Journal of Experimental Psychology:
Human Perception and Performance
2015, Vol. 41, No. 2, 306–323

0096-1523/15/$12.00 http://dx.doi.org/10.1037/a0038689

Metrical Expectations From Preceding Prosody Influence Perception of Lexical Stress

Meredith Brown and Anne Pier Salverda
University of Rochester

Laura C. Dilley
Michigan State University

Michael K. Tanenhaus
University of Rochester

Two visual-world experiments tested the hypothesis that expectations based on preceding prosody influence the perception of suprasegmental cues to lexical stress. The results demonstrate that listeners' consideration of competing alternatives with different stress patterns (e.g., *'jury*/*gi'raffe*) can be influenced by the fundamental frequency and syllable timing patterns across material preceding a target word. When preceding stressed syllables distal to the target word shared pitch and timing characteristics with the first syllable of the target word, pictures of alternatives with primary lexical stress on the first syllable (e.g., *jury*) initially attracted more looks than alternatives with unstressed initial syllables (e.g., *giraffe*). This effect was modulated when preceding unstressed syllables had pitch and timing characteristics similar to the initial syllable of the target word, with more looks to alternatives with unstressed initial syllables (e.g., *giraffe*) than to those with stressed initial syllables (e.g., *jury*). These findings suggest that expectations about the acoustic realization of upcoming speech include information about metrical organization and lexical stress and that these expectations constrain the initial interpretation of suprasegmental stress cues. These distal prosody effects implicate online probabilistic inferences about the sources of acoustic–phonetic variation during spoken-word recognition.

Keywords: prosody, spoken-word recognition, lexical stress, expectations, lexical competition

Across multiple cognitive and perceptual domains, stimulus processing is facilitated by the ability to detect temporal patterns as they unfold over time and project these patterns forward to predict upcoming information. For example, visual events that have a regular, predictable temporal structure are processed more quickly and accurately than visual events whose timing is relatively unpredictable (e.g., Olson & Chun, 2001; Rolke & Hofmann, 2007). In the auditory domain, listeners evaluate the pitch of tones more accurately when the tones occur in a rhythmically predictable sequence than when they do not (Jones, Moynihan, MacKenzie, & Puente, 2002).

Similar predictive processing may facilitate aspects of spoken-language comprehension. Across languages, speech is generally perceived as having rhythmic temporal organization. In English (and many other languages), this rhythmic organization centers on acoustically prominent or stressed syllables (e.g., Lehiste, 1977; Peelle & Davis, 2012). This perceived rhythmicity may facilitate spoken-word recognition by enabling listeners to generate expectations about the timing and acoustic properties of upcoming stressed syllables in the speech stream. In this article, we specifically examine whether the incremental evaluation of lexical candidates with different stress patterns during spoken-word recognition is influenced by the acoustic realization of stressed and unstressed syllables in preceding utterance context.

The involvement of expectations in language processing is well attested: Linguistic representations in long-term memory, such as lexical and syntactic knowledge, influence the processing of temporarily or globally ambiguous material (e.g., Altmann & Kamide, 1999; Levy, 2008). For example, as listeners incrementally process the utterance *The boy will eat the cake* while viewing a visual display containing a cake and some other (inedible) objects, they preferentially fixate the cake on processing the verb (i.e., before processing *cake*; Altmann & Kamide, 1999). This suggests that listeners can use verb-semantic information to anticipate that the

This article was published Online First January 26, 2015.

Meredith Brown and Anne Pier Salverda, Department of Brain & Cognitive Sciences, University of Rochester; Laura C. Dilley, Department of Communicative Sciences and Disorders, Michigan State University; Michael K. Tanenhaus, Departments of Brain & Cognitive Sciences and Linguistics, University of Rochester.

This research was supported by a National Science Foundation (NSF) predoctoral fellowship to Meredith Brown, NSF Grant BCS-0847653 to Laura C. Dilley, and National Institutes of Health Grants HD073890 and HD027206 to Michael K. Tanenhaus. We gratefully acknowledge Dana Subik for assistance with participant recruitment and testing and, for helpful discussions, the Tanenhaus lab and the audiences at the second conference on Experimental and Theoretical Advances in Prosody, September 2011, Montreal, Quebec, Canada; the 10th Annual Auditory Perception, Cognition, and Action Meeting, November 2011, Seattle, Washington; the 25th Annual City University of New York Conference on Human Sentence Processing, March 2012, New York; and the 34th Annual Meeting of the Cognitive Science Society, August 2012, Sapporo, Japan.

Correspondence concerning this article should be addressed to Meredith Brown, who is now at Tufts University and Massachusetts General Hospital, CNY-2, Department of Psychiatry, Building 149, 13th Street, Charlestown, MA 02129. E-mail: meredith@nmr.mgh.harvard.edu

domain of possible postverbal objects is likely to be restricted to the set of edible display items.

