Psychon Bull Rev (2015) 22:815–823
DOI 10.3758/s13423-014-0733-z

BRIEF REPORT

Distal prosody affects learning of novel words in an artificial language

Tuuli H. Morrill · J. Devin McAuley · Laura C. Dilley · Patrycja A. Zdziarska · Katherine B. Jones · Lisa D. Sanders

Published online: 23 September 2014

Abstract The distal prosodic patterning established at the beginning of an utterance has been shown to influence downstream word segmentation and lexical access. In this study, we investigated whether distal prosody also affects word learning in a novel (artificial) language. Listeners were exposed to syllable sequences in which the embedded words were either congruent or incongruent with the distal prosody of a carrier phrase. Local segmentation cues, including the transitional probabilities between syllables, were held constant. During a test phase, listeners rated the items as either words or nonwords. Consistent with the perceptual grouping of syllables being predicted by distal prosody, congruent items were more likely to be judged as words than were incongruent items. The results provide the first evidence that perceptual grouping affects word learning in an unknown language, demonstrating that distal prosodic effects may be independent of lexical or other language-specific knowledge.

Keywords Speech perception · Language acquisition · Word segmentation · Distal prosody

T. H. Morrill (✉)
Program in Linguistics, Department of English, George Mason University, 4400 University Drive, 3E4, Fairfax, VA 22030, USA
e-mail: tmorrill@gmu.edu

J. D. McAuley · P. A. Zdziarska · K. B. Jones
Department of Psychology, Michigan State University, East Lansing, MI, USA

L. C. Dilley
Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA

L. D. Sanders
Department of Psychology, University of Massachusetts, Amherst, MA, USA

Introduction

To learn a spoken language, listeners must be able to map portions of a continuous acoustic speech signal onto meaningful units (e.g., words). Because spoken language does not contain reliable pauses between words, infants acquiring their native language and adults learning a second language must use other cues to locate word boundaries. Two types of information that listeners use to segment words are statistical cues, such as the transitional probabilities between syllables or phonemes, and prosodic patterns, such as intonational phrasing and lexical stress.

Transitional probabilities can serve as a segmentation cue, since the transitional probabilities of units *within* a word tend to be larger than those *spanning* a word (Brent & Cartwright, 1996; Swingley, 2005). Both infants and adults can use this type of statistical information to extract words and learn phonotactic patterns from artificial language speech streams (e.g., Hay, Pelucchi, Estes, & Saffran, 2012; Newport & Aslin, 2004; Saffran, Newport, & Aslin, 1996). Statistical learning appears to be a very general cognitive mechanism that is not language-specific or restricted to language learning (Fiser & Aslin, 2001; Saffran, Johnson, Aslin, & Newport, 1999).

Prosodic word segmentation cues include cues that signal phrase boundaries, such as intonation contours and phrase-final lengthening, since phrase boundaries necessarily signal word boundaries. Prosodic cues also include word-level information, such as stressed syllables, in certain languages (see, e.g., Christophe, Gout, Peperkamp, & Morgan, 2003; Cutler & Butterfield, 1992; Jusczyk, Cutler, & Redanz, 1993). For example, in English, stressed syllables often signal the beginning of a word (Cutler & Carter, 1987), and both infants and adults use this information to recognize word boundaries or possible words (e.g., Cutler & Butterfield, 1992; Jusczyk et al., 1993; Mattys, Jusczyk, Luce, & Morgan, 1999).

Previous work investigating prosodic cues to word segmentation has focused primarily on local cues occurring at or adjacent to a word boundary; however, prosodic temporal and pitch patterning in the distal (nonlocal) context has recently been shown to affect segmentation of known lexical items. Dilley and McAuley (2008) proposed a perceptual grouping hypothesis, whereby the prosodic pattern of the beginning of an utterance was predicted to influence the grouping of later syllables into words. For example, when a repeating binary (e.g., high–low–high–low) pitch pattern occurs in the distal context preceding a syllable sequence containing ambiguous word boundaries (e.g., the sequence "*ti-mer-der-by*" can be "*tie murder bee*" or "*timer derby*"), listeners' perceptions of word boundaries depend on the distal prosodic context (Dilley, Mattys, & Vinke, 2010; Dilley & McAuley, 2008). When continuation of the grouping perceived in the distal prosodic context predicts that the final syllable will be isolated, listeners report more monosyllabic final words (e.g., *bee*) than when continuation of the grouping perceived in the distal prosodic context predicts that the final two syllables will be grouped (e.g., *derby*). These findings support the view that a general auditory perceptual grouping principle plays an important role in speech perception.

