Broersma, and Cho (2012), with Dutch and Korean speakers, and Frost, Monaghan and Tatsumi (2016), with English and Japanese speakers, showed that AL learning was boosted when the final syllable was lengthened. In all three AL studies (as in our experiments), increased learning – indexed by word recognition in a two-alternative forced-choice task following exposure to the language stream – is assumed to reflect enhanced segmentation during the exposure phase.

1.2.1. Perception-based accounts of domain-edge lengthening

The power of final lengthening for segmentation has been posited as a domain-general principle of auditory grouping, the "Iambic-Trochaic law": it is claimed that when alternate sounds are made prominent through variation along prosodic dimensions, sounds with higher amplitude or pitch are universally heard as sequence-initial, whilst sounds with greater duration are universally heard as sequence-final (Daniela, Nespor, & Toro, 2013; Hay & Diehl, 2007; Woodrow, 1909). However, the universality of the Iambic-Trochaic law has been questioned in studies indicating that the interpretation of such sequences is dependent on language-experience (Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Bion, Benavides-Varela, & Nespor, 2011; Iversen, Patel, & Ohgushi, 2008; Molnar, Carreiras, & Gervain, 2016). For example, whilst Italian-speaking adults better remembered disyllabic sequences when the duration pattern was short-long rather than long-short, Italian-learning seven-month-olds showed no bias towards the short-long pattern (although they did prefer sequences that had a high-low pitch pattern rather than low–high, Bion et al., 2011).

With specific regard to word segmentation, Ordin, Polyanskaya, Laka, & Nespor (2017) found that final-syllable vowel lengthening in trisyllables boosted artificial language learning for native German speakers and Gipuzkoan Basque-Spanish bilinguals, but did not benefit monolingual Spanish or Italian speakers. They interpret this as reflecting the competing interpretation of vowel lengthening as a cue to

stress. Italian and Spanish have predominant word-penultimate stress, with duration as an important cue given the lack of vowel reduction in unstressed syllables: the absence of significant vowel quality differences between stressed and unstressed syllables, for example in Dutch, makes suprasegmental cues to stress more important than in English, where vowel reduction is the primary indicator of unstressed syllables (Cutler & Van Donselaar, 2001; Cutler, 1986). Italian speakers, in particular, have a strong expectation of long vowels as marking penultimate stressed syllables, with final stress only marked by lengthening before phrase boundaries (Rogers & d'Arcangeli, 2004). Indeed, Italians' performance was boosted by lengthening of the penultimate vowels in the AL trisyllables (Ordin, Polyanskaya, Laka, & Nespor, 2017). By contrast, German lexical stress placement is more variable and duration is not the primary stress cue (Féry, Hörnig, & Pahaut, 2011; Nespor et al., 2008), whilst stress is non-contrastive in Gipuzkoan Basque (Elordieta & Hualde, 2014).

Thus, there appear to be language-specific amendments to the "general law that longer durations are perceived as sequence final events" (Frost, Monaghan, & Tatsumi, 2016). The status of this hypothesized general law is further challenged by the observation that lengthened consonants in syllable onsets are perceived as a cue to a preceding, rather than a following, boundary. There have been several demonstrations of the power of initial lengthening/strengthening for word segmentation, in natural speech when combined with other cues (e.g., preboundary strengthening, allophony): for example, for Dutch speakers (Quené, 1993; Shatzman & McQueen, 2006), Italian speakers (Tagliapietra & McQueen, 2010) and English speakers (Gow & Gordon, 1995). Using an AL paradigm, White et al. (2015) found that English speakers showed learning benefits, in the absence of other segmentation cues beyond the statistical contingencies of the language, when trisyllabic AL words had lengthened initial consonants. Moreover, for Dutch speakers, lengthening of the voice-onset time (VOT) of word-initial voiceless stops was shown to promote segmentation in AL learning, despite these speakers' experience of VOT *shortening* at higher prosodic boundaries in their natural speech (Kim, Cho, & McQueen, 2012): in this case, the segmentation boost provided by initial lengthening overrode language-specific regularities.

1.3. The present study

As described above, final lengthening has been widely investigated cross-linguistically, and our primary goal is to explore the use of initial lengthening in diverse languages and compare its power with that of final lengthening. We consider the hypothesis that, given the additional benefit for lexical access of hyperarticulated (longer/stronger) word-initial gestures, onset effects may be more robust cross-linguistically than final lengthening as segmentation cues.

We used an artificial language paradigm to test whether onset consonant lengthening is a sufficient segmentation cue for speakers of three different languages: English, Hungarian and Italian. These languages were chosen because of their diverse consonant timing and prosodic characteristics (see below). We also compared listeners' use of word-initial consonant lengthening with their interpretation of vowel lengthening

in word-initial and word-final syllables. Ordin, Polyanskaya, Laka, & Nespor (2017) called into question the universality of final-vowel lengthening as a segmentation cue; in particular, they suggested that language-specific use of vowel lengthening as a cue for lexical stress contrast may override any language-independent tendency to interpret lengthening as indicative of an upcoming boundary. The use of speakers of English, Hungarian and Italian allows us to validate and further explore that hypothesis, as the three languages also have different patterns of lexical stress and distinct uses of vowel duration.

With regard to segmental timing, English has only singleton consonants. In many English accents, tense-lax vowel pairs contrast in duration, but this is accompanied by quality differences (Giegerich, 1992). By contrast, there is a binary phonemic consonant length in Italian (singletons *vs* geminates, Rogers & d'Arcangeli, 2004; additionally with post-lexical geminates in non-northern accents, Nespor & Vogel, 2007), whilst Hungarian has binary phonemic length distinctions for both consonants and vowels (Szende, 1994). As for stress, Italian stress is lexically contrastive and (as above) strongly marked by vowel lengthening when on the word-penultimate syllable (the predominant location for stress), but not when word-final (Bertinetto, 1980; Rogers & d'Arcangeli, 2004), and only marginally in antepenultimate syllables (d'Imperio & Rosenthall, 1999). Hungarian stress is non-contrastive, always occurring on the word-initial syllable, and duration is not a primary marker of stress (Fónagy, 1958), except where accompanied by phrasal accent (Szalontai, Wagner, Mády, & Windmann, 2016). English has predominant word-initial stress (Cutler & Carter, 1987), with lengthening of vowels within stressed syllables (Klatt, 1976), although vowel quality is claimed to be more important than duration as a stress cue for English listeners (Cutler, 1986).

Final lengthening effects have been found for all three languages, at utterance-edges and, variably, also at lower level of prosodic structure (e.g., English: Klatt, 1975; Wightman et al., 1992; Hungarian: Hockey and Fagyal, 1999; White & Mády, 2008; Italian: D'Imperio & Fivela, 2003; White, Payne, & Mattys, 2009). Lengthening of consonants in word-initial position is well documented for English (e.g., Oller, 1973) and hyperarticulation of word-initial consonants has been shown for Italian (Farnetani & Vayra, 1996), although as noted above, interpretation of Italian consonant duration is complicated by post-lexical processes, primarily in non-northern accents (for further discussion, see Tagliapietra & McQueen, 2010, who report that singleton consonants in Italian are subject to word-initial lengthening). There are few data on Hungarian onset consonant duration, although Mády (in preparation) found 50% lengthening of word-initial voiceless stops relative to medial position.

In accordance with the observed timing effects in production and the lexical access requirement in perception, we predict that word-initial lengthening of consonants should be perceived by speakers of all three languages as a cue to a preceding boundary. An alternative hypothesis is that speakers of languages which have a consonant length distinction between singletons and geminates (Hungarian, Italian) may not use timing as a segmentation cue. For word-final vowel lengthening, most foregoing work has suggested this as a universal cue to word boundaries; however, in line with recent results (e.g., Bhatara,

Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Iversen, Patel, & Ohgushi, 2008; Ordin, Polyanskaya, Laka, & Nespor, 2017), it may be that interpretation of vowel lengthening (examined here in word-initial and word-final syllables) is conditioned by language-specific influences on vowel duration, as a cue to stress (Italian) or segmental identity (Hungarian).

