Psychon Bull Rev (2011) 18:1189–1196
DOI 10.3758/s13423-011-0167-9

BRIEF REPORT

Expectations from preceding prosody influence segmentation in online sentence processing

Meredith Brown · Anne Pier Salverda · Laura C. Dilley · Michael K. Tanenhaus

Published online: 4 October 2011

Abstract Previous work examining prosodic cues in online spoken-word recognition has focused primarily on local cues to word identity. However, recent studies have suggested that utterance-level prosodic patterns can also influence the interpretation of subsequent sequences of lexically ambiguous syllables (Dilley, Mattys, & Vinke, *Journal of Memory and Language*, *63*:274–294, 2010; Dilley & McAuley, *Journal of Memory and Language*, *59*:294–311, 2008). To test the hypothesis that these distal prosody effects are based on expectations about the organization of upcoming material, we conducted a visual-world experiment. We examined fixations to competing alternatives such as *pan* and *panda* upon hearing the target word *panda* in utterances in which the acoustic properties of the preceding sentence material had been manipulated. The proportions of fixations to the monosyllabic competitor were higher beginning 200 ms after target word onset when the preceding prosody supported a prosodic constituent boundary following *pan-*, rather than following *panda*. These findings support the hypothesis that expectations based on perceived prosodic patterns in the distal context influence lexical segmentation and word recognition.

M. Brown (✉) · A. P. Salverda · M. K. Tanenhaus
Department of Brain and Cognitive Sciences,
University of Rochester,
Meliora Hall, Box 270268, Rochester, NY 14627-0268, USA
e-mail: mbrown@bcs.rochester.edu

L. C. Dilley
Department of Communicative Sciences & Disorders,
Michigan State University,
116 Oyer Center, East Lansing, MI 48824, USA

Keywords Prosody · Expectations · Spoken-word recognition · Lexical competition · Perceptual organization · Visual-world paradigm

Expectation and prediction are increasingly playing major roles in models of language processing (e.g., Jurafsky, 1996; Levy, 2008). Effects of expectations based on lexical and (morpho)syntactic properties are well-documented (Kamide, 2008). In contrast, relatively little work has focused on prosodic expectations, including prosodic phenomena distal (i.e., several syllables removed) from the locus of processing. Preceding prosodic phenomena might influence interpretation of proximal material by contributing to listeners' expectations about the prosodic organization of upcoming material. Here, we provide evidence that perceived prosodic patterns generate expectations that can constrain word segmentation and recognition.

Previous work has shown that listeners detect pitch, temporal, and/or amplitude patterning in nonlinguistic auditory stimuli ranging from simple tone sequences to musical passages. This patterning influences metrical and grouping structures that listeners perceive downstream (Handel, 1989). For example, listeners tend to perceptually organize a sequence of tones of equal duration, temporal separation, and amplitude differing only by alternating in pitch height (high/low) into either high–low or low–high groups, with the first element in each group perceived as accented (Thomassen, 1982; Woodrow, 1911). To describe such perceptual proclivities in music, Lerdahl and Jackendoff (1983) proposed parallelism preference rules, which state that when a musical passage contains segments perceived as similar or repetitive, parallel parts of segments will be construed as having similar metrical and grouping structure. These studies suggest that distal

regularities in auditory stimuli can influence processing of proximal material.

Likewise, speech prosody often exhibits regularities in pitch, duration, and/or amplitude that listeners perceive as patterning (Couper-Kuhlen, 1993; Dainora, 2001; Pierrehumbert, 2000). For example, in English and other languages, stressed syllables tend to be perceived as perceptually isochronous—that is, occurring at regular intervals (e.g., Lehiste, 1977; Patel, 2008). Moreover, English and other languages tend to exhibit recurring prenuclear pitch accents within intonational phrases (Couper-Kuhlen, 1993; Crystal, 1969; Dainora, 2001; Pierrehumbert, 2000).