Recent research showing effects of distal prosody on word segmentation has raised the question of whether or not perceptual grouping induced by distal context might also play a role in language acquisition. If so, listeners should be able to use distal prosodic cues to segment speech for which they do not have previously stored lexical representations. In general, investigations pertaining to the use of prosodic cues in the segmentation of an unknown language have yielded varied results (e.g., Adams, 2010; Akker & Cutler, 2003; Creel, Tanenhaus, & Aslin, 2006; Cutler, Mehler, Norris, & Segui, 1992; Dupoux, Pallier, Sebastian, & Mehler, 1997; Goetry & Kolinsky, 2000; Sanders & Neville, 2002, 2003). Some studies employing artificial-language paradigms have suggested that some possibly universal indices of prosodic structure, such as intonation contours or pauses, can affect word learning (e.g., E. K. Johnson & Seidl, 2008; Langus, Marchetto, Hoffmann Bion, & Nespor, 2012; Shukla, Nespor, & Mehler, 2007; Shukla, White, & Aslin, 2011). By applying intonational phrase contours to a speech stream, Shukla et al. (2007) showed that listeners were less accurate at identifying "words" (syllable sequences with high transitional probabilities) that straddled a prosodic boundary than those that were not interrupted by a prosodic boundary. In these studies, boundaries were marked by pitch patterns and duration cues, and further studies demonstrating this effect have also used overt pitch and/or duration cues to signal the presence of prosodic boundaries (e.g., E. K. Johnson & Seidl, 2008; Langus et al., 2012; Shukla et al., 2007; Shukla et al., 2011). However, in addition to local cues to prosodic boundaries, it is possible that distal prosodic cues may be used for the identification of candidate words in a novel language.

To investigate a potential role for distal prosodic cues in language learning, we considered the possibility that perceptual grouping induced by distal prosodic context might influence which words individuals learn in an unknown language. We used an artificial-language paradigm, in order to control for the segmental, syllabic, and prosodic characteristics of the language. Participants were exposed to syllable sequences that contained embedded disyllabic words. The distal prosody of the initial portion of the "phrases" was manipulated, while acoustic characteristics and the transitional probabilities between syllables of words were held constant. After exposure, listeners judged whether disyllabic test items were words in the language. The test items were sequences that had been heard with a pitch pattern that was either aligned (i.e., congruent) with the distal prosody, incongruent with the distal prosody, or were non-words, syllable sequences that had not occurred during exposure. According to a perceptual-grouping hypothesis, listeners should use distal prosodic cues to extract words of the language and should be more likely to judge as words disyllabic sequences that are congruent with distal prosody during exposure, compared to disyllabic sequences that are incongruent with distal prosody. This would indicate that, during exposure, listeners were grouping together co-occurring syllables and identifying them as possible words according to expectations based on the preceding prosodic patterning, despite a lack of any overt segmentation cues at the word boundaries themselves. If, on the other hand, distal prosodic cues are not powerful enough to influence word segmentation and learning, we should find no difference in word judgments for congruent and incongruent test items.

Experiment 1

Method

Participants Thirty nine native English speaking undergraduate students (26 female, 13 male; 18–22 years, M = 19, SD = 1.6) with self-reported normal hearing participated in the experiment for partial course credit. The data from two additional participants were eliminated from the final sample—one for a self-described cognitive deficit, and another for failure to follow instructions. The participants varied in years of music training (M = 2.3, SD = 3.0).

Stimuli The stimuli were constructed from 48 syllables. All of the syllables had a CV or CVC structure derived from the consonants [p, b, t, d, k, g, m, f, v, s, z,] and the vowels [i, ɛ, , u,]. Syllables were recorded by an English-speaking adult female. The final tokens were selected from multiple recordings in order to achieve the most isochronous distribution of syllable durations, and the syllables were concatenated into phrases using Praat (Boersma & Weenink, 2012). Previous studies employing similar stimulus creation methods (with naturally produced syllables) have successfully elicited evidence of word learning (e.g., Toro, Pons, Hoffmann Bion, & Sebastián-Gallés, 2011). Since our syllables had been recorded in isolation, all concatenated phrase durations were multiplied by a factor of 0.8 using PSOLA resynthesis (i.e., to 80 % of their original durations), for a more natural, fluent rate. The final syllable durations, including initial stop closures, averaged 392 ms (*SD* = 54). Intensity was normalized to 70 dB.

Twelve critical syllables were selected to form 12 disyllabic words in two artificial languages (see Table 1). Note that the words in both languages consisted of the same syllable pairs, but in opposite orders. Participants were randomly assigned to Language 1 (n = 20) or Language 2 (n = 19).