2. Experiment 1

We wished to determine if lengthening of a word-initial consonant could serve as a sufficient cue to word boundaries: (a) consistently between languages, regardless of whether they have a phonemic consonant length distinction; (b) in the absence of other segmentation-relevant information (allophony, intonation, preboundary lengthening, etc.). We adopted an artificial language learning methodology, as used in several previous studies of speech timing and segmentation (e.g., Saffran, Newport, & Aslin, 1996; for a recent review, see Frost, Monaghan, & Tatsumi, 2016), in which syllable transition probabilities provide a consistent guide to lexical structure across all timing conditions. Note that, in our experiments, comparisons between timing conditions are between-subjects. Specifically, we followed the procedure of White et al. (2015), who showed that English listeners were better at recognising AL words at test when the duration of the initial consonant of each trisyllable was lengthened during exposure. We tested native speakers of Hungarian and Italian, both languages in which consonant duration indicates phonemic identity; we also tested native English speakers, to validate the results of White et al. (2015) with our new artificial language and to compare the strength of effects with those found for Hungarian and Italian.

2.1. Method

2.1.1. Participants

We tested 80 native speakers of English (at Plymouth University, UK), 80 native speakers of Italian (University of Padua, Italy) and 80 native speakers of Hungarian (Hungarian Academy of Sciences, Budapest, Hungary), with no reported speech or hearing problems. All received either course credit or a small honorarium for their participation. No participants had been raised as functionally bilingual (i.e., they did not grow up speaking an additional language, although most were exposed to other languages during their secondary school education and beyond). These participant characteristics were equivalent for all three experiments. For each language group, participants were randomly allocated to one of two timing conditions (40 per language per condition).

2.1.2. Design

We constructed an artificial language stream of 12 consonant–vowel (CV) syllables. These were formed into four trisyllabic words: *bitusa*, *nudopa*, *ripolu*, *subako*, having first obtained native speaker judgements to eliminate any words that closely resembled real words for any of our three language groups. For each word, we chose onset consonants that were not voiceless stops (two continuants, one nasal, one voiced stop), to obviate the possibility that consonant lengthening could perceptibly increase the silent closure interval across a word boundary in voiceless stops.

The four words were pseudo-randomly ordered into a 500-word stream (125 repetitions of each word), with the constraint that no word was immediately repeated. Thus, the transition probability from the final syllable of one word to the first syllable of the next was always 0.33. As no syllables were used in more than one word, the transition probability between syllables within words was always 1 (e.g., *bi-tu-sa*).

There were two timing conditions. In the Flat condition, all segments (consonants and vowels) were 120 ms long. In the Syl1-C condition, the onset consonant of the first syllable of each word was 170 ms long (underlined in *bitusa*, *nudopa*, *ripolu*, *subako*) and all the other segments were 110 ms long. These durations were selected to keep total word duration equal between conditions; hence total AL stream length was six minutes for each condition. Five-second ramps were used to fade the streams in and out, so that no words were immediately preceded or followed by clear silence.

The AL streams for the two timing conditions were generated using the diphone synthesizer MBROLA (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996). We used the *en1* male British English voice, with a constant fundamental frequency of 120 Hz. Given that the voice is native for one speaker group, it was possible that the English could have an overall advantage for learning the AL. Furthermore, as this experiment is not designed to test statistical learning based on syllable transition probabilities (Saffran et al., 1996), we did not construct multiple AL streams. Based on an extensive foregoing literature, we fully expected that speakers of all three languages would be able learn the AL words. Linguistic idiosyncrasies might be manifest in a main effect of native language; however, our experimental hypotheses are concerned with differences between timing conditions. Furthermore, as reported below, the English speakers did not show a native speaker advantage for AL learning.

Foils for the test phase following AL exposure were constructed from the syllables of the four words, following the procedure of White et al. (2015). Part-word foils comprised syllables from the end of a word, followed by syllables from the start of another word: e.g., *tusanu* from *bitusa*...*nudopa*; thus, these foils were occasionally heard within the AL stream, but only a third as often than the words themselves. Non-word foils were constructed from three syllables in the AL that were never heard contiguously, e.g., *lukopa*. In the test phase, each word was paired with three foils (two part-words and one non-word), in two different orders (word-foil, foil-word) for each pairing. Thus, there were 24 test trials altogether, comprised of all these word-foil pairings/orderings. Words and foils for the test phase were always synthesised with each segment having 120 ms duration.

2.1.3. Procedure

Participants were told that they would hear an artificial language through headphones for six minutes and that their task was just to listen and try to discover the words in the language. Following this exposure phase, they were given the instructions for the test phase, in which participants were told that they would hear a pair of sound sequences on each trial and they had to decide which of the pair was a word in the AL. This two-alternative forced choice method is typical of AL learning experiments (e.g., Saffran et al., 1996). Here however we used

a somewhat shorter test phase, not pairing all words with all foils, as is typical when the goal is to examine statistical learning based on syllable transition probabilities. In this case, our interest was in performance differences between versions of the same AL that differed in terms of timing (as in White et al., 2015); we were not concerned with statistical learning *per se*.

There was a response time-out of 10 s. When participants timed out on more than one trial, we took this as evidence that they had not understood the procedure correctly. These participants were thus replaced, prior to analysis, with new participants in the same condition, so that there were always 40 participants in each condition for each language (see below).

2.1.4. Statistical analysis

Mixed-effect regression models, with random intercepts for participant and test item, were constructed incrementally. The null model included just the random effects and a null intercept. We first added the primary experimental factor of Timing and compared this to the null model using a log-likelihood χ^2 test. We then added the factor of Language to the Timing model, and compared those models. We finally added the Timing/Language interaction to the Timing + Language model and compared those models. To give a standardized estimate of effect size, we calculated Cohen's-*d* for all pairwise comparisons of mean scores.

2.2. Results

One English participant in the Flat timing condition made no response within ten seconds on one trial and so was timed out: this response was treated as a missing datum. Two Hungarian participants in the Flat condition had multiple timeouts and were replaced with two new participants.

Logistic regression showed that overall recognition of the words in the test phase was above chance in both the Flat condition, $M = 64\%$, $z = 4.43$, $p < 0.0001$, and the Syl1-C condition, $M = 74\%$, $z = 9.03$, $p < 0.0001$. Performance was significantly higher in the Syl1-C, however, as indicated by a main effect of Timing: $\chi^2(1) = 26.92$, $p < 0.0001$, $d = 0.23$. Thus, lengthening of the initial consonant in each trisyllable of the AL stream improved subsequent recognition of those words.

There was also a main effect of Language, $\chi^2(2) = 19.39$, $p < 0.001$, and a marginal interaction between Timing and Language, $\chi^2(2) = 5.28$, $p = 0.071$. As shown in Table 1, correct responses in the Syl1-C condition were significantly higher than in the Flat condition for speakers of English, Hungarian and Italian taken separately. This replicates the previous finding for English speakers (White et al., 2015), and additionally shows that both Hungarian and Italian speakers also used lengthening of the word-initial consonant as a cue for AL segmentation. Finally, as indicated in Table 2, all speaker groups had recognition scores significantly better than chance in both the Flat and Syl1-C conditions, in accordance with the extensive research demonstrating listeners' learning of artificial languages based on syllable transition probabilities (e.g., Saffran et al., 1996).

Inspecting the mean scores, the Language effect and the marginal Timing × Language interaction are likely to relate to the word-initial lengthening boost being larger for Hungarian

Table 1
Mean correct word recognition scores and pairwise planned comparisons between timing conditions (Flat vs Syl1-C) for Experiment 1.