During sentence processing, listeners may also develop expectations about the fine-grained acoustic realization of upcoming speech sounds or words. Listeners interpret phonetic cues to segment identity with respect to surrounding speech context (e.g., Cole, Linebaugh, Munson, & McMurray, 2010; Ladefoged & Broadbent, 1957; McMurray & Jongman, 2011). Effects of speech context on listeners' expectations are not limited to cues in the immediate context: Manipulations of pitch and duration early in an utterance influence the segmentation of temporarily or globally ambiguous material several syllables downstream (Breen, Dilley, McAuley, & Sanders, 2014; Brown, Salverda, Dilley, & Tanenhaus, 2011; Dilley, Mattys, & Vinke, 2010; Dilley & McAuley, 2008). For example, Brown et al. (2011) manipulated the acoustics of prosodic constituents preceding a target word like *panda* such that preceding prosodic words exhibited regular pitch patterns and temporal characteristics that were either similar to the pitch patterns and temporal characteristics across the target word or were dissimilar, encouraging a different perceptual grouping of syllables (e.g., [*saw that*] [*panda*] vs. [*saw*] [*that pan*] [*da in*]). Crucially, the target word and its proximal context (i.e., the syllables immediately adjacent to it) were acoustically identical across conditions. On hearing the target word *panda*, participants were more likely to initially perceive its first syllable as corresponding to the monosyllabic word *pan* when the pitch and timing characteristics of the target word were dissimilar to those of preceding constituents than when the pitch and timing characteristics of the target word were similar to those of preceding constituents. This suggests that listeners rapidly develop expectations about aspects of the acoustic manifestation of word boundaries on the basis of perceived prosodic patterning within an utterance.

Perceptual expectations may include not only aspects of segmental realization and acoustic signatures of word boundaries but also acoustic characteristics associated with syllable prominence. In English, lexical stress is signaled by both vowel quality (specifically, unstressed syllables tend to have reduced vowels) and suprasegmental acoustic cues such as higher pitch and amplitude and increased duration (e.g., Cutler, Dahan, & van Donselaar, 1997). In conversational English, approximately 90% of content words have initial stress (Cutler & Carter, 1987). This distributional bias supports a well-documented preference to posit word boundaries prior to stressed syllables (Cutler & Norris, 1988). For example, listeners more frequently misperceive phrases like "she'll officially" as "Sheila Fishley" than phrases like "in closing" as "enclosing" (Cutler & Butterfield, 1992).

Lexical stress is not only a useful cue for segmentation but also a potentially informative cue for spoken-word recognition more generally. For instance, cues to lexical stress might facilitate discrimination between lexical alternatives with lexical competitors that phonemically overlap at onset but differ in the stress of their initial syllables, such as *jury* and *giraffe*. Indeed, this is the case in Dutch, in which lexical stress is conveyed primarily via suprasegmental cues as opposed to vowel quality. Visual-world studies indicate that stress cues influence the dynamics of competition between Dutch lexical alternatives like *DEcibel* and *deCIsie* prior to phonemic disambiguation (Reinisch, Jesse, & McQueen, 2010), suggesting that Dutch listeners can use suprasegmental stress information incrementally as words unfold (see also Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; van Donselaar, Koster, & Cutler, 2005). Previous work investigating the role of suprasegmental stress cues in English, however, has yielded equivocal results. Vowel quality shifts (e.g., the contrast between the full vowel in the initial syllable of *CONflict* and the reduced vowel in the initial syllable of *conFLICT*) have strong effects on the speed and accuracy of spoken-word recognition (e.g., Fear, Cutler, & Butterfield, 1995). Suprasegmental cues alone, on the other hand, appear to have weaker effects or effects that are only detectable in certain situations. For example, cross-splicing a vowel from an unstressed syllable into a strong syllable generally reduces acceptability ratings only when the weak vowel is reduced (Fear et al., 1995). Further, minimal pairs with differing stress but similar vowel quality, like *DIScount* and *disCOUNT*, pattern like homophones in a cross-modal priming task (Cutler, 1986). However, suprasegmental stress cues do influence the degree to which word fragments like *mus-* prime words with initial primary stress (e.g., *music*) versus words whose initial syllables bear secondary or null stress (e.g., *museum*; N. Cooper, Cutler, & Wales, 2002; Mattys, 2000). In addition, stress mispronunciations involving only suprasegmental characteristics (e.g., *fab-U-lous*) result in longer response latencies in shadowing and lexical-judgment tasks relative to correctly pronounced stresses (Slowiaczek, 1990). We suggest that the generally mixed patterns of results in the literature may have resulted in part from examining the recognition of words in isolation, with little or no context in which stress cues could be interpreted relationally.

The percept of utterance-level suprasegmental patterning or rhythmicity in English arises from alternations between stressed and unstressed syllables (reviewed in Peelle & Davis, 2012). English listeners tend to hear stressed syllables as perceptually isochronous (i.e., occurring at regular intervals) even when the acoustic basis of these perceived regularities eludes characterization (Lehiste, 1977). The tendency for metrically prominent syllables to be associated with recurring sequences of pitch accents (Couper-Kuhlen, 1993; Pierrehumbert, 2000) may likewise contribute to perceived suprasegmental patterning within an utterance.

These observations raise the possibility that perceived rhythmicity in speech informs listeners' expectations about the locations and acoustic characteristics of stressed syllables within upcoming words. Judgments about the locations of stressed syllables in German words are modulated by surrounding prosody in off-line tasks (Niebuhr, 2009). Further, listeners are faster to process information within syllables that are expected to bear stress on the basis of intonational or metrical patterns in the preceding context (Cutler, 1976; Pitt & Samuel, 1990). For example, Cutler (1976) spliced neutral recordings of monosyllabic words containing a target phoneme (e.g., /d/ in *dirt*) into two versions of a carrier utterance: one in which the target word bore contrastive focus (e.g., *She managed to remove the DIRT from the* rug [*but not the berry stains*]) and one in which another sentence element bore contrastive focus (e.g., *She managed to remove the dirt from the RUG* [*but not from their clothes*]). Listeners were faster to detect the word-initial target phoneme when the neutral word was spliced into a contrastively focused position such that preceding intonation predicted a high degree of prominence on the target word. These results suggest that listeners focus their attention on speech material that they expect to be prominent on the basis of preceding prosodic information.