During the exposure phase, words were presented within critical-syllable sequences of four syllables; each critical syllable occurred in each position within the sequence, yielding 12 unique critical-syllable sequences (see Tables 1 and 2). The transitional probabilities within each disyllabic word were .75, and those for nonwords were 0.0.

Each critical-syllable sequence within a language was appended to each of six 5-syllable carrier sequences (Table 3); on half of the trials, an additional phrase-final syllable occurred (all of the phrases were either nine or ten syllables) so that backward grouping of phrase-final syllables would not be a reliable cue to word boundaries. This yielded a total of 144 (12 *critical-syllable sequences* × 6 *carrier sequences* × 2 *presence or absence of phrase-final syllable*) stimulus phrases. None of the syllables from the carrier sequences or the phrase-final syllables occurred within the critical-syllable sequences.

Distal prosody was manipulated by varying the pitch and timing patterns of the five-syllable carrier sequences to elicit perceptual grouping of the syllables (Fig. 1); the prosodic patterns of the critical-syllable sequences remained constant in both of the distal prosodic manipulation conditions. Repeated patterns of high (H) and low (L) tones (either H–L or L–H groupings) were applied to the five-syllable carrier sequences (the H and L tones were approximately 245 and 225 Hz, respectively). Participants heard both tonal patterns. For L–H groupings, each syllable was assigned one tone, for an L–H–L–H–L pattern; the durations of the syllables were unaltered. For H–L groupings, the first four syllables bore an H–L–H–L pattern, and the fifth syllable included two tones for its own H–L grouping; the fifth syllable was lengthened by a factor of 1.8 to accommodate the two tones.

Table 1 Words for Languages 1 and 2

Language 1 Words		Language 2 Words	
bɑ di	kɑ gu	dibɑ	gukɑ
butɛ	pikɔ	tɛbu	kɔpi
dutɑ	gɔ pɛ	tɑdu	pɛgɔ
dikɑ	gubɑ	kɑdi	bɑgu
tɛ pi	kɔ bu	pitɛ	bukɔ
tɑ gɔ	pɛ du	gɔtɑ	dupɛ

Each row makes up a critical-syllable sequence, which was rotated as presented in Table 2

Table 2 Example of rotation of syllables forming critical-syllable sequences

Order 1	**tɑ**	**gɔ**	pɛ	du
Order 2	du	**tɑ**	**gɔ**	pɛ
Order 3	pɛ	du	**tɑ**	**gɔ**
Order 4	**gɔ**	pɛ	Du	**tɑ**

Each word is heard in every position; e.g. see the word "*tɑgɔ*"

The disyllabic words in the critical-syllable sequence (which always had an H–L–H–L tonal pattern) were either congruent or incongruent with the distal prosody (Fig. 1). For each participant, half of the words were always congruent with the predicted perceptual grouping based on distal context, and half were incongruent with the predicted grouping. For example, when *tag pɛdu* was presented after carrier sequences with H–L groupings, the words *tag* and *pɛdu* were congruent with the distal prosody (i.e., aligned with the H–L grouping), whereas *g pɛ* was incongruent (since it contained an L–H pattern). When the critical-syllable sequence *tag pɛdu* was presented after carrier sequences with L–H groupings, the word *g pɛ* (also L–H) was congruent, whereas *tag* and *pɛdu* were incongruent. Crucially, the acoustics and transitional probabilities of the syllables within the critical-syllable sequences were identical in both contexts (and always had an H–L–H–L tonal pattern). In addition, for any given participant, each congruent word occurred with both distal prosodic pitch patterns—that is, congruent L–H and H–L contexts. Each word was always either congruent or incongruent for a given listener, but the congruent and incongruent words were counterbalanced, so that across participants each word was heard in both congruent and incongruent contexts. Because the critical-syllable sequences were identical across stimuli, any differences in how listeners responded to congruent and incongruent words in the test could only be attributed to the distal prosodic manipulation in the stimulus phrases.