	Mean Flat (%)	Mean Syl1-C (%)	$\chi^2(1)$	p	d
English	64	72	6.69	0.010	0.17
Hungarian	67	82	16.66	<0.0001	0.35
Italian	60	68	6.38	0.012	0.17

Table 2
z-values and p-values comparing correct word recognition to chance performance (50%) for individual language/timing condition groups.

	Flat		Syl1-C		Syl1-V		Syl3-V	
	z	p	z	p	z	p	z	p
English	4.20	<0.0001	6.01	<0.0001	0.97	0.324	7.12	<0.0001
Hungarian	4.56	<0.0001	7.80	<0.0001	2.06	0.040	5.24	<0.0001
Italian	2.58	0.001	5.13	<0.0001	1.76	0.078	2.56	0.010

speakers than for the other two language groups (Table 1). To further assess these patterns, we analysed performance differences between the three languages separately for the two timing conditions. In the Flat condition, there was no reliable difference in correct scores between English speakers and those of either of the other two languages, but Hungarians' performance was higher than that of Italians, $\chi^2(1) = 3.88$, $p = 0.049$, $d = 0.13$. In the Syl1-C condition, there was no reliable difference in correct scores between English and Italian speakers, but Hungarians' overall performance was higher than both Italian speakers, $\chi^2(1) = 14.34$, $p = 0.0002$, $d = 0.31$, and English speakers, $\chi^2(1) = 8.10$, $p = 0.0044$, $d = 0.22$. These comparisons between speaker groups are post-hoc, but even using a conservative Bonferroni-corrected alpha level of 0.008, the Hungarian speakers out-performed the other two groups in the Syl1-C condition: possible interpretations of this advantage are explored below.

2.3. Discussion

Experiment 1 showed that speakers of English, Hungarian and Italian all learnt the artificial language better when the onset consonant of each word-initial syllable was lengthened. As with other studies of segmentation using AL methodology, we take this to mean that the available segmentation cue – onset consonant lengthening – was interpreted as indicating the presence of an immediately preceding word boundary by speakers of all three languages. For English, this replicates White et al. (2015) with a different AL, one with no word-initial voiceless stops in this case. For Italian and Hungarian, an alternative hypothesis was that the presence of phonemic consonant length distinctions – short singletons and long geminates – would mean that speakers of these languages did not also use consonant duration for determining word boundary location. The results strongly suggest, however, that listeners can interpret consonant duration as a cue to a boundary, even when it is also used in their native language as a cue to phonemic identity (noting that this artificial language did not require discrimination between singleton and geminate consonants).

For Italians, this result is in line with Tagliapietra and McQueen's (2010) interpretation of their data, derived from experiments using natural speech with the possibility of other segmentation cues being available. Here we eliminated all cues except onset consonant lengthening and syllable transition probabilities (the latter being constant between timing conditions). Thus, the difference between conditions strongly shows that consonant length is a cue to word boundaries. This finding accords with the cross-linguistic observation of word-initial lengthening (Keating, Cho, Fougeron, & Hsu, 2004).

With regard to the performance differences between languages: Hungarian speakers were overall better at the task in the Syl1-C condition than the other two groups and marginally better than Italians in the Flat condition. In a between-subjects design – as cross-linguistic experiments necessarily are – it is difficult to interpret the source of such differences. The difference in the Syl1-C condition is unlikely to be due to Hungarian speakers' greater sensitivity to consonant duration due to their phonemic consonant length distinction, as a similar singleton/geminate distinction is also present in Italian. We will refrain from speculating about the sources of these cross-linguistic differences, however: firstly, the Timing × Language interaction was only marginally significant; secondly, between-group factors that are independent of language (variation in participant motivation, for example) could be at play.

So far, we have shown that word-initial consonant lengthening boosts segmentation for all three languages, in the absence of additional cues. We have not shown, however, that the lengthening has to be localized on the consonant. Firstly, it has been claimed that any regular lengthening effect could serve, in sequences of AL words of the same syllable length, as a metronomic cue to structure (Tyler & Cutler, 2009), in which case vowel lengthening in the first syllable should boost all language groups in similar fashion. Secondly, it is possible that lengthening of the initial syllable vowel would also be a linguistic segmentation cue, particularly in Hungarian, which has fixed word-initial stress, marked by lengthening in words with phrasal stress (Szalontai, Wagner, Mády, & Windmann, 2016). If so, we would expect to see cross-linguistic variation in listeners' use of initial syllable vowel (rather than consonant) lengthening, according to language-specific distribution and phonetic marking of lexical stress.

3. Experiment 2

We tested the impact on segmentation performance of localized lengthening of the vowel in the first syllable of each AL trisyllable. Firstly, we needed to check that listeners' interpretation of word-initial lengthening – as found in Experiment

1 – relies on the lengthening being localized on the onset consonant, in line with the production of word-initial lengthening in English (e.g., Oller, 1973). Secondly, we needed to compare listeners' interpretation of vowel lengthening between languages. Hungarian has fixed word-initial stress, which could bias listeners to interpret vowel lengthening as a cue to an immediately-preceding word boundary. There is also a phonemic vowel length contrast in Hungarian, however, which might militate against using vowel length as a segmentation cue. For Italian, stressed vowels in trisyllable-initial position are not greatly lengthened, unlike vowels in the predominant penultimate stress position in trisyllables, so the first syllable vowel lengthening is unlikely to be interpreted as a stress cue boosting segmentation. For English, most words begin with stressed syllables, but lengthening is not the primary perceptual indicator of stress; indeed, Tyler and Cutler (2009) found that lengthening of the first syllable vowel was unhelpful for segmentation with English listeners. Thus we expect that for languages like English, where word-final lengthening is a cue to segmentation, vowel lengthening in the word-initial syllable should actually reduce segmentation performance, as it suggests to the listener that a boundary follows the word-initial syllable, in conflict with the statistical cues to lexical structure.

3.1. Method

We tested 40 new participants each for English, Hungarian and Italian. The procedure was as for Experiment 1, but in this case, all participants heard an AL with vowel lengthening in the initial syllable. Thus, in this Syl1-V condition, the vowel of the first syllable (underlined in _bitusa_, _nudopa_, _ripolu_, _subako_), was 170 ms long, whilst all other segments were 110 ms long. Participants heard the six-minute AL stream thus created, followed by the 24-trial two-alternative forced-choice test phase, as for Experiment 1.

3.2. Results

As in Experiment 1, cases where participants timed out on a single trial (i.e., made no response within ten seconds) were treated as missing data (English: 3; Italian 4). There were also 3 English and 2 Italian participants with multiple time-outs: these were replaced to retain 40 participants in each condition for each language.

Overall word recognition was above chance in a logistic regression in the Syl1-V condition (lengthening of vowel in initial syllable), M = 55%, z = 2.00, p = 0.045. Performance across all participants was lower than the baseline Flat condition in Experiment 1, however, $\chi^2(1)$ = 23.86, p < 0.0001, d = 0.17. As shown in Fig. 1, the correct response rate in every language was lower for Syl1-V than for the Flat condition. Thus, whilst lengthening of the consonant in the word-initial syllable acted as a strong cue to a preceding word boundary, lengthening of the vowel in the same syllable actually served to decrease listeners' ability to infer the statistically-defined word boundaries in the AL stream.

There was a marginal effect of Language, $\chi^2(2)$ = 4.61, p = 0.09987, and no interaction between Timing and Language, $\chi^2(2)$ = 3.65, p = 0.160. Exploring differences between languages further, the effect of the Syl1-V condition is summa-

rized in Table 3. Syl1-V performance is numerically worse than in the Flat condition for all three, although this difference is not statistically significant for Italian speakers. As shown in Table 2, the effect of lengthening the vowel in the first syllable of every word in the AL stream is to reduce subsequent recognition performance by English speakers to chance level. Italian participants' performance is only marginally better than chance, while Hungarian participants do perform above chance in the Syl1-V condition.