Further evidence that contextual information influences the perception of suprasegmental stress cues comes from experiments by Reinisch, Jesse, and McQueen (2011) showing that perception of duration as a cue to lexical stress in Dutch is dependent on context speech rate. Reinisch et al. (2011) presented participants with word fragments embedded within a rate-manipulated carrier phrase and asked them to select one of two lexical alternatives: one with initial primary stress (e.g., *'alibi*) and one whose primary stress fell on the second or third syllable (e.g., *a'linea*). The results showed that the speech rate of preceding context influenced how listeners interpreted the duration of the initial syllable as a cue to the stress pattern of the target word fragment. Participants were more likely to choose the alternative with initial stress when the speech rate of the carrier phrase was resynthetically sped up (such that the initial syllable of the target sequence was long relative to the speech rate of surrounding context) than when it was slowed down. This finding may reflect effects of overall rate of articulation on expectations about the duration of upcoming stressed versus unstressed syllables. However, there are two alternative explanations for this result. First, it is possible that these effects have a proximal locus, driven by the manipulation of material adjacent to the target syllable. For example, Newman and Sawusch (1996) found that perception of a target phoneme is influenced by speech-rate manipulations within an adjacent time window of limited duration, typically including one or two phonemes, but not by more distal manipulations (though Dilley & Pitt, 2010, and other related studies have observed distal effects of speech rate on the perception of highly coarticulated function words). Second, compressing speech rate may have decreased overall intelligibility and increased reliance on prior knowledge about the statistical likelihood that Dutch nouns begin with a stressed syllable, biasing listeners to perceive the onset of the target word as stressed.

In light of the relevance of stress information for lexical segmentation and processing, it is important to characterize the mechanisms by which listeners perceive stressed syllables in online speech processing. The equivocal effects of suprasegmental stress cues in English observed in previous studies raise the possibility that stress expectations based on distal prosody play a significant role in enabling listeners to successfully resolve stress information during online lexical segmentation and processing. Indeed, several studies supposedly examining perception of proximal stress cues (i.e., cues on or immediately surrounding the syllable of interest) have potentially confounded distal prosody and proximal stress (e.g., Soto-Faraco et al., 2001; van Donselaar et al., 2005). That is, each of these studies used naturally produced carrier phrases whose distal prosodic characteristics are likely to have differed across conditions; thus, distal prosody could have contributed to results that were attributed to proximal stress characteristics. The role of distal prosody in the perception of lexical stress in English therefore remains an important unresolved question.

We conducted two experiments using the visual-world paradigm (R. Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to test the hypothesis that expectations based on the suprasegmental characteristics of stressed versus unstressed syllables in preceding utterance context influence the perception of suprasegmental cues to lexical stress downstream. Both experiments investigated the perception of English word pairs with phonemically overlapping initial syllables but different initial stress patterns, such as *jury* and *giraffe* (whose initial syllables overlap segmentally despite differing orthographically). These word pairs are referred to throughout as *alternate-stress cohorts*. Experiment 1 verified that alternate-stress cohorts compete for recognition during spoken-word recognition. Experiment 2 further demonstrated that fundamental frequency (F0) and syllable timing patterns across material preceding the target word influenced the dynamics of lexical competition of alternate-stress cohorts. When the suprasegmental characteristics of the initial syllable of the target word were more similar to preceding stressed syllables than unstressed syllables, participants were initially biased to fixate pictures of initially stressed cohorts than pictures of initially unstressed cohorts. This initial bias was not observed, however, when the suprasegmental characteristics of the initial syllable of the target word were more similar to preceding unstressed syllables than stressed syllables. These findings suggest that expectations about the acoustic realization of upcoming speech include information about expected phonetic characteristics of metrically stressed and unstressed syllables and that these expectations constrain the initial interpretation of lexical stress cues during spoken-word recognition.

Experiment 1

The goal of Experiment 1 was to establish that alternate-stress cohorts compete for recognition when segmental and suprasegmental stress cues on the initial syllable of a spoken target word are ambiguous. The use of materials with ambiguous stress cues facilitated examination of distal prosody effects in Experiment 2. We confirmed that stress cues in our materials were perceived as ambiguous by comparing patterns of fixations elicited by the first syllable of initially stressed alternatives (e.g., *jur-* from *jury*) versus the first syllable of initially unstressed alternatives (e.g., *gir-* from *giraffe*). In addition, given the strong tendency for English nouns to have initial stress (e.g., Cutler & Carter, 1987), it was important to determine whether listeners exhibit initial biases toward lexical alternatives with initial stress. The absence of a strong bias was a prerequisite for testing the effect of preceding prosody on the dynamics of lexical competition between alternate-stress cohorts (Experiment 2), because a strong bias could overwhelm consideration of initially unstressed alternatives in the initial moments of processing.

Method

Participants. Sixteen University of Rochester students took part in the experiment. All were native English speakers with normal hearing and normal or corrected-to-normal visual acuity and received $7.50 for participating.

Materials. The 24 test stimuli used in the experiment were spoken sentences containing either a word beginning with a strong–weak (SW) stress pattern (e.g., *jury*) or a word with a phonetically similar onset beginning with a weak–strong (WS) pattern (e.g., *giraffe*; see Appendix). SW and WS target words did not differ significantly in log-transformed lexical frequency, paired $t(23) = -0.24$, $p > .10$, as indexed by the *Corpus of Contemporary American English* (Davies, 2008). The preceding utterance context contained at least two disyllabic words with initial stress followed by one or two monosyllabic words (e.g., *Heidi sometimes saw*). This distal context was followed by another monosyllabic

word containing a full vowel (e.g., *that*) followed by the target word. Whereas the context preceding both SW and WS target words was the same, the sentence material following the target word differed for purposes of semantic coherence (e.g., *jury leaving the courthouse* vs. *giraffe in the city zoo*).