On the basis of these observations, Dilley and McAuley (2008) proposed a *perceptual grouping hypothesis* claiming that distal prosodic regularities in some *earlier* context that are associated with (1) distinct perceived groupings of syllables into prosodic constituents and (2) distinct patterns of perceptual isochrony might also generate different expectations about the constituency and rhythmic structure of *following* material in speech, similar to perceived regularities observed in nonspeech auditory processing. These expectations then constrain spoken-word recognition by favoring lexical candidates that are consistent in lexico-prosodic constituency and/or metrical structure with the anticipated prosodic structure.

Supporting this hypothesis, Dilley and McAuley (2008) conducted an experiment using eight-syllable strings, beginning with two-syllable words with initial stress and ending with syllables with lexically ambiguous structure (e.g., *channel dizzy foot–note–book–worm*, for which the final syllables could be structured as *footnote–bookworm*, *foot–notebook–worm*, etc.). They showed that manipulating the fundamental frequency (*f*0) and/or the timing of the initial five syllables influenced listeners' perceptions of the lexical constituency of the subsequent syllables—for instance, *note–book–worm*—which were held acoustically identical. Participants reported hearing more final two-syllable words (e.g., *bookworm*) when the preceding prosody encouraged constituency and metrical structure expectations for the final ambiguous syllables to form two-syllable units than when the prosody did not encourage such structuring. Manipulating either *f*0 or durational cues alone produced differences in the rates of final two-syllable words reported, but combining the cues led to the largest differences across conditions in the rates of final two-syllable words heard. These findings were replicated and extended by Dilley, Mattys, and Vinke (2010), who demonstrated that distal prosody is a robust cue to lexical organization even when semantic context or proximal acoustic cues suggest a different structure.

The locus of these distal prosody effects, however, remains an open question. According to the perceptual grouping hypothesis, prosodic patterns influence listeners' expectations about the prosodic structure of upcoming material, which influence the online perception of spoken words. However, another possibility consistent with previous findings is that distal prosody effects have a post-perceptual locus. Using a cross-modal identity priming paradigm, Dilley et al. (2010) found that distal prosody influences lexical segmentation within the first second after hearing a string of lexical items ending in potentially final-embedded words (e.g., *turnip* vs. *nip*). Although this finding demonstrates rapid effects of preceding prosody on interpreting lexically ambiguous material, these effects could arise from post-perceptual reinterpretation. No studies have directly addressed the prediction of the perceptual grouping hypothesis that distal prosody influences listeners' online expectations about the prosodic structuring of subsequent material. In addition, distal prosody effects on subsequent input have so far been demonstrated only in the processing of lists of words lacking grammatical structure. It is important to evaluate whether listeners are sensitive to distal prosodic patterns in more ecologically valid contexts—for example, spoken sentences.

We used the visual-world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to assess the online effects of distal prosody on the processing of utterances containing words such as *panda*, which are temporarily ambiguous due to the presence of an onset-embedded word (*pan*). We manipulated the acoustics of each utterance such that prosodic constituents preceding the target word had temporal and *f*0 characteristics that were either congruent or incongruent with the prosodic characteristics of the target word, which were identical across conditions (e.g., *Heidi sometimes saw that panda*… vs. *Heidi sometimes saaaw that panda*, where underlining denotes high *f*0). The perceptual grouping hypothesis makes two clear predictions: First, if distal prosody influences listeners' expectations about the prosodic organization of upcoming lexical material, we should observe greater or lesser lexical activation of the embedded competitor word (as indexed by fixations to its referent—e.g., a visually depicted pan), depending on whether the distal prosodic context biases the listener to expect a prosodic constituent boundary and/or stressed syllable following the first syllable of the target word. Second, effects of distal prosody on fixation patterns should emerge in the earliest signal-driven fixations after target word onset.

Method

Participants

A group of 43 participants from the University of Rochester and the surrounding community took part in the experi-

ment. All were native English speakers with normal hearing and normal or corrected-to-normal vision.