Comparing performance between the three language groups in the Syl1-V condition, Hungarians' performance was marginally higher than that of the English speakers, $\chi^2(1)$ = 3.52, p = 0.060, d = 0.10, but there was no significant difference between the Italian speakers' performance and that of the other two groups. Applying a Bonferroni correction for multiple post-hoc comparisons (alpha level of 0.017), we conclude that there are no differences between language groups to be interpreted.

3.3. Discussion

Taken together with Experiment 1, the results of Experiment 2 clearly showed the importance of the localization of lengthening effects. A relatively long consonant in syllable onset position is interpreted as a cue to an immediately preceding boundary, even in the absence of other segmental or prosodic cues. By contrast, lengthening the vowel in the same word-initial syllable is at best neutral and at worst a hindrance to appropriate segmentation by listeners. This combination of results, congruent with previous findings for English (White et al., 2015), strongly indicates the specific importance of word onsets for initiating the activation of lexical hypotheses.

An alternative interpretation of word-initial lengthening could have been that it provides listeners with a metrical cue to segmentation, either specifically in the rather constrained case of artificial language learning or with respect to the metrical structure of listeners' native languages. In the specific case, it has been argued that prosodic cues in AL streams in which each successive word is trisyllabic serve simply to draw attention to the regularity of lexical structure (Tyler & Cutler, 2009). If this were the case, however, it should not matter whether the consonant or the vowel were made more salient through lengthening, and Experiment 2 demonstrates that lengthening of every third vowel was either neutral or actually unhelpful for segmentation when that vowel was in the word-initial syllable.

With regard to linguistically-motivated metrical segmentation, one would expect that initial-syllable lengthening, interpreted as a cue to stress, would serve selectively to promote segmentation for speakers of languages with word-initial stress. All content words in Hungarian (Fónagy, 1958) and most in English have stress on the first syllable (Cutler & Carter, 1987), whilst in Italian the predominant stress pattern is penultimate, with relatively few trisyllabic words having strong initial syllables (d'Imperio & Rosenthall, 1999; Rogers & d'Arcangeli, 2004). The pattern of results – onset consonant lengthening boosting segmentation, vowel lengthening in the word-initial syllable neutral or hindering segmentation – was however similar for all three languages. This strongly argues against a stress-based interpretation of such lengthening cues.
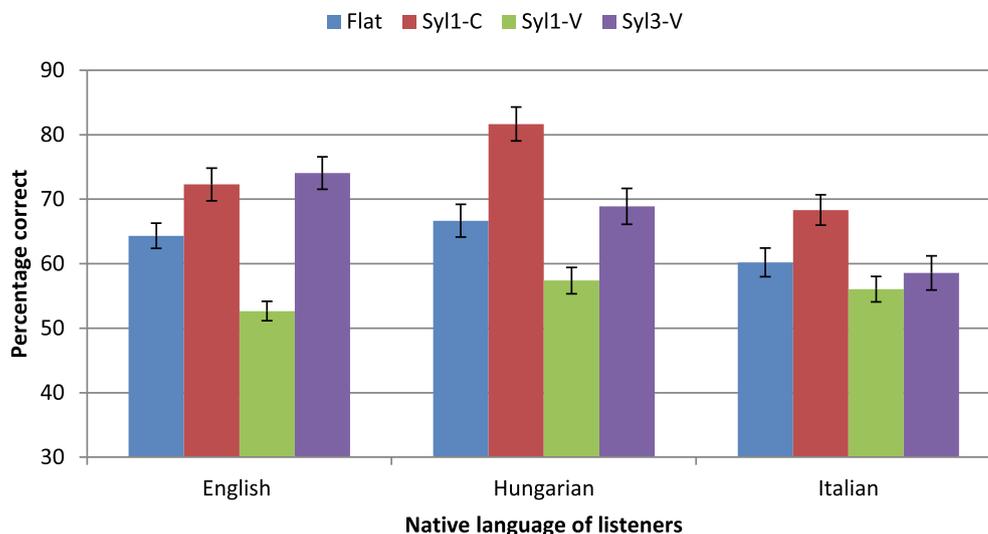
Fig. 1. Mean correct recognition of words in artificial language – comprised of four trisyllables: *nudopa, subako, bitusa, ripolu* – after six minutes' exposure. Timing conditions: **Flat**: even timing of all segments; **Syl1-C**: lengthening of consonant in first syllable (*nudopa*); **Syl1-V**: lengthening of vowel in first syllable (*nudopa*); **Syl3-V**: lengthening of vowel in third syllable (*nudopa*). Bars are one standard error each side of the mean.

Table 3
Mean correct word recognition scores and pairwise planned comparisons between the Experiment 2 Syl1-V timing condition and the Flat baseline (Experiment 1).

	Mean flat (%)	Mean Syl1-V (%)	$\chi^2(1)$	p	d
English	64	53	20.32	<0.0001	0.24
Hungarian	67	57	8.06	0.005	0.19
Italian	60	56	2.07	0.151	–

It is notable that the lengthening of the vowel in the first syllable was not ignored, but blocked the identification of the preceding statistically-defined word boundaries (partially for Hungarian speakers, marginally for Italian speakers, completely for English speakers). This suggests that our final hypothesis relating to Experiment 2 may be correct: if vowel lengthening is interpreted as a cue to an upcoming boundary, it will be unhelpful for segmentation when it occurs word-medially (i.e., between the first and the second syllable). For Italian speakers, Shukla, Nespor, and Mehler (2007) found that a combination of final lengthening and intonational cues, either native or non-native, effectively blocked learning of an artificial language where the statistically-defined words straddled the prosodic breaks. That provides negative evidence that word-final lengthening may be a segmentation cue for Italian speakers, although testing that hypothesis more directly, Ordin, Polyanskaya, Laka, & Nespor (2017) did not find that Italians exploited this cue in AL learning. We address this question directly by creating a fourth AL stream, this time with all word-final vowels lengthened.

4. Experiment 3

We tested whether speakers of English, Hungarian and Italian all exploit word-final lengthening to a comparable degree. In line with many previous findings (Saffran, 2002; Tyler & Cutler, 2009; Wightman et al., 1992), we strongly expected that English speakers will use final lengthening to infer an upcoming boundary. The status of this cue in Italian and Hungarian remains to be adjudicated however: the universal hypothesis (Tyler & Cutler, 2009) suggests that results for Italian and Hun-

garian should parallel the well-established findings for English. Alternatively, language-specific uses of vowel duration – for signalling stress in Italian and vowel length in Hungarian – may override the universal tendency for final lengthening in production. In that case, we expect to see an interaction between timing (specifically, the contrast between the final-lengthening and flat timing conditions) and the participants' native languages.

4.1. Method

We tested 40 new participants each for English, Hungarian and Italian. The procedure was as for Experiments 1 and 2, but here all participants heard an artificial language with vowel lengthening in the final syllable. Specifically, in this Syl3-V condition, the vowel of the final syllable (underlined in *bitusa*, *nudopa*, *ripolu*, *subako*), was 170 ms long, whilst all other segments were 110 ms long. Participants heard the six-minute AL stream thus created, followed by the 24 trial two-alternative forced choice test phase, as for Experiments 1 and 2.

4.2. Results

As in the previous experiments, cases where participants timed out on a single trial during the experiment were treated as missing data (Hungarian 2; Italian 4). There was one Italian and one English participant with multiple time-outs: these were replaced to retain 40 participants for each language.