Meredith Brown recorded multiple tokens of all critical and filler sentences as 44.1 kHz waveform audio files, producing each token with minimal F0 excursions and slight F0 declination. The initial syllables for each alternate-stress cohort pair were pronounced as similarly as possible (e.g., with respect to vowel quality and suprasegmental stress cues). The speech editor Praat (Boersma & Weenink, 2010) was used to alter the perceived rhythm of the context material by lengthening the rime of the first monosyllabic word following the disyllabic words in the distal context (e.g., *saw* in *Heidi sometimes saw that . . .*), subsequently referred to as the *temporal-manipulation syllable*. In particular, the duration of the rime was increased to 150% of the mean duration of the preceding four and following two intervocalic intervals (i.e., the intervals spanning the vowel onsets of successive syllables). The purpose of this manipulation was to minimize the extent to which the rhythmic properties of the context favored SW alternatives over WS alternatives, or vice versa. The choice of a factor of 150% was motivated by the fact that it is halfway between a 1:1 and a 2:1 ratio of syllable durations—that is, it corresponds to a ratio of 1.5:1. Thus, this lengthening factor was expected to result in a disrupted rhythm that lacked clear beat structure for syllables following the temporal-manipulation syllable. To highlight the between-items variance, consider the example six-syllable stimulus context *Heidi sometimes saw that.* Here, the first, third, and fifth syllables are expected to carry sentence-level stress, giving rise to an SWSWSW rhythmic structure; thus, an SW target word (e.g., *jury*) would be more consistent with the foregoing metrical pattern than a WS target word (e.g., *giraffe*). In contrast, for the seven-syllable stimulus context *Marcus really wants that new . . .*, the first, third, fifth, and seventh syllables are expected to carry sentence-level stress, giving rise to an SWSWSWS rhythmic structure; thus, a WS target word would be more consistent with preceding prosody in this particular context. Lengthening the temporal-manipulation syllable to 150% of the duration of surrounding syllables should disrupt the perceived rhythmic structure of metrically prominent syllables in a way that should reduce this between-items source of variance, permitting us to focus on lexical competition between SW and WS alternatives in contexts whose timing properties do not strongly support one alternative or the other. This manipulation was performed for all critical items; that is, the duration of a particular item's temporal-manipulation syllable was identical across conditions.

Two recordings of each sentence were split into two fragments at the end of the first syllable of the target word, at a point in the waveform with zero amplitude (to avoid amplitude discontinuities in the cross-spliced materials). Fragments from different recordings were recombined to create *identity-spliced* and *cross-spliced* versions of the sentences corresponding to the SW target word (e.g., *jury*) and WS target word (e.g., *giraffe*). This manipulation allowed us to examine the extent to which we were successful in producing SW and WS words with minimal suprasegmental and segmental stress cues on the initial syllable. This was important because other factors might influence lexical competition between SW and WS alternatives, such as the predominance of nouns with initial stress in English. The first fragment of an SW recording (e.g., *Heidi sometimes saw that jur-*) was spliced together with (a) the second fragment of another SW recording (e.g., *-y leaving the courthouse*) to create an identity-spliced SW item and (b) the second fragment of a WS recording (e.g., *-affe in the city zoo*) to create a cross-spliced WS item. Likewise, the first fragment of a WS recording (e.g., *Heidi sometimes saw that gir-*) was spliced together with each of the second fragments used for the SW items to create (a) a cross-spliced SW item and (b) an identity-spliced WS item, respectively.

On each trial, four colored clip art pictures were presented in fixed positions within a grid (see Table 1 for a full list of pictures used in critical trials, and Figure 1 for an example display). On experimental trials, two pictures depicted each of the alternate-stress cohorts. The remaining two distractor images were selected such that they were distinct from the two potential target pictures with respect to visual and semantic properties and their labels' phonological properties.

Critical test stimuli were intermixed with 48 filler trials for which the visual display contained two pictures whose labels had a range of phonological relations. Fillers were included to minimize the possibility that participants might develop expectations about likely targets on the basis of the names of the pictures in the display and notice the experimentally relevant phonological relations between them (i.e., the contrast in initial stress for experimental items). In 33 filler trials, the related stimuli were polysyllabic words containing onset-embedded words, like *antlers* and *ant*; in 10, they were cohort competitors with the same stress pattern, like *candy* and *cannon*, and in five, they were cohort competitors with different stress patterns but marked vowel quality differences, like *balloon* and *bullets*. The target words used in filler

Table 1
List of Pictures Displayed in Each Critical Trial

SW target	WS target	Distractors	
alligators	alpacas	cookies	batteries
bunnies	bananas	keys	snowshoes
cantaloupe	canteen	anchor	baby
cashew	cashier	broom	scissors
compass	computer	rose	bridge
corset	corsage	ax	lantern
eagle	eclipse	bottle	cart
gokart	goatee	bell	matches
harpist	harpoon	boots	toucans
jury	giraffe	chimney	fiddle
lizard	lasagna	baseball	chest
magnet	magnolia	camera	palace
mermaid	marina	waffles	chess
mushroom	machine	cupboard	house
mustard	masseuse	tulips	clock
peanut	piano	grass	mask
pirate	pipette	bamboo	wallpaper
pulley	police	carton	cheesecake
racket	raccoon	pillow	canoe
Reese's	receipt	kite	bucket
tortoise	tornado	fairy	clown
trafficlight	trapeze	hive	cow
umpire	umbrella	collie	robot
volume	volcano	cabin	grill

Note. SW = strong–weak initial stress pattern; WS = weak–strong initial stress pattern.