Table 3 Carrier sequences and corresponding phrase-final 10th syllables (n = 6)

1) dɔ sɑ b ku pɑ zi ________	(fɛ)
2) ti gɑb vu kɔs dɑ ________	(gɑ)
3) pɔ dɑɹ gi bɛm tu ________	(ki)
4) bɛ dɑn vi gɔb tɔ ________	(ɹɑ)
5) fu gɑɹ zɔ bɑk dɛ ________	(mu)
6) tɑk pu vɑm si bɔ ________	(vɔ)

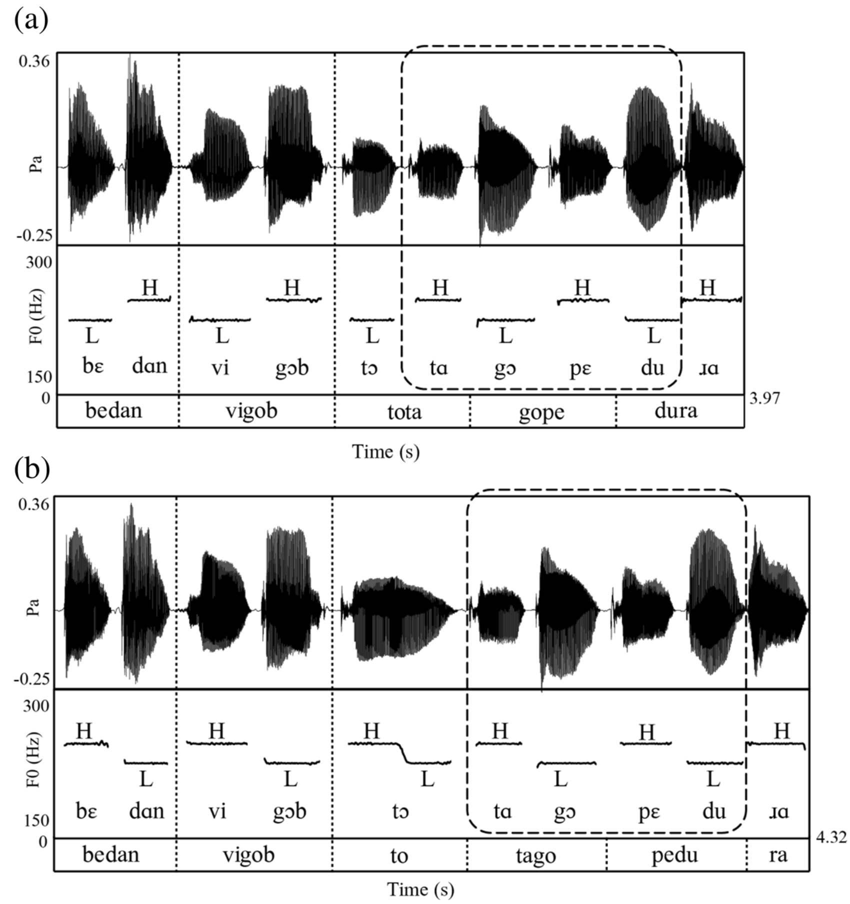

Fig. 1 Example stimuli illustrating distal prosody manipulations. (**a**) Low–High distal prosodic pitch pattern in the carrier sequence; in this distal prosodic context, the word [g pɛ] is congruent with the distal prosody, whereas [tag] and [pɛdu] are incongruent with the distal prosody. (**b**) High–Low distal prosodic pitch pattern in the carrier sequence, with fifth-syllable lengthening; in this distal prosodic context, the words [tag] and [pɛdu] are congruent with the distal prosody, whereas the word [g pɛ] is incongruent with the distal prosody. For any given participant for whom [g pɛ] was a congruent word, [g pɛ] would occur as congruent in both types of distal prosodic pitch patterns—the Low-High and High-Low contexts

Apparatus The experiment was run on a Dell PC with E-Prime 2.0.8.22 software (Psychology Software Tools, Inc.). All stimuli were presented over Sennheiser HD-280 Pro headphones (Old Lyme, CT) at a comfortable listening level.

Procedure During the exposure phase, participants listened to two blocks of 180 trials. Within each block, they heard the 144 phrases described above and 36 filler phrases that included a repeated pair of syllables. Fillers were created by combining the carrier sequences with other four- and five-syllable sequences in which a pair of syllables from the carrier was repeated (for 12 unique fillers). Fillers were used to hold attention. On each trial, participants were asked to listen carefully and to indicate whether they heard a repeated pair of syllables, using a response box. Trials were presented in pseudorandom order, and every ten trials included two fillers. Across both exposure blocks, participants heard each word 72 times.

During the test, participants heard 24 disyllabic test items; half were disyllabic words from the exposure phase, and half consisted of nonwords. Listeners judged whether each item was a "word" or "nonword" using a six-point scale (1 = *definitely nonword*, 2 = *likely nonword*, 3 = *maybe nonword*, 4 = *maybe word*, 5 = *likely word*, and 6 = *definitely word*). All test items were presented with an H–L pitch pattern. Of the 12 test items that were words, six had been congruent with distal prosody during exposure and six had been incongruent; all had been presented with both H–L and L–H pitch patterns. The nonwords consisted of the same syllables as the congruent and incongruent items, but presented in reverse order (so that the transitional probabilities during the exposure phase had been 0). After the experiment, participants completed a demographic survey and provided information about any strategies used.