Materials

The 20 speech stimuli were grammatical declarative sentences containing a target word with an onset-embedded competitor word (e.g., *panda*; see Table 1). The distal context preceding the target word consisted of two disyllabic words with initial primary stress, followed by one monosyllabic word (e.g., *Heidi sometimes saw*). The proximal context contained another monosyllabic function word (e.g., *that*) followed by the target word (e.g., *panda*). The target word was followed by a 3- to 5-syllable continuation. Each stimulus was associated with four pictures: a target, a competitor, and two distractors with names that were phonologically unrelated to the target word.

To prevent participants from developing expectations that target words corresponded to the longer of two phonologically related pictures, there were 20 filler trials with a target word referring to the shorter of two phonologically related pictures. Similarly, to discourage participants from developing expectations that phonologically related pictures were likely targets, 20 filler trials had two phonologically related distractors. The structure of the filler items (e.g., the number of syllables) was varied to reduce the overall predictability of the position of the target word in the sentence.

The first author recorded all items using a Marantz PMD660 digital recorder sampling at 32 kHz in a sound-attenuated booth. Sentences were produced at an approximately uniform rate, with minimal *f*0 excursions and slight *f*0 declination across each utterance. The average durations of the target and the embedded words were 675 and 366 ms, respectively.

The pitch-synchronous overlap-and-add algorithm (Moulines & Charpentier, 1990) was used in Praat (Boersma & Weenink, 2010) to create two resynthesized versions of each recording with different prosodic patterns across the first five syllables of each sentence (i.e., the distal context). Importantly, the acoustic characteristics from the start of the syllable preceding the target word (i.e., the proximal context) through to the end of the utterance were held constant. The word preceding the target word was associated with high *f*0, and the target word itself with a low–high *f*0 sequence. Controlling the acoustic characteristics of the target word and the proximal context ensured that differences in participants' fixation patterns between the experimental conditions could not be attributed to proximal prosodic cues.

The *f*0 manipulations involved shifting all pitch points within the vowel of each syllable up by 35 Hz (for high syllables) or down by 25 Hz (for low syllables), enabling preservation of natural *f*0 declination from the original recording while imposing periodic alternations. The *f*0 contours across nonvocalic portions of speech were

Table 1 List of stimuli

Stimulus	Target	Competitor
Andy really got his antlers from a mantelpiece.	antlers	ant
Many people said that beaker was on the top shelf.	beaker	bee
Cindy's brother took that boulder for his rock garden.	boulder	bowl
Mr. Johnson gave his candy to those students.	candy	can
Amy's photo with her captain was lost when she moved.	captain	cap
Maybe Carlos bought this carpet on his vacation.	carpet	car
Maybe Emma got her dolphin poster at Sea World.	dolphin	doll
Marty's father got that hamster for his son's birthday.	hamster	ham
Cara never found that leaflet from the agency.	leaflet	leaf
Tyler never got his nectarine smoothie yesterday.	nectarine	neck
Heidi sometimes saw that panda in the city zoo.	panda	pan
Marcus really wants that pirate costume for his son.	pirate	pie
Jamie's family bought that pumpkin at a local farm.	pumpkin	pump
Jenna never liked that reindeer theme for the party.	reindeer	rain
Mr. Morris wants one soldier to guard the entrance.	soldier	sole
Timmy sometimes sees one spider sitting in its web.	spider	spy
Many people said those taxi cabs are often late.	taxi	tacks
Sometimes people want one toaster in their cabinet.	toaster	toe
People sometimes see that tractor on Mr. Smith's farm.	tractor	track
Betsy never found that welder very friendly.	welder	well

Target and competitor words are provided after each sentence

determined by monotonic interpolation between immediately preceding and following *f*0 values.

In the low–high (LH) grouping condition, the first five syllables contained prosodic constituents with initial stress whose boundaries were temporally aligned with edges of roughly perceptually isochronous low–high tone sequences (Fig. 1a). This timing was achieved by modifying the duration of the fifth rime such that the fifth intervocalic interval (i.e., the interval between the vowel onsets of the fifth and sixth syllables) was equivalent to the mean duration of the sixth and seventh intervocalic intervals (cf. Dilley et al., 2010; Dilley & McAuley, 2008). The LH-grouping context was predicted to promote the perception of prosodic boundaries preceding and following the target word, consistent with either *pan* or *panda*.