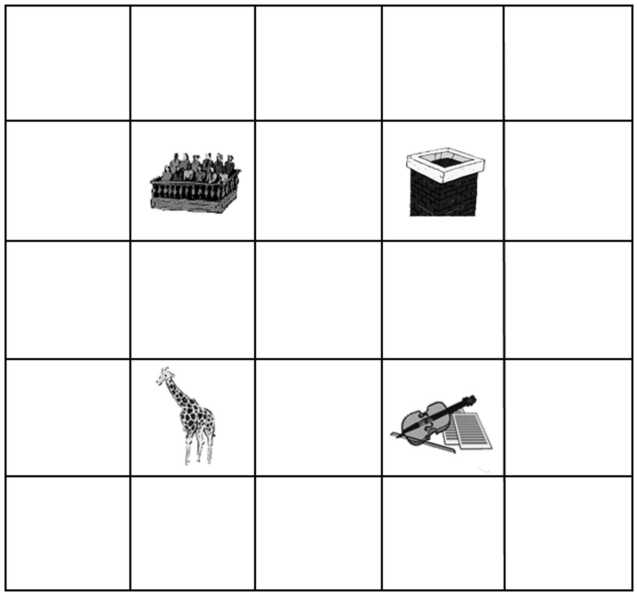

Figure 1. Example of the visual displays in Experiments 1 and 2.

trials comprised an equal number of SW, WS, and monosyllabic words. In half of the filler trials, the target word was one of the two phonologically related words. The sentences in eight filler trials were identity-spliced between the first and second syllables of the target word, whereas the rest were spliced at some point in the preceding context.

Procedure. Throughout the experiment, eye movements were recorded using a head-mounted SR Research (Mississauga, Ontario, Canada) EyeLink II system sampling at 250 Hz, with a drift-correction procedure performed every five trials. The eye tracker was fitted and calibrated at the start of the experiment.

The experiment was divided into three phases. In the picture-exposure phase, each of the 304 clip art pictures used in the experiment was presented to the participant in the center of the screen, with a corresponding label printed underneath, for a minimum of 3 s. Participants proceeded at their own pace through the sequence of picture–label pairs by pressing the space bar. Participants were preexposed to the pictures and their names to encourage them to identify the pictures as intended.

The eye-tracking phase began immediately following the picture-exposure phase. Each trial started with the presentation of a visual display containing four pictures from the picture-exposure phase. On critical trials, two of these pictures corresponded to the phonologically related SW and WS words. After 500 ms of display preview, participants heard a spoken sentence, and their task was to click on the picture mentioned in the sentence. They received no feedback on their performance.

Two lists of stimuli were constructed by randomizing the positions of the images on the screen within each trial and pseudorandomizing the order of trials within the list such that experimental trials were nonadjacent. Within each list, target stress pattern (SW, WS) was crossed with splice condition (identity-spliced, cross-spliced). Each list consisted of 12 items with SW target words and 12 with WS target words, an equal number of which were cross-spliced versus identity-spliced. The assignment of items to each condition was counterbalanced across the two lists of pseudorandomized stimuli for a total of eight lists. Two participants were randomly assigned to each list. The experiment list began with four practice trials to familiarize participants with the task.

Immediately after the completion of the eye-tracking phase, we assessed participants' ability to generate the intended labels for both members of each stress-alternating word pair (48 pictures in total). The purpose of this posttest was to allow us to identify trials on which participants did not associate both critical pictures with their intended labels. Participants were given a sheet of paper on which the critical pictures were printed in color in a pseudorandom order, subject to the constraint that phonologically related pictures were separated by at least four intervening pictures. Participants were asked to verbally recall the label that had been associated with each picture during the exposure phase and to continue to the next picture if they did not successfully recall the label of a picture within approximately 3 s. Responses were considered correct if they preserved the phonemes and stress pattern across the initial two syllables (e.g., *jury*, *jury-box*, and *jury-members* were all considered correct responses for the picture of the jury, but *jurors* was not).

Analysis. For each 4-ms interval, eye gaze was coded either as being directed at one of the four grid cells containing the target, competitor, or distractor pictures (if eye-gaze coordinates fell within the coordinates of the grid cell containing the relevant picture; cf. Figure 1) or as being directed at some other position on the screen. Fixations were coded from the start of the utterance until the participant clicked on one of the four pictures. Proportions of fixations to the target, competitor, and distractor pictures on experimental trials were averaged across two windows of interest. The *early window* began 200 ms after target-word onset and ended 200 ms after the onset of segmental information distinguishing the target word from the competitor. The *late window* began at the end of the early window and ended 200 ms after the offset of the target word. The mean durations of the early and late windows were 375 ms and 388 ms, respectively.