Results

Word ratings (from 1 to 6) were normalized for each participant by *z*-scoring their responses, in order to account for differences in using the range of the scale (K. Johnson, 2011); Fig. 2 shows normalized word ratings for the test items (for the raw scores, see the Appendix). To examine the effect

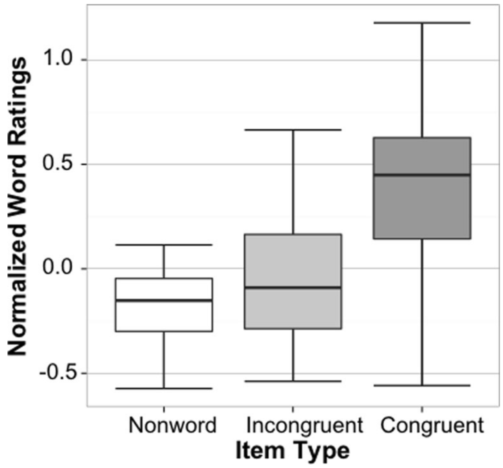

Fig. 2 Normalized ratings (means for each participant) of test items by item types. Horizontal black bars represent the medians, and boxes and whiskers represent the quartiles

of distal prosody on word learning, a mixed-effects logistic regression model analysis was performed in R (Bates, Maechler, & Bolker, 2012). Test item type (incongruent, congruent, or nonword) was included as a fixed effect, and subjects and items were included as random effects (Table 4).

With the fixed-effect reference level as the nonword test item type, congruent test items ($\beta = 0.56$, $t = 7.72$, $p < .001$) and incongruent test items ($\beta = 0.14$, $t = 1.96$, $p < .05$) both exhibited significantly higher ratings (i.e., more word-like) than the nonwords. Treatment coding, with incongruent test items as the reference level (Table 4), showed that the difference between the congruent test items and the incongruent test items was also significant ($\beta = 0.42$, $t = 4.98$, $p < .001$), indicating that participants were more likely to give high ratings to items that had been presented as congruent with distal prosody during exposure than to those presented as incongruent with distal prosody (Fig. 2).[1]

Discussion

Consistent with an effect of distal prosody on word learning, participants gave higher ratings to words if they had been congruent with distal prosody than if they had been incongruent with distal prosody. These results provide the first evidence that word learning can be influenced by prosodic information in the nonadjacent (i.e., distal) context. An alternative possibility, however, is that the observed effects might not have been generated by the distal context, but by the prosodic manipulation on the adjacent, preceding syllable. This possibility seems unlikely, given that in a previous study, Dilley and McAuley (2008) showed that distal prosodic effects on lexical perception were much larger than those observed with a truncated (single-syllable) prosodic context. Nonetheless, to ensure that word learning in the present study was influenced by the distal context, and not only by the prosodic manipulation on the adjacent syllable, we conducted a second experiment in which the stimuli were truncated; the prosodic context consisted of only the syllable preceding the critical-syllable sequence. If distal prosody and perceptual grouping affect word learning, then we would expect to see smaller effects of prosodic context with truncated stimuli than were observed in Experiment 1 with the full (distal) prosodic context.

Table 4 Experiment 1: Mixed-effects logistic regression model with coefficient estimates, standard errors, *t* values, and the significance level of each predictor

	Estimate	Std. Error	*t* Value	*p* Value
(*Reference level*: Incongruent words)	–.03299	.08733	–0.378	.7057
Congruent words	.41576	.08347	4.981	.0000
Nonwords	–.14190	.07224	–1.964	.0498

Test item types: incongruent with distal prosody, congruent with distal prosody, and nonword

Experiment 2

Method

Participants A total of 33 native English speaking undergraduate students (21 female, 12 male; 18–23 years, $M = 19$, $SD = 1.4$) with self-reported normal-hearing participated in the experiment for partial course credit. The data from two additional participants were eliminated from the final sample—one due to self-reported dyslexia, and another for failure to follow instructions. The participants varied in years of music training ($M = 2.7$, $SD = 3.5$).