In the high–low (HL) grouping condition, the first five syllables contained prosodic constituents with initial stress whose boundaries were aligned with edges of high–low tone sequences (Fig. 1b). The rime of the fifth syllable was lengthened such that the total duration of the fifth intervocalic interval was equivalent to the mean of the first two interstress intervals (measured between vowel onsets). This duration manipulation enabled the fifth syllable to be paired with two tones instead of one while maintaining approximately regular timing for each HL tone group. The HL-grouping condition was predicted to bias listeners to perceive the target word as two prosodic units—for instance, *pan-da*, with a boundary following *pan*. This prosodic boundary was predicted to increase the activation of the embedded word, which should result in more fixations to the competitor picture in the HL-grouping condition.

The *f*0 and duration manipulations were also performed on the fillers. To eliminate a consistent association between *f*0 and duration patterns, half of the fillers beginning with HL tone sequences contained one shortened syllable at a variable position within the sentence, whereas the other half contained one lengthened syllable. Likewise, half of the fillers starting with LH sequences contained one shortened syllable, and the rest contained one lengthened syllable.

Procedure

Eye movements were recorded using a head-mounted SR Research EyeLink II system sampling at 250 Hz. Drift

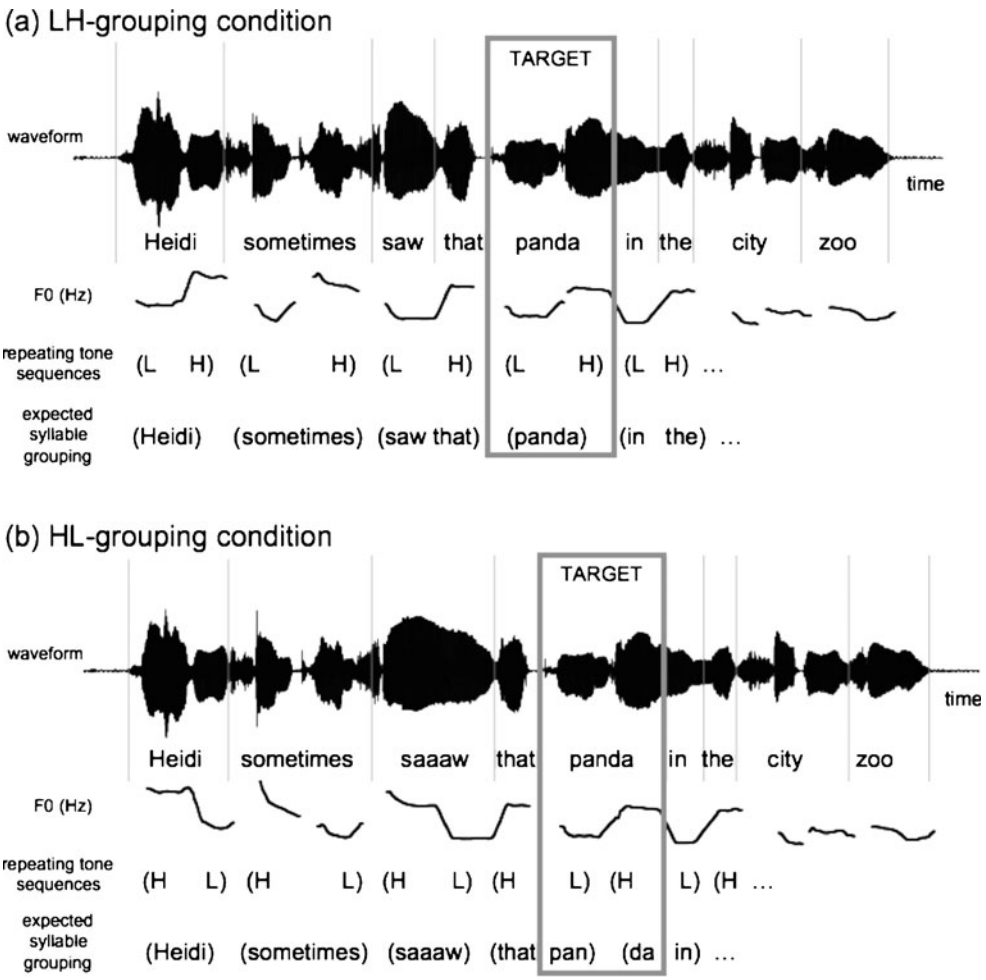

Fig. 1 Explanation of the resynthesis methods performed to generate stimuli in the low–high (LH) grouping condition (**a**), and the high–low (HL) grouping condition (**b**). The *f*0 of the first five syllables and the duration of the fifth syllable were manipulated to discourage or encourage, respectively, the perception of a prosodic boundary and stressed syllable following the embedded word *pan*. The acoustic properties of the stimuli were held constant from the onset of the sixth syllable through the end of the utterance