For each experimental trial, we computed the ratio of the average target fixation proportion to the sum of the average target and competitor fixation proportions (*target–competitor ratios*[1]) within the two analysis windows. These values were then transformed using the empirical logit function (Cox, 1970), with the number of observations set to the number of 50-ms time intervals spanned by the analysis window. This procedure resulted in a dependent variable reflecting the relative proportion of fixations to the target picture compared with the competitor picture while processing the initial or final part of the target word on a given trial. Logit-transformed target–competitor ratios were analyzed using multilevel linear regression analysis using the lme4 package in R (Bates, Maechler, & Bolker, 2011; R Development Core Team, 2012). Fixed effects included target-word type (SW vs. WS), splicing condition (identity-spliced vs. cross-spliced), performance in the recall task (i.e., whether the participant correctly named both of the two alternate-stress cohorts in the picture-recall phase), analysis

[1] This ratio has also often been referred to in the literature as the *target-advantage ratio*. We favor the more descriptive term *target–competitor ratio* because the ratio relates target fixations to competitor fixations.

window (early vs. late), and interactions between these factors. All factors were sum coded. Each observation was weighted by the reciprocal of the variance (Barr, 2008). The initial model included by-participant and by-item random intercepts and slopes for target-word type, splice condition, analysis window, and their interactions. If the model containing full random effects failed to converge, random slopes were removed stepwise, starting with the highest order interaction terms accounting for the lowest proportion of variance in logit-transformed fixation proportions, until the resulting model converged. In addition, fixed effects were removed stepwise, and each reduced model was compared with the more complex model using the likelihood ratio test (Baayen, Davidson, & Bates, 2008). This procedure was also used to estimate *p* values for each fixed effect. Final models included only fixed effects that contributed significantly to model fit. We examined collinearity between fixed effects by calculating the condition number using the *kappa* function in R. The condition number was less than 10 for all models, indicating that collinearity was unlikely to have affected the reliability of model estimates for the fixed effects and their interactions.

Results and Discussion

Our analyses sought to investigate whether and to what extent SW and WS alternatives competed for recognition during the processing of our spoken stimuli. In addition, we examined whether the initial syllables of our materials were perceived as having ambiguous stress cues.

We excluded from analysis trials on which participants did not click on the target picture in the eye-tracking phase (2.3% of the data). The proportions of fixations to the target, competitor, and distractor pictures as a function of condition starting at the onset of the target word are shown in Figure 2. At the onset of the target word, listeners showed a tendency to fixate pictures corresponding to WS words (i.e., the competitor picture in the case of SW target words and the target picture in the case of WS target words) relative to pictures corresponding to SW words. Approximately 250 ms after the onset of the target word, fixation proportions to pictures depicting SW words began to diverge from distractor fixation proportions. At the same time, proportions of fixations to pictures depicting WS words also rose slightly. This rise in fixations had approximately the same magnitude and time course for SW and WS target words. At approximately 600 ms after target-word onset, participants started to favor the target picture, as reflected by steep increases in proportions of fixations to target pictures and decreases in proportions of fixations to competitor pictures for both SW and WS target words.

The results of multilevel linear regression analyses of logit-transformed target–competitor ratios are presented in Table 2. During the processing of the target word, the proportion of fixations to the target relative to the proportion of fixations to the competitor did not differ as a function of the lexical-stress pattern of the target word. Only window of analysis contributed significantly to model fit, with lower target–competitor ratios in the early analysis window than in the late window. Neither target-word stress pattern, nor splicing condition, nor performance in the recall task (or their interactions) contributed significantly to model fit after taking into account by-participant and by-item random intercepts and slopes. Most relevant, target–competitor ratios did not differ significantly for SW or WS target words (untransformed ratios averaged across both analysis windows were 0.54 and 0.60, respectively) or for identity- versus cross-spliced target words (the untransformed mean ratio was 0.57 in both cases). Thus, hearing the initial sounds of a target word with ambiguous stress cues resulted in similar proportions of looks to SW and WS candidates, whether the target word was an SW word or a WS word.[2] Further, the time course of competition between SW and WS candidates was unaffected by whether the initial syllable of the target word was excised from an SW word or a WS word.

Taken together, the fact that participants considered both alternate-stress cohorts for recognition to a similar degree across splicing conditions suggests that we succeeded in creating materials with ambiguous stress cues on the initial syllable of the target word. Crucially, consideration of WS alternatives did not differ significantly in magnitude from consideration of SW alternatives, indicating that the predominance of nouns with initial stress in English did not favor an interpretation of the onset of the target word as corresponding to the SW alternative. In light of the strong skew toward SW nouns in the lexicon, it may seem surprising that participants did not exhibit a stronger bias toward SW candidates relative to WS candidates and, in fact, exhibited a tendency to fixate pictures corresponding to WS pictures prior to the processing of the speech signal. It is possible that the acoustic realization of the initial syllable of both SW and WS target words was somewhat more characteristic of an unstressed syllable than a stressed syllable, counteracting a prior bias toward interpreting the onset of the noun as stressed. We return to the interaction between prior distributional biases toward SW alternatives and suprasegmental (and segmental) stress cues in the General Discussion. However, for our purposes, having materials that elicited similar initial looks to SW and WS alternatives put us in a good position to investigate effects of distal prosody on the time course of interpretation of SW and WS words.

Experiment 2

We examined whether expectations based on the suprasegmental characteristics associated with stressed and unstressed syllables early in an utterance influence the perception of proximal suprasegmental cues to lexical stress within words encountered downstream. The timing and pitch contours of the stimuli from Experiment 1 were acoustically manipulated such that stressed syllables in the preceding context were roughly isochronous and all had similar pitch characteristics (i.e., for an F0 contour alternating between F0 maxima and minima, all stressed syllables within a stimulus occurred at maxima or else all occurred at minima). We hypothesized that this context manipulation would bias listeners to perceive as lexically stressed downstream syllables whose timing and pitch characteristics were similar to preceding stressed syllables. We also hypothesized that listeners would perceive as lexically unstressed downstream syllables

[2] In addition to the nonsignificant interaction between target-word stress and analysis window, a multilevel regression analysis of data within the early analysis window alone further confirmed that neither stress nor splice condition significantly contributed to model fit during the processing of phonemically overlapping material ($\beta = 0.16$, $SE = 0.30$, $p > .10$).