Logistic regression showed that overall word recognition was above chance in the Syl3-V condition (word-final-vowel lengthening), $M = 67\%$, $z = 6.97$, $p < 0.0001$. This was higher than the

baseline Flat condition ($M = 64\%$), an effect of Timing which was statistically marginal, $\chi^2(1) = 3.39$, $p = 0.066$, $d = 0.07$. As shown in Fig. 1, the relative correct response scores in the Syl3-V lengthening condition and the Flat condition varied substantially according to Language. Indeed, there was both an effect of Language, $\chi^2(2) = 17.74$, $p = 0.0001$, and an interaction between Timing and Language, $\chi^2(2) = 6.24$, $p = 0.044$.

The effects of Timing are summarized in Table 4 for the individual languages. The marginal effect of Timing and the Timing/Language interaction are due to the fact that English speakers showed higher performance in the Syl3-V condition relative to the Flat condition, but for both Hungarian and Italian speakers, performance was similar in the two conditions. In line with these results, performance was statistically well above chance for all language groups (Table 2).

Comparing directly between these groups, English speakers' performance was better than that of the Italian speakers, $\chi^2(1) = 16.68$, $p < 0.0001$, $d = 0.33$, but similar to the Hungarian speakers; in line with this, Hungarian speakers outperformed the Italians, $\chi^2(1) = 7.16$, $p < 0.007$, $d = 0.22$.

Pooling the results of the three experiments, Fig. 1 shows the performance of all three language groups on all four timing conditions. Table 5 reports the statistical comparisons between the three non-baseline timing conditions for each language (the baseline comparisons having been reported above) and Table 6 shows the overall ranking of performance for all timing conditions.

4.3. Discussion

In contrast to fairly consistent patterns across English, Hungarian and Italian in Experiments 1 and 2, the responses to word-final vowel lengthening in Experiment 3 were highly variable between languages. Contrary to what has previously been suggested (Tyler & Cutler, 2009), but in line with recent findings for Italian and Spanish (Ordin, Polyanskaya, Laka, & Nespor, 2017), word-final lengthening does not appear to be interpreted universally as a cue to segmentation. Rather, how listeners interpret vowel lengthening with respect to speech structure appears to depend on the specific prosodic characteristics of their native language.

The results for English are congruent with existing production and perception data: large lengthening effects have been consistently reported phrase- and utterance-finally (reviewed in Fletcher, 2010) and lengthening of word-final vowels has been shown to be a cue to an upcoming boundary with similar consistency (e.g., Saffran et al., 1996; Tyler & Cutler, 2009).

With regard to Italian, as discussed by Ordin, Polyanskaya, Laka, & Nespor (2017), the results are compatible with the fact that word-final vowels in Italian are not lengthened phrase-medially, even when stressed (Bertinetto, 1980; Rogers & d'Arcangeli, 2004). Thus, when tracking word boundaries in artificial languages, vowel lengthening is not perceived as informative about structure, at least when incongruent with the predominant Italian pattern of stress on the penultimate syllable. Two caveats are worth mentioning in this regard. First, Shukla et al. (2007) found that a combination of final lengthening plus phrase-final intonation did encourage Italian speakers to perceive a prosodic break. Thus, in combination with another cue, the interpretation of timing appears to be different: it remains to be demonstrated whether the intonational cue is a sufficient indicator of a boundary for Italian speakers, or whether lengthening and intonation are required together. Secondly, Bion et al. (2011) showed that when Italians segment disyllabic words they actually interpret an alternating short-long-short-long pattern of syllables as having the longer syllables sequence-finally, suggesting that they do not simply ignore vowel-lengthening information during segmentation. This result is in line with the Iambic-Trochaic law, as discussed above, and it may be that alternating disyllabic sequences constitute a special case in this regard. The contrast between Bion et al. (2011) and our findings, also those of Ordin, Polyanskaya, Laka, & Nespor (2017), implies that the immediate alternation of short and long elements may facilitate segmentation due to general perceptual processes that override language-specific use of cues. Indeed, perception of longer elements in alternating sequences as being constituent-final has been shown for non-linguistic sequences (Bion et al., 2011; Daniela et al., 2013; Hay & Diehl, 2007) and even in the visual domain (Peña, Bion, & Nespor, 2011). Thus, it appears a broad perceptual principle may be overridden by language-specific prosodic characteristics, depending on the structure of the elements and the nature of the task.

For Hungarian, it is established that vowels lengthen in phrase-/utterance-final position (e.g., Hockey and Fagyal, 1999; White & Mády, 2008), so it seems initially surprising that final lengthening does not operate as a clear segmentation cue. In Hungarian, however, the smallest prosodic boundary is between accentual phrases, and this boundary requires the presence of post-boundary phrase accent (Beňuš, Reichel, & Mády, 2014). Thus, in keeping with the language-specific patterns, listeners may need to perceive a syllable as accented in order to make use of the lengthening in the preceding syllable as a boundary cue. In our materials, fundamental frequency does not vary, and so supporting intonational cues are absent. As with Italian, it seems likely that the combination of boundary intonation plus final lengthening is a strong cue for Hungarian listeners, but the timing cue on its own is ineffective. For Hungarian, this may be due to the linguistically significant use of vowel duration as a cue to phonemic length; this contrasts, for example, with French, which lacks a purely duration-based vowel length distinction and where listeners do interpret final-vowel lengthening as an upcoming boundary cue (Tyler & Cutler, 2009).

Thus, a weaker version of the hypothesis that vowel lengthening universally indicates an upcoming boundary may

Table 4

Mean correct word recognition scores and pairwise planned comparisons between the Experiment 3 Syl3-V timing condition and the Flat baseline (Experiment 1).

	Mean flat (%)	Mean Syl3-V (%)	$\chi^2(1)$	p	d
English	64	74	9.76	0.0017	0.21
Hungarian	67	69	0.46	0.499	–
Italian	60	59	0.19	0.661	–

Table 5
Pairwise planned comparisons between three non-baseline timing conditions for all three experiments.

	Syl1-C vs Syl1-V			Syl1-C vs Syl3-V			Syl1-V vs Syl3-V		
	$\chi^2(1)$	p	d	$\chi^2(1)$	p	d	$\chi^2(1)$	p	d
English	35.36	<0.0001	0.41	0.27	0.60	–	40.77	<0.0001	0.46
Hungarian	40.04	<0.0001	0.55	11.00	<0.001	0.30	10.99	<0.001	0.24
Italian	14.62	0.0001	0.25	7.16	0.007	0.20	0.66	0.41	–

Table 6
Ranking of performance for all timing conditions within each language group.

English				
Syl1-C = Syl3-V	>	Flat	>	Syl1-V
Hungarian				
Syl1-C	>	Flat = Syl3-V	>	Syl1-V
Italian				
Syl1-C	>	Flat = Syl3-V = Syl1-V		

be correct: in languages where vowel duration is a primary cue to some phonological distinction (stress, segmental identity), vowel lengthening only operates as a phrase boundary cue when accompanied by other sources of segmentation information, notably intonation.

5. General discussion

Reviewing the results language-by-language, the native English speakers showed the clearest and – based on foregoing findings – most predictable pattern of performance. Test recognition of the AL words was boosted both in the Syl1-C and the Syl3-V conditions compared to the Flat timing and Syl1-V conditions. Thus, word-initial consonant lengthening and word-final vowel lengthening functioned, when congruent with the statistics of the artificial language, as effective segmentation cues, as previously found (e.g., Saffran et al., 1996; White et al., 2015). For English speakers, the worst – chance level – performance was found in the Syl1-V condition, which reinforces the power of vowel lengthening as a cue to an upcoming boundary: where the vowel is lengthened medially in a statistically-defined word, English speakers are inhibited in their segmentation of the AL words (Tyler & Cutler, 2009; White et al., 2015).