correction was performed every five trials. Each trial began with the presentation of a visual display containing four colored clip-art pictures (Fig. 2). After 500 ms of preview, a spoken sentence was played. Participants were instructed to click on the picture referred to in the sentence and were not given feedback on their performance. Pilot testing indicated that participants generally noticed that the speech had been manipulated, so we employed the cover story that we were evaluating the comprehensibility of synthesized speech stimuli.

Four lists were constructed by randomizing picture positions within each trial and pseudo-randomizing the order of trials. Within each list, half of the experimental trials had HL-grouping prosody, and the other half had LH-grouping prosody. The assignment of stimuli to each type of grouping was counterbalanced, resulting in eight lists. Four practice trials were included at the start of the experiment to familiarize participants with the procedure.

Results

Three of the participants were excluded from the analysis because they did not meet the inclusion criteria or did not complete the experiment. These participants were replaced such that analyses of the fixation patterns included data from 5 participants for each of the eight lists. Overall, participants found the referent identification task easy and clicked on the incorrect picture on fewer than 3% of experimental trials. These trials were excluded from further analysis.

Figure 3 shows the proportion of fixations over time as a function of condition for each type of picture starting at target word onset. The perceptual grouping hypothesis predicts an early competitor bias and more looks to the competitor in the HL-grouping condition than in the LH-grouping condition. The data confirmed both predictions. In the HL-grouping condition, fixation patterns revealed a competitor bias that emerged around 200 ms following onset of the target word (i.e., approximately when signal-driven fixations should emerge, assuming a delay of 200 ms to program an eye movement; Matin, Shao, & Boff, 1993) and persisted until approximately 700 ms. In the LH-grouping condition, however, fixation proportions to target and competitor pictures were roughly equivalent until 600 ms (approximately 200 ms following the offset of the embedded word), when participants began to converge on the target picture.

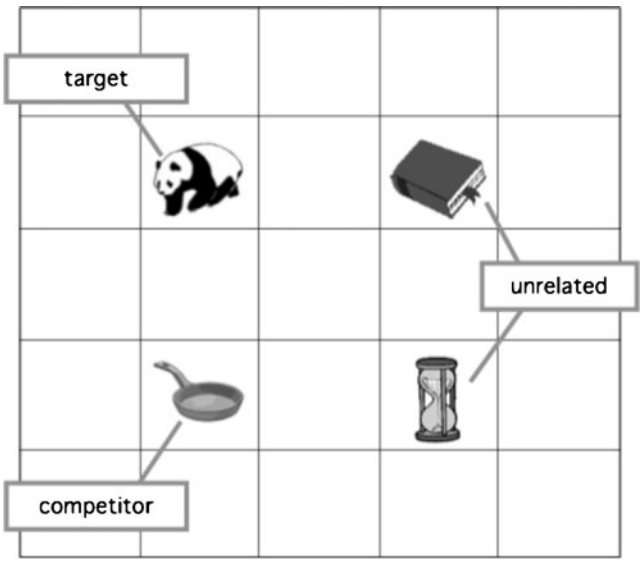

Fig. 2 The visual display for an example trial in the experiment, with picture labels added for clarity