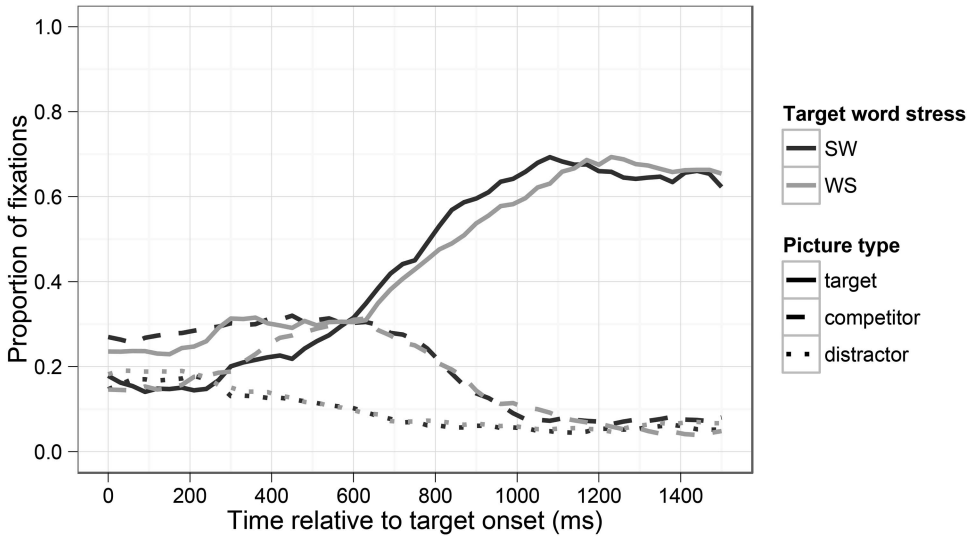

Figure 2. Proportions of fixations to the target, competitor, and averaged distractors for strong–weak (SW) and weak–strong (WS) items in Experiment 1 relative to the onset of the target word.

whose timing and pitch characteristics were similar to preceding unstressed syllables.

Method

Participants. Thirty-two University of Rochester students meeting the inclusion criteria for Experiment 1 took part in the experiment.

Materials. The stimuli were created by manipulating prosodic characteristics of the identity- and cross-spliced sentences from Experiment 1 using the pitch synchronous overlap-and-add algorithm in Praat (Moulines & Charpentier, 1990). Two versions of each item were created in which syllable timing and pitch contours varied across the distal context preceding the target word; in contrast, the acoustic characteristics of the syllable preceding the target word (i.e., the proximal context) and all subsequent material were kept identical across conditions (see Figure 3). In the *SW-biasing condition*, we manipulated the pitch contour and timing attributes of distal context to support the initial interpretation of the target word as stress-initial (e.g., making listeners more likely to interpret the initial syllable of the target word as corresponding to that in *jury* as opposed to that in *giraffe*). Conversely, in the *WS-biasing condition*, we manipulated the distal context to bias listeners toward a WS interpretation (e.g., *giraffe* as opposed to *jury*).

In the SW-biasing condition, distal context syllables with lexical and sentence-level stress (e.g., *Heidi sometimes saw*) had the same F0 level as the initial syllable of the target word (e.g., *jury* or *giraffe*; cf. Figure 3). Whether the first syllable of the target word had low or high F0 varied between items.[3] The manipulation thus contrasted with that of Experiment 1, in which the pitch contour across target words was not manipulated and had little pitch excursion and slight overall declination (as did the surrounding utterance context). The pitch contour across the target word for each particular item in Experiment 2 was kept the same across conditions and across lists. For instance, in the SW-biasing condition, items with high F0 on the initial syllable of the target word also had high F0 on all metrically prominent syllables in preceding context; likewise, items with low F0 on the initial syllable of the target word had low F0 on all metrically prominent syllables in preceding context. F0 manipulations involved removing nonvocalic pitch points and then alternating between shifting vocalic pitch points within each syllable up by 35 Hz or down by 25 Hz, alternating between successive syllables. Nonvocalic pitch points were removed such that the F0 between vocalic regions was linearly interpolated. This manipulation preserved natural microvariation and F0 declination from the original recording while imposing salient periodic alternations onto the F0 contour. These alternating F0 manipulations were performed for all syllables through at least the second syllable following the offset of the target word.

In addition, the duration of the temporal-manipulation syllable in the SW-biasing condition was either shortened or lengthened

Table 2
Parameters of Final Linear Regression Model of Logit-Transformed Target–Competitor Ratios in Experiment 1

Predictor	β	*SE*	*t*	*p*
Intercept	−0.22	0.11	−1.92	
Analysis window = late	0.39	0.09	4.28	<.0001

Note. The final model included random intercepts and random slopes as specified in the model formula. The kappa of the final model was 1.00. Formula: lmer (value ~ region + (1 + stress + splice + window + stress:splice | subj) + (1 + stress + splice + window + stress:splice + splice:window | item), data = expt1data, weights = 1/wts).

[3] Although high pitch accents are more common in English, low pitch accents are also felicitous and attested, and distal prosody effects have previously been documented using target words with low pitch accents on prominent syllables (e.g., Brown et al., 2011; Dilley et al., 2010; Dilley & McAuley, 2008).