For Hungarian speakers, performance on the Syl1-C condition was better than on all the others, clearly supporting word-initial consonant lengthening as a segmentation cue. Furthermore, the Syl1-V condition elicited the worst performance of all conditions. Together, these results rule out two alternative interpretations of the word-initial consonant lengthening boost to segmentation. Firstly, the effect is not due to a general metronomic pattern, a regular lengthening which could point to the repeated trisyllabic structure of the AL. Secondly, and language-specifically, this is not due to perception of word-initial (vowel or consonant) lengthening as a stress cue, which would accord with the fixed location of Hungarian stress. An explanation of the inhibitory effect of initial-syllable vowel lengthening for Hungarian could have been, as for English, that vowel lengthening is perceived as being a cue to an upcoming boundary; however, actual final vowel lengthening (Syl3-V) was no better than the Flat timing condition.

This mixed pattern of Hungarian results points to, firstly, a potentially universal use of initial consonant lengthening as a segmentation cue, underpinned by production and perception constraints, as discussed below. Beyond that, the language-specific use of vowel duration should be considered. As described above, duration is not used to mark Hungarian stress, except when the word is additionally phrase-accented (Fónagy, 1958; Szalontai, Wagner, Mády, & Windmann, 2016). Furthermore, because lexical stress is fixed, it is possible that it is ignored by Hungarian speakers as a source of structural information. Indeed, Peperkamp and Dupoux (2002) found evidence of "stress-deafness" in Hungarian speakers, who – like Finnish and French speakers – were relatively poor at discriminating minimal pairs (e.g., *míku – mikú*) purely on the basis of stress cues (see also Rahmani, Rietveld, & Gussenhoven, 2015, on the nature of "stress deafness").

In the light of all this evidence, it is perhaps unsurprising that initial-syllable vowel lengthening is not helpful for Hungarian speakers, but it remains unclear why it should be actually inhibitory for segmentation, given the lack of use of final vowel lengthening as a cue to an upcoming boundary. (We incidentally note that apparent "stress deafness" in French speakers is not incongruent with their use of final-vowel lengthening as a boundary cue, e.g., Tyler & Cutler, 2009.) One possibility that we tentatively suggest is that a longer Syl1 vowel could affect the perception of the preceding consonant, which would appear relatively short by contrast and thus less likely to be word-initial. Further work would be required to establish the validity of such a mechanism. There is evidence that the perceived presence of prosodic boundary-related lengthening can affect listeners' categorisation of durationally-cued phonemic distinctions (Kim, Mitterer, & Cho, 2018; Mitterer, Cho, & Kim, 2016; Steffman, 2019) and it seems plausible that interpretation of prosodic timing effects is likewise subject to local contextual modulation.

An apparently surprising result for Hungarian, given the robust evidence of preboundary lengthening vowel effects in production (Hockey and Fagyal, 1999; White & Mády, 2008), is that Syl3 vowel lengthening does not promote segmentation. Because vowel length is perceptually important for distinguishing vowel identity, it may be that it is not interpreted as a cue to structure in the absence of other cues, such as phrase-final intonation or a post-boundary accent, as considered above. (Additionally, Hungarian accentual phrases often end with an unstressed article before the accent that initiates a new AP.)

This interpretation appears somewhat at odds with the findings for consonant duration, which clearly serves as a cue to both consonant identity and the presence of a preceding boundary, but one reason for this difference may be the importance of word-initial consonants for lexical access (see below).

The pattern of results was somewhat similar for Italian speakers to that of Hungarians, but for Italian, vowel duration appeared to have no impact on listeners' perception of lexical structure. Performance in the Syl1-V and Syl3-V lengthening conditions was statistically equivalent to that in the Flat condition, where segments have equal duration. This is initially surprising given that phrase- and utterance-final vowel lengthening effects are clearly observed in accents of Italian (e.g., D'Imperio & Fivela, 2003; White, Payne, & Mattys, 2009). For Syl3 vowel lengthening, however, our findings accord with the perceptual results of Ordin, Polyanskaya, Laka, & Nespor (2017), who used different artificial languages and a French synthetic voice: in both their study and ours, no recognition difference was found for AL words between the Flat timing and Syl3-V lengthening conditions.

Thus, Italian listeners (recruited and tested in north-eastern Italy for both the present study and Ordin, Polyanskaya, Laka, & Nespor, 2017), do not appear to use final lengthening for segmentation in the absence of other prosodic cues. Unlike the Hungarian speakers, however, the explanation for Italian may well relate to the use of duration as a strong cue to lexical stress in the predominant penultimate position. As discussed above, this is the only stress position in which vowels are strongly marked by lengthening: word-final stressed vowels are relatively short (Bertinetto, 1980; Rogers & d'Arcangeli, 2004), at least intonation-phrase-medially. Thus, when presented with a sequence of trisyllables, it may be that the only pattern of vowel lengthening congruent with Italian listeners' expectations would be on the penultimate syllable, a condition found to boost segmentation by Ordin, Polyanskaya, Laka, & Nespor (2017). Unlike for English and Hungarian listeners, duration is the primary cue to Italian stress (Bertinetto, 1980), with stress itself being lexically contrastive to a much more productive degree than English (Hungarian having no lexically-contrastive stress). Given that, it is perhaps unsurprising that Italian listeners do not also interpret lengthening as a pre-boundary cue, at least in the absence of intonational support.

It should be noted that, for the current study, Italian performance on the Flat timing condition was numerically but not statistically better than the Syl1-V lengthening conditions; for Ordin, Polyanskaya, Laka, & Nespor 2017, the difference was statistically robust. Given the disparity, we do not offer a strong interpretation of this finding: if Syl1 vowel lengthening does actually inhibit segmentation, it may be related to a comparative timing mechanism, as discussed above for Hungarian, but this is speculative and requires further empirical investigation. It may also be that vowel lengthening was less conspicuous in our between-subjects design than in Ordin et al.'s repeated measures experiment, wherein subjects were exposed to the artificial languages in all timing conditions, potentially providing a point of reference and contrast as regard vowel duration.

Overall, as for English and Hungarian, results for Italian clearly support word-initial consonant lengthening as strong segmentation cue, discussed further below. Clearly, more experiments with a range of prosodically diverse languages are required to determine the generalizability of these findings

5.1. Is word-initial consonant lengthening a universal segmentation cue?

Speech comprehension requires the listener to determine the structure of the speech stream ("where") and identify the words therein ("what"). Models of speech perception typically characterise word recognition as a competition between lexical candidates on the basis of their compatibility with incoming segments, whilst prosody provides cues to the structure of the stream (e.g., Cho, McQueen, & Cox, 2007). In clear listening conditions with appropriate contextual information available, explicit structural analysis may be bypassed at the lexical level: thus, potential non-lexical cues to word segmentation are overlooked because competition provides a lexical solution that is sufficiently congruent with the segmental input (Mattys, White, & Melhorn, 2005). However, in certain speech processing contexts – difficult listening conditions, cases of lexical ambiguity, learning of novel input – prosodic and other non-lexical cues to structure become more important for listeners, restricting the set of lexical candidates under consideration for recognition (Mattys et al., 2005).

Constituent onsets are particularly important for both word segmentation and recognition, and as discussed above, the hypothesized Syllable Onset Segmentation Heuristic prioritizes identification of potential word onsets in speech recognition (Content et al., 2001).

Furthermore, in established phonological theory, the unmarked syllable structure is regarded as being CV, with a consonantal onset (e.g., the Obligatory Onset Principle) and a vocalic rhyme. Indeed, glottalization of onsetless syllables can be seen as device to include a consonantal gesture before the vowel (Hooper, 1972; McCarthy & Prince, 1994), with more likelihood of glottalization after higher-level boundaries (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996); moreover, glottalization is interpreted as a cue to the preceding boundary (e.g., Newman, Sawusch, & Wunnenberg, 2011), potentially analogous to consonant lengthening. Indeed, infants have a preference for filled onsets, with a relative delay in segmentation of vowel-initial compared with consonant-initial words (Jusczyk & Aslin, 1995; Mattys & Jusczyk, 2001). Thus, listeners have a perceptual bias for paying attention to syllable-onset consonants as possible starts of new lexical items.