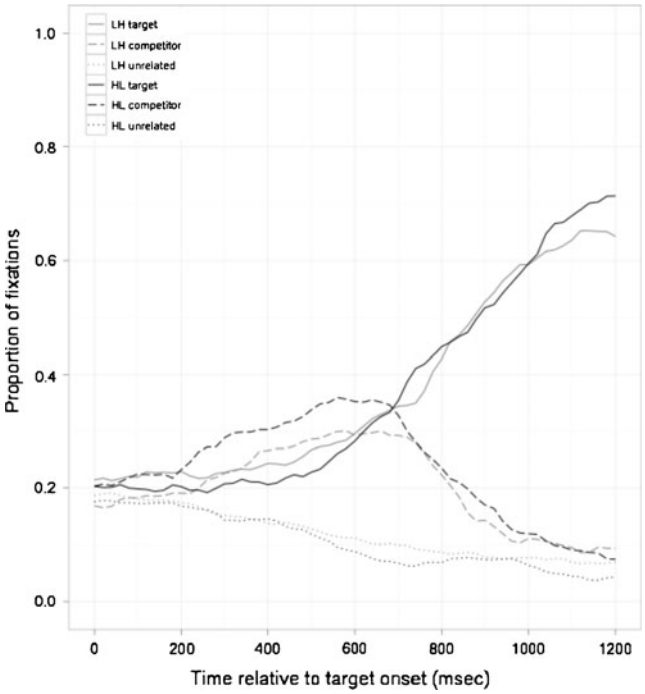

Fig. 3 Proportions of fixations to the target, competitor, and unrelated pictures in the HL-grouping and LH-grouping conditions, from the onset of the target word until 1,200 ms following target onset. The mean duration of the embedded word was 366 ms, and that of the target word 675 ms

For statistical analysis of these fixation curves, we employed growth curve analysis (GCA; Mirman, Dixon, & Magnuson, 2008), a technique that is increasingly being applied to visual-world data. GCA is a variant of multilevel regression modeling in which orthogonal power polynomial terms are fit to time series data to model variations in curve parameters (e.g., intercept and slope) that can be attributed to independent variables and/or individual differences in processing. GCA models the trajectory of change over time through the estimation of curve parameters. This approach avoids several disadvantages of standard ANOVAs on

Table 2 Results of growth curve analyses comparing fixations to target and competitor pictures in the low–high and high–low grouping conditions

		Competitor			Target		
		B	*t*	*p*	*B*	*t*	*p*
Intercept							
	Subjects	–.054	−2.652	<.05	.030	1.935	<.10
	Items	–.052	−3.208	<.005	.027	1.422	n.s.
Linear							
	Subjects	.019	0.257	n.s.	.022	0.282	n.s.
	Items	.023	0.309	n.s.	.024	0.279	n.s.
Quadratic							
	Subjects	–.035	−2.914	<.005	.002	0.136	n.s.
	Items	–.033	−3.186	<.005	.000	0.000	n.s.
Cubic							
	Subjects	–.056	−4.669	<.0001	–.029	−2.469	<.05
	Items	–.056	−5.445	<.0001	–.029	−3.567	<.005

All models included data from 200 to 566 ms after the onset of the target word. *B* = parameter estimate

averaged proportions of fixations, including violation of independence assumptions and loss of fine-grained temporal detail. GCA is also an appropriate analysis technique for the present study because effects of prosody on the interpretation of spoken language can differ across individuals (e.g., Fraundorf & Watson, 2010). GCA takes individual variation into account by explicitly incorporating parameters that estimate effects of experimental factors for each participant.

Fixations to target and competitor pictures were analyzed in separate models including data from 200 to 566 ms after the onset of the target word (i.e., from the earliest point at which signal-driven fixations were expected until 200 ms after the mean offset of the embedded word).[1] Separate analyses assessed by-subject and by-item effects on target and competitor fixations. All models included intercept, linear, quadratic, and cubic terms. Model comparison verified that higher-order terms contributed significantly to model fit. Additional analyses were performed on fixations between 0 and 200 ms; no significant effects of condition on baseline competitor and target activation were found.