Stimuli Stimulus phrases of five and six syllables were constructed as in Experiment 1, except that the carrier sequences were reduced from five syllables to one syllable. The truncated carriers consisted of only the final syllable of the original carriers. The four-syllable critical-syllable sequences were identical to those in Experiment 1. Each critical-syllable sequence was appended to each single-syllable carrier sequence, as in Experiment 1, for a total of 144 stimulus phrases. The prosodic manipulation on the syllable preceding the critical-syllable sequence was identical to that of the final syllable preceding the critical-syllable sequence in Experiment 1. Twelve filler phrases also contained five or six syllables, including a repeated pair of syllables.

[1] The same pattern of results emerged in an analysis conducted on the raw ratings scores, with the difference between congruent and incongruent ratings and the difference between congruent and nonword ratings being significant ($p < .001$) and the difference between incongruent and nonword ratings approaching significance ($p = .07$).

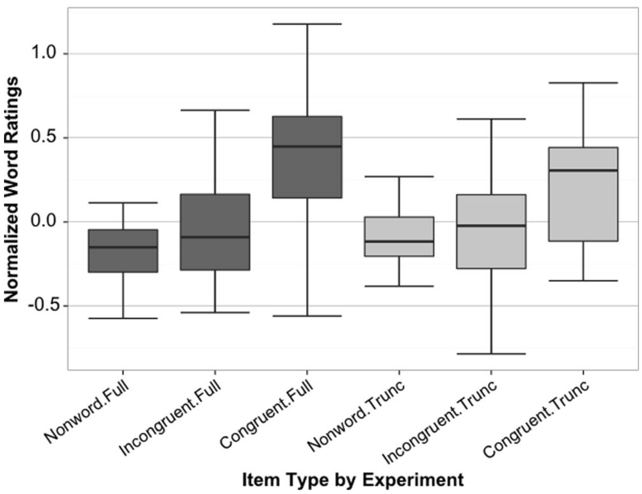

Fig. 3 Normalized ratings (means for each participant) of test items by item type, for Experiment 1 ("Full") and Experiment 2 ("Trunc"). Horizontal black bars represent the medians, and boxes and whiskers represent the quartiles

Apparatus The apparatus was identical to that of Experiment 1.

Procedure The procedure was identical to that of Experiment 1, except that during the exposure phase, the truncated (five- and six-syllable) stimulus phrases replaced the original (nine- and ten-syllable) stimulus phrases.

Results

Figure 3 shows normalized word ratings for the test items in both Experiments 1 and 2. Overall, the effects of the truncated prosodic context in Experiment 2 were weaker than the effects of the full prosodic context in Experiment 1. To examine the reliability of the effect of truncated prosodic context on word learning, a mixed-effects logistic regression model analysis was performed with test item type (congruent, incongruent, or nonword) as a fixed effect and subjects and items as random effects (Tables 5 and 6).

For the truncated prosodic context (Experiment 2), with the fixed-effect reference level as the nonword test item type, congruent test items received significantly higher ratings (i.e., more word-like) than nonwords ($\beta = 0.30$, $t = 3.68$, $p < .001$), whereas incongruent test items did not ($\beta = 0.03$, $t = 0.33$, $p = .739$). Treatment coding with the incongruent test items as the reference level showed that the difference between the congruent and incongruent test items was also significant ($\beta = 0.27$, $t = 2.90$, $p < .05$; see Table 5). For the analysis comparing the effects of distal prosodic context (full stimuli: Exp. 1) with the truncated prosodic context (truncated stimuli: Exp. 2), the predicted interaction between experiment and item type was reliable ($\beta = -0.25$, $t = -2.235$, $p < .05$; Table 6); the effect of prosodic context was significantly weaker with the truncated than with the full stimuli.

Discussion

The results demonstrated that the prosodic manipulation in the syllable adjacent to the critical-syllable sequences does affect listeners' segmentation of the sequences, with participants assigning higher word ratings to items that were congruent with the prosody in the adjacent syllable. However, crucially, the effect of the truncated prosodic context in Experiment 2 was weaker than the effect of the full prosodic context in Experiment 1. The results confirm that distal prosodic context influences word learning in a manner consistent with the perceptual grouping hypothesis.

General discussion

In this study, we investigated whether distal prosodic context can induce perceptual grouping and influence the segmentation and learning of novel words in an unknown language. Experiment 1 showed that syllable pairs were more likely to be perceived as words if they had been presented as congruent with distal prosody than if they had been incongruent with distal prosody. Experiment 2 showed that the effects of prosodic context were weaker with a truncated context than with the full prosodic context examined in Experiment 1. Taken together, these results provide the first evidence that word learning can be influenced by prosodic information in the nonadjacent (i.e., distal) context.