As discussed, two parallel and related speech processing goals are associated with the articulation of onset consonants. Word recognition, specifically the initiation of new lexical hypotheses, is boosted by gestural strengthening, whilst segmentation is boosted by lengthening, as shown in the current study. Furthermore, the widespread and possibly universal occurrence of these effects in speech production, and their interpretation in speech perception, may be motivated by "perceptuomotor analogies".

As discussed by Dingemanse et al. (2015), perceptuomotor analogies allow listeners to interpret speech sounds by reference to the process of their production. These analogies may be seconded for signalling not only lexical meaning, but also speech structure. For example, the Respiration Code (formerly the "Production Code", Gussenhoven, 2002, 2016) relies

on the positive association between subglottal air pressure and pitch: because pressure is higher at the start of breath groups than at the end, higher pitch tends to indicate onset of utterances, whilst lower pitch is associated with finality. This declination trend arises from intrinsic articulatory processes, but has become formalized cross-linguistically as a cue to structure (Gussenhoven, 2002, 2016). It is similarly plausible that perceptuomotor analogies may condition listeners' interpretation of speech timing cues: for example, the minimum jerk principle (as discussed above) would predict that onsets and offsets should be slower (hence longer) than equivalent constituent-medial units. This tendency, arising as it does from a very general principle of primate movement, would clearly be subject to linguistically-derived modulation: for example, in English, whilst word-initial lengthening is restricted to the onset consonant, phrase-final lengthening may begin with the final stressed vowel and include subsequent unstressed syllables (White, 2014, for a review).

By default, gestural lengthening and strengthening tend to co-occur, more extreme gestures tending to take longer to achieve, although they are not necessarily correlated (Fougeron & Keating, 1997). As further discussed by Fougeron and Keating, hyperarticulation of word-initial consonantal gestures potentially provides a lexical access boost for listeners, by reducing coarticulatory influences and making segmental identity maximally clear. This would naturally be more useful earlier in utterances, when context has yet to allow listeners to generate top-down expectations about lexical identity: as highlighted earlier, a corollary of this would be supralaryngeal declination (e.g., Hinton, 1996), mediated by syntactic/prosodic structure.

That the listener's requirement for segmentation cues (lengthening) and lexical access (strengthening) are somewhat separable is evident at utterance-initial boundaries. It is frequently observed that segments are shorter utterance-initially than at the start of other prosodic domains, a fact that White (2002, 2014) suggests may be related to the obvious lack of need for an overt segmentation cue at the onset of articulation: whatever follows a (non-hesitation) pause must be the start of a new word. Independent of durational patterns, however, articulatory strengthening is observed utterance-initially (Fougeron & Keating, 1997): manifestly, the lexical access requirement for maximally helpful information about word identity pertains regardless of whether the presence of the boundary has been made clear or not. Incidentally, this functional interpretation implies that differences may be found between consonants in whether they are relatively short utterance-initially: those consonants for which identification is substantially boosted by lengthening, independent of gestural strengthening, would be longer. By contrast, consonants for which gestural strengthening should suffice to boost identification may be shortened when the segmentation function of lengthening has been subserved by the preceding pause.

5.2. Why is final vowel lengthening Fletcher, 2010; White, Payne, & Mattys, 2009 a language-specific segmentation cue?

In many languages, word-final vowel lengthening is interpreted as a cue to an upcoming boundary (e.g., Tyler & Cutler, 2009). Indeed, White et al. (2015) suggested that final vowel lengthening and onset consonant lengthening together may be interpreted by listeners through a delay in the occurrence of the perceptual centre (P-Centre, Morton, Marcus, & Frankish, 1976) of the upcoming syllable. This synergistic exploitation of cues would accord with the broader principle that the unmarked syllable structure is CV, as discussed above; thus, consonants tend to be identified with onsets and vowels with offsets.

The evidence from this study and from Ordin, Polyanskaya, Laka, & Nespor, 2017 strongly suggests, however, that the interpretation of word-final vowel lengthening as a cue to an upcoming boundary is not universal, at least in the absence of supporting cues such as intonation. The selectivity of the cue appears to arise because languages exploit vowel duration to signal a range of contrasts and gradient effects. As discussed, Italian relies on duration as the primary cue to stress, itself an important aspect of lexical identity in the Italian vocabulary. Although durational variation can be interpreted to multiple ends in the same language (e.g., consonant duration signalling both quality and structure in Italian, Tagliapietra & McQueen, 2010), a heavy functional load on one feature, such as vowel duration, may mean that the segmentation cue is only perceived by listeners when accompanied by additional cues or, potentially, when the cue is unambiguously related to structure through contextual support.

6. Conclusion

As adults listening to a familiar language in good listening conditions, we tend to rely on language-specific knowledge, particularly of words and syntactic structures, to infer the structure of the speech stream; we fall back on segmental-acoustic and prosodic cues where noise, ambiguity and other challenges to segmentation arise (Mattys et al., 2005). Prosody has long been identified, however, as having a central role in early vocabulary development (the "bootstrapping" hypothesis, e.g., Cutler, 1996). Indeed, infants have been show to use domain-edge lengthening effects as structural cues early in language acquisition (e.g., Gout, Christophe, & Morgan, 2004). It may be that the early interpretation of final vowel lengthening as a cue to an upcoming boundary is universal and only later loses its power in certain languages, such as Italian and Hungarian, where vowel duration is used as a primary cue to other aspects of meaning. We propose, however, that initial consonant lengthening may be likely to maintain a universal structural function because of the critical importance of word onsets for the entwined processes of speech segmentation and word recognition.

References

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*(6), 644–663.

Beňuš, Š., Reichel, U. D., & Mády, K. (2014). Modeling accentual phrase intonation in Slovak and Hungarian. In L. Veselovská & M. Janebová (Eds.), Complex Visibles Out There. Proceedings of the Olomouc Linguistics Colloquium 2014. (pp. 677–689). Palacký University.

Berkovits, R. (1994). Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech, 37*(3), 237–250.

Bertinetto, P. M. (1980). The perception of stress by Italian speakers. *Journal of Phonetics, 8*(4), 385–395.

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America, 134*(5), 3828–3843.

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54*(1), 123–140. https://doi.org/10.1177/0023830910388018.

Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics, 64*, 71–89. https://doi.org/10.1016/j.wocn.2016.12.003.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35*(2), 210–243.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language, 45*(2), 177–199.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and Speech*. Harvard University Press.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201–220.

Cutler, A. (1990). From performance to phonology: Comments on Beckman and Edwards's paper. In J. Kingston & M. E. Beckman (Eds.). *Between the Grammar and Physics of Speech* (Vol. 1, pp. 208–214). Cambridge University Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3), 133–142.

Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*(2), 171–195.

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 87–99). Lawrence Erlbaum Associates.

D'Imperio, M., & Fivela, B. G. (2003). How many levels of phrasing? Evidence from two varieties of Italian. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation, papers in laboratory phonology VI* (pp. 130–144). Cambridge University Press.

d'Imperio, M., & Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology, 16*(01), 1–28.

Daniela, M., Nespor, M., & Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic–trochaic law? *Attention, Perception, & Psychophysics, 75*(1), 92–100.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*(4), 423–444.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences, 19*(10), 603–615.

Elordieta, G., & Hualde, J. I. (2014). Intonation in Basque. In S.-A. Jun (Ed.), Prosodic Typology II (pp. 405–464).

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceedings of the Fourth International Conference on Spoken Language Processing, 3*, 1393–1396.