The primary goal of these analyses was to determine whether competitor fixations were significantly higher in the HL-grouping condition than in the LH-grouping condition during the processing of the onset of the target word. The perceptual grouping hypothesis most directly predicts differences in the intercept term, which represents the average height of each curve, though any differences in curve parameters between conditions are of interest because they can be attributed solely to distal context.

[1] The same patterns of results were obtained using *t* tests on fixation proportions across the analysis window, so the main findings are not dependent on the use of GCA.

The results are shown in Table 2. For competitor fixations, the distal-prosody condition had a significant effect on the intercept term, indicating that the mean proportions of fixations to the competitor were significantly higher in the HL-grouping than in the LH-grouping condition. The effect of condition on the linear term was not significant, suggesting a similar rate of change in fixation proportions in the HL-grouping and LH-grouping conditions. A significant effect of condition on both the quadratic and cubic terms revealed that the steepness of the curve preceding and following the inflection points in the competitor fixation curve was greater in the HL-grouping condition than in the LH-grouping condition.

For target fixations, condition did not have a significant effect on the intercept term, indicating that the mean proportions of fixations did not differ between conditions. Likewise, condition did not have a significant effect on the linear term: Within the analysis window, the slope of the target fixation curve did not differ between conditions. Condition significantly affected the cubic term, but not the quadratic term, suggesting that the shape of the curve at the edges of the analysis window differed between conditions.

In summary, early effects of distal prosody on lexical processing were observed primarily in fixations to the competitor. Competitor fixations had a higher mean value and a steeper curve surrounding inflection points in the HL-grouping than in the LH-grouping condition, consistent with the predictions of the perceptual grouping hypothesis.

Discussion

The results from our visual-world experiment supported the predictions of the perceptual grouping hypothesis. Starting around 200 ms after onset of a temporarily ambiguous

target word (e.g., *panda*), the proportion of fixations to a visually displayed competitor (e.g., *pan*) was higher when the preceding prosody supported a prosodic boundary and a stressed syllable following *pan-* than when these conditions did not hold. These results suggest that the embedded competitor word was more strongly activated in the former condition during processing of the onset of *panda*. The observed differences cannot be attributed to the acoustic–phonetic characteristics of the target word or to its proximal context, since they were identical across conditions. We therefore conclude that distal prosodic patterns can influence listeners' expectations about the prosodic organization of upcoming material and that these expectations interact with the processing of the proximal input during online lexical segmentation and word recognition.

These findings are congruent with studies showing that other types of distal information influence word recognition. For example, speech processing can be influenced by nonlinguistic pitch information in the preceding context: Distal sequences of sine wave tones with different spectral properties influence speech categorization (Holt, 2005). Moreover, McAuley, Dilley, Rajarajan, and Bur (2011) recently replicated the results of Dilley et al. (2010) with complementary stimuli in which all low pitches were replaced with high pitches, and vice versa. This adds support for the perceptual grouping hypothesis by suggesting that the effects of distal prosody on perceived lexical organization are not due to a specific pairing of high or low pitch with particular syllabic positions. In addition, distal speaking rate influences the perception of function words in English. Listeners are less likely to perceive a function word that is present in the signal when the preceding and following speech rate is relatively slow, and they are more likely to perceive a function word that is absent from the signal when the surrounding speech rate is relatively fast (Dilley & Pitt, 2010). The segmentation of lexically ambiguous sequences of syllables in Dutch is also influenced by the distal speaking rate, which modulates the perceived duration of phonemes at junctures between words (Reinisch, Jesse, & McQueen, 2011).