Previous research has shown that local prosodic cues that conflict with transitional probabilities can interfere with statistical learning (e.g., E. K. Johnson & Jusczyk, 2001; E. K. Johnson & Seidl, 2008; Langus et al., 2012; Shukla et al., 2007; Shukla et al., 2011). Dilley and McAuley (2008)

Table 5 Experiment 2: Mixed-effects logistic regression model with coefficient estimates, standard errors, *t* values, and the significance level of each predictor

	Estimate	Std. Error	*t* Value	*p* Value
(*Reference level*: Incongruent word)	–.05504	.08353	–0.659	.5118
Congruent word	.27487	.09468	2.903	.0038
Nonword	–.02736	.08214	–0.333	.7391

Test item types: incongruent with truncated prosodic context, congruent with truncated prosodic context, and nonword

Table 6 Experiment 2: Mixed-effects logistic regression model with coefficient estimates, standard errors, *t* values, and the significance level of each predictor

	Estimate	Std. Error	*t* Value	*p* Value
(*Reference level*: Full stimuli, Nonwords)	–.17634	.06852	–2.574	.0134
Truncated stimuli	.08905	.06333	1.406	.1599
Incongruent words	.14770	.07423	1.990	.0468
Congruent words	.55767	.07423	7.512	$.35 \times 10^{-14}$
Experiment (*Truncated*): Item type (*Incongruent*)	–.11097	.10972	–1.011	.3120
Experiment (*Truncated*): Item type (*Congruent*)	–.24521	.10972	–2.235	.0256

Experiment: full vs. truncated stimuli. Test item types: incongruent with truncated prosodic context, congruent with truncated prosodic context, and nonword

showed that, in one's native language, different distal prosodic patterns lead to distinct parsings of lexically ambiguous syllable sequences, and they proposed a perceptual grouping hypothesis to account for this finding. The novel contribution of the present study is that it has provided the first evidence that *expectations* about how syllables are grouped on the basis of distal prosodic cues, without support from lexical content, can enhance the learning of novel words temporally removed from the context. The finding that perceptual grouping affects the learning of words complements recent evidence that perceptual salience and organization can affect statistical learning of linguistic information, such as sound categories (Emberson, Liub, & Zevinc, 2013) and possibly rule learning (Endress, Scholl, & Mehler, 2005). In other domains, temporal predictability has been shown to enhance statistical learning of visual and auditory patterns (Barakat, Seitz, & Shams, 2013). The present results thus add to this literature by showing that perceptual grouping based on distal prosodic patterning can also affect the extraction of statistical regularities in speech, supporting the view that general principles of auditory perceptual organization may affect language learners at the earliest stages of acquisition. Future research will be needed to determine the extent to which infants demonstrate use of distal prosodic cues in word learning, and whether native speakers of languages other than English can also perceive distal pitch and timing cues in an unknown language.

One open question pertains to the recognition of test items that participants rated highly. Since each item was presented with the H–L pitch pattern (and listeners heard both H–L and L–H patterns during exposure), this raises the question of whether listeners would still recognize test items presented with a pattern not heard during exposure. In other words, would listeners have access to the representation of the syllable sequence, even without a specific pitch pattern as a cue? To address this question, we conducted a follow-up experiment in which the test items were presented with a monotone, flat pitch pattern (235 Hz) not heard during the exposure phase. Participants (n = 17) were exposed to the artificial language as in Experiment 1, and rated the monotone test items. The results for the monotone test items show the same pattern as those for the H–L pitched test items (Fig. 4): Congruent words were rated as being more word-like than both nonwords and incongruent words ($p < .05$).

Thus, these results replicate the finding that listeners use distal prosodic context to perceptually group ambiguously segmentable syllables into candidate words. Moreover, the recognition of these words in a later test phase appears to be independent of the pitch cues associated with the word during exposure.

Conclusion

The present study is the first to demonstrate that distal prosodic cues can influence the learning of novel words. Consistent with a perceptual-grouping hypothesis, candidate words that were congruent with the grouping predicted by the distal prosodic patterning were more likely to be judged as words than were candidate words that were incongruent with the distal prosodic patterning, despite the fact that (1) the critical-syllable sequences (which contained the candidate words) in the congruent and incongruent conditions were acoustically identical, and (2) all of the candidate words had identical transitional probabilities between adjacent syllables. Moreover, because weaker effects of