Farnetani, E., & Vayra, M. (1996). The role of prosody in the shaping of articulation in Italian CV syllables. In P Perrier, F Bustarret, & R Laboissière (Eds.), *Proceedings of the 1st ESCA Tutorial and Research Workshop on Speech Production Modeling and 4th Speech Production Seminar* (pp. 9–12). Institut National Polytechnique de Grenoble, France.

Féry, C., Hörnig, R., & Pahaut, S. (2011). Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. In C. Gabriel & C. Lleó (Eds.), *International phrasing in Romance and Germanic: Cross-linguistic and Bilingual Studies* (pp. 11–41). John Benjamins.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience, 5*(7), 1688–1703.

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), The Handbook of Phonetic Sciences, Second Edition (pp. 521–602). Hoboken, NJ: Wiley-Blackwell.

Fónagy, I. (1958). A hangsúlyról [On stress]. *Nyelvtudományi Értekezések, 18*, 1–76.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America, 101*(6), 3728–3740.

Fourakis, M., & Monahan, C. B. (1988). Effects of metrical foot structure on syllable timing. *Language and Speech, 31*(3), 283–306.

Frost, R. L. A., Monaghan, P. J., & Tatsumi, T. (2016). Domain-general mechanisms for speech segmentation: The role of duration information in language learning. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 466–476.

Giegerich, H. J. (1992). *English Phonology: An Introduction*. Cambridge: Cambridge University Press.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language, 51*(4), 548–567. https://doi.org/10.1016/j.jml.2004.07.002.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 344–359.

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science, 8*(2), 425–434. https://doi.org/10.1111/tops.12197.

Gussenhoven, C., & Rietveld, A. C. M. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics, 20*(3), 283–303.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard, 4*(s2). https://doi.org/10.1515/lingvan-2017-0027.

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Attention, Perception, & Psychophysics, 69*(1), 113–122.

Hinton, V. A. (1996). Interlabial pressure during production of bilabial phones. *Journal of Phonetics, 24*(3), 337–349.

Hockey, B. A., & Fagyal, Z. (1999). Phonemic length and pre-boundary lengthening: An experimental investigation on the use of durational cues in Hungarian. 14th International Congress of Phonetic Sciences, Berkeley, University of California, 313–316.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *Journal of Neuroscience, 4*(11), 2745–2754.

Hooper, J. B. (1972). The syllable in phonological theory. *Language, 48*(3), 525–540.

Houlihan, K. (1975). *The role of word boundary in phonological processes [PhD]*. University of Texas at Austin.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *1st International Conference on Speech Prosody* (pp. 7–57).

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124*(4), 2263–2271. https://doi.org/10.1121/1.2973189.

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

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 143–161). Cambridge University Press.

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition, 34*(03), 415–444. https://doi.org/10.1017/S0272263112000137.

Kim, S., Cho, T., & McQueen, J. M. (2012). Phonetic richness can outweigh prosodically-driven phonological knowledge when learning words in an artificial language. *Journal of Phonetics, 40*(3), 443–452. https://doi.org/10.1016/j.wocn.2012.02.005.

Kim, S., Mitterer, H., & Cho, T. (2018). A time course of prosodic modulation in phonological inferencing: The case of Korean post-obstruent tensing. *PLOS ONE, 13*(8). https://doi.org/10.1371/journal.pone.0202912 e0202912.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3*(3), 129–140.

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

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology, 45*(1), 73–87.

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 644–655. https://doi.org/10.1037/0096-1523.27.3.644.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477.

McCarthy, J. J., & Prince, A. S. (1994). The emergence of the unmarked: Optimality in prosodic morphology. Proceedings of the North East Linguistic Society, 24, 333–379.

Mitterer, H., Cho, T., & Kim, S. (2016). How does prosody influence speech categorization? *Journal of Phonetics, 54*, 68–79. https://doi.org/10.1016/j.wocn.2015.09.002.

Molnar, M., Carreiras, M., & Gervain, J. (2016). Language dominance shapes non-linguistic rhythmic grouping in bilinguals. *Cognition, 152*, 150–159. https://doi.org/10.1016/j.cognition.2016.03.023.

Moon, S., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America, 96*(1), 40–55. https://doi.org/10.1121/1.410492.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review, 83*(5), 405–408.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics, 46*(2), 135–147.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. Lingue e Linguaggio, 7(2), 139–168. https://doi.org/10. 1418/28093

Nespor, M., & Vogel, I. (2007). *Prosodic Phonology*. Berlin: Walter de Gruyter.

Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory and Language, 64*(4), 460–476. https://doi.org/10.1016/j.jml.2010.11.004.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America, 54*(5), 1235–1247.

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition, 45*, 863–876. https://doi.org/10.3758/s13421-017-0700-9.

Peña, M., Bion, R. A. H., & Nespor, M. (2011). How modality specific is the iambic–trochaic law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1199–1208. https://doi.org/10.1037/a0023944.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (Eds.), Laboratory phonology (Vol. 7, pp. 203–240). Mouton de Gruyter.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America, 90*(6), 2956–2970.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *The Journal of the Acoustical Society of America, 94*(4), 2027–2035.

Rahmani, H., Rietveld, T., & Gussenhoven, C. (2015). Stress "deafness" reveals absence of lexical marking of stress or tone in the adult grammar. *PLOS ONE, 10* (12). https://doi.org/10.1371/journal.pone.0143968 e0143968.

Rogers, D., & d'Arcangeli, L. (2004). Italian. *Journal of the International Phonetic Association, 34*(1), 117–121. https://doi.org/10.1017/S0025100304001628.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*(1), 172–196. https://doi.org/10.1006/jmla.2001.2839.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606–621.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*(1), 1–16.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54*(1), 1–32. https://doi.org/10.1016/j.cogpsych.2006.04.002.

Steffman, J. (2019). Intonational structure mediates speech rate normalization in the perception of segmental categories. *Journal of Phonetics, 74*, 114–129. https://doi.org/10.1016/j.wocn.2019.03.002.

Szalontai, Á., Wagner, P., Mády, K., & Windmann, A. (2016). Teasing apart lexical stress and sentence accent in Hungarian and German. *Proc. Phonetik Und Phonologie Im Deutschsprachigen Raum*, 215–218.

Szende, T. (1994). Hungarian. *Journal of the International Phonetic Association, 24*(2), 91–94. https://doi.org/10.1017/S0025100300005090.

Tagliapietra, L., & McQueen, J. M. (2010). What and where in speech recognition: Geminates and singletons in spoken Italian. *Journal of Memory and Language, 63* (3), 306–323. https://doi.org/10.1016/j.jml.2010.05.001.

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

Vaissière, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 53–66). Springer.

Vayra, M., & Fowler, C. A. (1992). Declination of supralaryngeal gestures in spoken Italian. *Phonetica, 49*(1), 48–60. https://doi.org/10.1159/000261902.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication, 63–64*, 38–54. https://doi.org/10.1016/j.specom.2014.04.003.

White, L., Mattys, S. L., Stefansdottir, L., & Jones, V. (2015). Beating the bounds: Localized timing cues to word segmentation. *The Journal of the Acoustical Society of America, 138*(2), 1214–1220. https://doi.org/10.1121/1.4927409.

White, L., & Mády, K. (2008). The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. 4th Speech Prosody Conference, Campinas, Brasil, 363–366.

White, L. (2002). *English Speech Timing: A Domain and Locus Approach*. [University of Edinburgh] http://www.cstr.ed.ac.uk/projects/eustace/fulltext/lswdiss.pdf.

White, L., Payne, E., & Mattys, S. L. (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigário, S. Frota, & M. J. Freitas (Eds.), *Phonetics and Phonology: Interactions and Interrelations* (pp. 137–158). Amsterdam: John Benjamins.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America, 91*(3), 1707–1717.

Woodrow, H. (1909). A quantitative study of rhythm: The effect of variations in intensity, rate and duration. *Archives of Psychology, 14*, 1–66.