Given Reinisch et al.'s (2011) results, it is important to consider the potential effects of distal speaking rate on the observed differences in lexical activation between conditions in our experiment. The duration manipulations caused items with HL-grouping prosody to have slower distal speaking rates than did items with LH-grouping prosody. This speech rate difference may have caused listeners to perceive the target word and its proximal context as having a relatively short duration in the HL-grouping as compared to the LH-grouping condition. Note, however, that differences in the perceived duration of the first syllable of the target word would be predicted to lead to more activation of the monosyllabic competitor in the LH-grouping condition than in the HL-grouping condition, given evidence that listeners use segmental lengthening as a cue to an upcoming prosodic boundary (Salverda, Dahan, & McQueen, 2003; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, & McDonough, 2007). But this is contrary to what we observed. Dilley and McAuley (2008) further showed that truncating the first four syllables of five-syllable distal contexts reduced distal prosody effects, suggesting that the effects were not solely attributable to duration manipulations on the fifth syllable. Taken together, the differences in distal speaking rates between conditions most likely cannot account for our results.

Our data show that the results from previous studies with word lists generalize to more natural utterances. In future work, it will be fruitful to explore the effects of distal prosody using natural, conversational speech. It will also be important to determine specifically which context patterns listeners' expectations are sensitive to. Although our preferred interpretation is that listeners are sensitive to global $f0$ and temporal patterns across prosodic constituents within an utterance, a closely related alternative is that listeners' expectations are based on the co-occurrence of high or low pitches with word-initial or stressed syllables. Both accounts are compatible with our results and make similar claims that listeners develop expectations on the basis of perceived patterns in the preceding context.

To conclude, this work demonstrates that pitch and duration patterns within an utterance can influence listeners' expectations about the prosodic structure of upcoming material. These expectations can constrain listeners' initial segmentation of temporarily ambiguous material. Taken together with other results, our study demonstrates that global prosodic context influences not only the interpretation of local prosodic cues, but also lexical segmentation and recognition. More generally, these findings add to a growing body of work demonstrating that listeners integrate various top-down sources of information with proximal acoustic–phonetic cues to prosodic structure and word identity in the earliest moments of spoken-word recognition.

Author Note This research was supported by a Javits fellowship and an NSF predoctoral fellowship to M.B., NSF Grant BCS-0847653 to L.C.D., and NIH Grants HD27206 and DC0005071 to M.K.T. We gratefully acknowledge Dana Subik for assistance with participant recruitment and testing, and the Tanenhaus lab and the audiences at AMLaP 2010 and CogSci 2011 for helpful discussions.

References

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved 20 January, 2010, from www.praat.org

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam: Benjamins.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. Ph.D. dissertation, University of Chicago.

Dilley, L. C., Mattys, S. L., & 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. doi:10.1016/j.jml.2010.06.003

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59,* 294–311. doi:10.1016/j.jml.2008.06.006

Dilley, L., & Pitt, M. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664–1670.

Fraundorf, S., & Watson, D. (2010). *Who cares about prosody? Predicting individual differences in sensitivity to pitch accent in online reference resolution*. Poster presented at Architectures and Mechanisms for Language Processing 2010, York, U.K.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Holt, L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16,* 305–312.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137–194. doi:10.1016/S0364-0213(99)80005-6

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 2,* 647–670.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5,* 253–263.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177. doi:10.1016/j.cognition.2007.05.006

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53,* 372–380.

McAuley, J., Dilley, L., Rajarajan, P., & Bur, K. (2011). Effects of distal pitch and timing of speech and nonspeech precursors on word segmentation. *Journal of the Acoustical Society of America, 129,* 2683.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59,* 475–494. doi:10.1016/j.jml.2007.11.006

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453–467.

Patel, A. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment. Studies presented to Gösta Bruce* (pp. 11–36). Dordrecht, The Netherlands: Kluwer.

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 978–996.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90,* 51–89. doi:10.1016/S0010-0277(03)00139-2

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition, 105,* 466–476. doi:10.1016/j.cognition.2006.10.008

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. doi:10.1126/science.7777863

Thomassen, J. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America, 71,* 1596–1605.

Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review, 18,* 54–77. doi:10.1037/h0075201