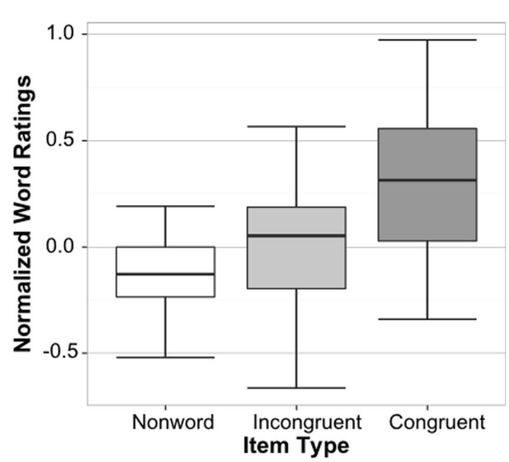

Fig. 4 Mean normalized ratings of monotone test items, by item types

prosodic context were observed for truncated stimuli than for the full prosodic context, our results confirm that word learning in the present study was influenced by the distal context, and not just by the prosodic manipulation on the immediately preceding syllable, or the adjacent context. Extending previous research on the effects of distal prosodic context to the learning of words in an artificial language also provides evidence that the effects of distal prosody may be independent of lexical or other language-specific knowledge. Taken together, these findings support the view that general principles of auditory perceptual organization play an important role in speech perception and in language acquisition. Thus, this study serves as a step to eventually examining the role of distal prosody in more naturalistic language-learning environments (for both first and second language learning), as well as for examining the generalizability of the role of a perceptual grouping mechanism across languages.

Author note We thank Neelima Wagley, Ashley Elliston, Brian Chivers, and Mitchell Reddan for assistance with the stimulus creation and experiment running. This work was partially supported by NSF CAREER Grant No. BCS-0874653 to L.C.D., the Department of Psychology at Michigan State University, and a Michigan State University Provost Undergraduate Research Initiative grant to J.D.M. and P.A.Z.

Appendix

Table 7 Raw mean ratings by subjects and by items for test items

By Subjects				By Items			
	Congruent	Incongruent	Nonword		Congruent	Incongruent	Nonword
Experiment 1: Full Stimuli							
Mean	4.35	3.73	3.52	Mean	4.35	3.73	3.53
StDev	0.65	0.91	0.72	StDev	0.77	0.73	0.61
Range	3.17	4.17	2.92	Range	3.11	2.93	2.64
Experiment 2: Truncated Stimuli							
Mean	3.95	3.45	3.45	Mean	3.96	3.48	3.48
StDev	1.24	0.97	0.96	StDev	0.97	0.78	0.57
Range	5.00	3.83	4.50	Range	3.22	2.67	2.04

References

Adams, T. M. (2010, May). *Prosodic transfer and phonological learning in a second language fluent speech segmentation task*. Paper presented at Speech Prosody 2010, 5th International Conference, Chicago, IL.

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Language and Cognition, 6,* 81–96.

Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representation: Predictable items are enhanced even when not predicted. *Cognition, 129,* 205–211. doi: 10.1016/j.cognition.2013.07.003

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes [Software]. Retrieved from http://CRAN.R-project.org/package=lme4

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer [Computer program]. Retrieved from www.praat.org

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93–125.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics, 31,* 585–598.

Creel, S. C., Tanenhaus, M. K., & Aslin, R. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 15–32.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

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

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24,* 381–410.

Dilley, L. C., Mattys, S., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63,* 274–294.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation & lexical processing. *Journal of Memory and Language, 59,* 294–311.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language, 36,* 406–421. doi:10.1006/jmla.1996.2500

Emberson, L. L., Liub, R., & Zevinc, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition, 128,* 82–102.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134,* 406–419.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499–504.

Goetry, V., & Kolinsky, R. (2000). The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners. *Psychologica Belgica, 40,* 115–152.

Hay, J., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2012). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63,* 93–106.

Johnson, K. (2011). *Quantitative methods in linguistics*. Malden, MA: Blackwell.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy, 13,* 440–455.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental Science, 12,* 131–141. doi:10.1111/j.1467-7687.2008.00740.x

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687. doi:10.1111/j.1467-8624.1993.tb02935.x

Langus, A., Marchetto, E., Hoffmann Bion, R. A., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66,* 285–306. doi:10.1016/j.jml.2011.09.004

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465–494.

Newport, E. L., & Aslin, R. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

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

Sanders, L. D., & Neville, H. J. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language, and Hearing Research, 45,* 1301–1321.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: II. Segmentation, semantics and syntax in non-native speakers. *Cognitive Brain Research, 15,* 214–227.

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

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences, 108,* 6038–6043.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50,* 86–132.

Toro, J. M., Pons, F., Hoffmann Bion, R. A., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically-coherent word candidates. *Journal of Memory and Language, 64,* 171–180.