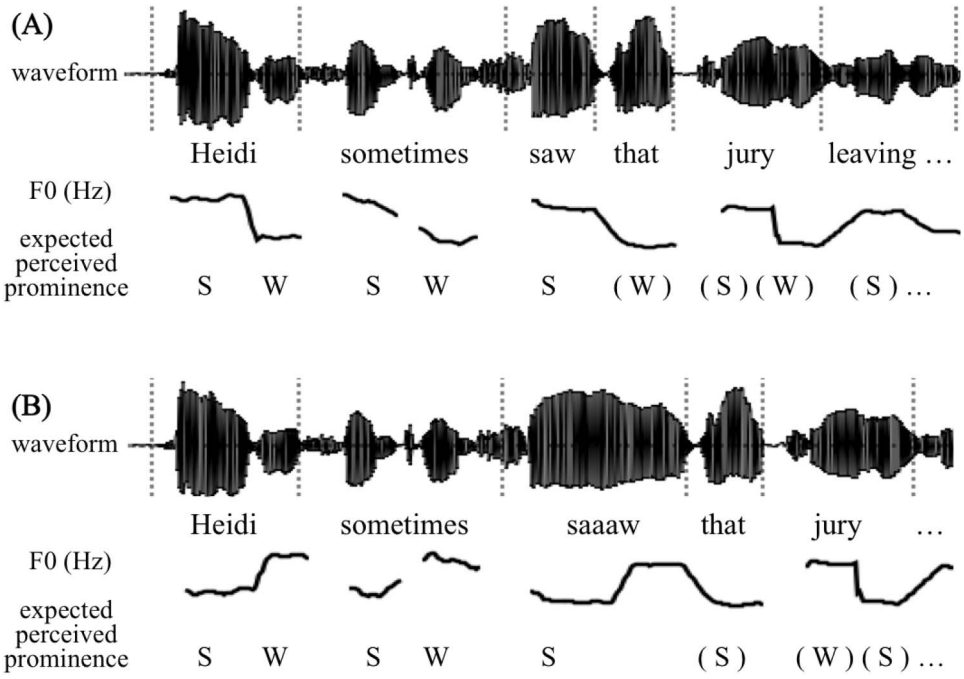

Figure 3. Explanation of the resynthesis methods used to perform the distal prosody manipulation in the SW-biasing prosody condition (A) and the WS-biasing prosody condition (B). The F0 of the distal context syllables (e.g., *Heidi sometimes saw*) and the duration of the temporal-manipulation syllable (e.g., *saw*) were manipulated to encourage (A) or discourage (B) the perception of a stressed syllable following the proximal context syllable (e.g., *that*). The acoustic properties of stimuli were held constant from the onset of the proximal context syllable through the end of the utterance. SW = strong–weak initial stress pattern; WS = weak–strong initial stress pattern.

such that the first syllable of the target word fit into the approximately regular temporal pattern established by prominent syllables in distal context. Whether this regularity was achieved using a short or long temporal-manipulation syllable depended on the structure of the preceding context: The syllable was short when the distal context ended in one monosyllabic word (e.g., *Heidi sometimes saw* [*that jur- . . .*]) and long when it ended in two (e.g., *Marcus really wants that* [*new mer- . . .*]). Short temporal-manipulation syllables were created by compressing the syllable's rime such that the duration between its vowel onset and the following vowel onset matched the mean duration of the following two intervocalic intervals, following Dilley and McAuley (2008; see also Figure 3). Short temporal-manipulation syllables thus conformed to a 1:1 duration pattern compared with the average of the following two syllable durations. Long temporal-manipulation syllables were created by expanding the rime such that the duration between its vowel onset and the following vowel onset was twice the mean duration of the preceding four intervocalic intervals. Long temporal-manipulation syllables thus conformed to a 2:1 duration pattern compared with the average of the preceding four syllable durations. These small whole-number duration ratios were expected to induce distinct patterns of downstream rhythmic expectations (e.g., Large & Jones, 1999) such that the initial syllable of the target word would occur in a position expected to be metrically prominent.

In the WS-biasing condition, stressed syllables in preceding context had the opposite relative F0 level from the initial syllable of the target word. For example, if the initial syllable of the target word had a high F0 level relative to the syllables immediately preceding and following it, stressed syllables in preceding context would have low F0 and unstressed syllables would have high F0 (see Figure 3). In addition, the duration of the temporal-manipulation syllable was altered such that the first syllable of the target word was out of phase with the quasi-regular timing of stressed distal-context syllables. When the distal context ended in one monosyllabic word (e.g., *Heidi sometimes saw* [*that jur- . . .*]), the temporal-manipulation syllable was long, conforming to a 2:1 duration pattern with respect to the average of the preceding four syllable durations. When it ended in two monosyllabic words (e.g., *Marcus really wants that* [*new mer- . . .*]), the temporal-manipulation syllable was short, conforming to a 1:1 duration pattern respect to the average of the following two syllable durations. These manipulations were expected to induce different patterns of downstream rhythmic expectations from the SW-biasing condition. Specifically, in the WS-biasing condition, the initial syllable of the target word would occur in a position expected to be metrically nonprominent with respect to preceding context.

Predictions. SW-biasing prosody was predicted to make listeners more likely to perceive the initial syllable of the target word as lexically stressed (e.g., biasing them to perceive it as the first syllable of *jury* as opposed to that of *giraffe*). For stimuli containing an SW target word (e.g., *jury*), this bias was predicted to decrease the initial proportions of fixations to the WS competitor (*giraffe*) elicited by the target word relative to the SW target (*jury*).

For stimuli containing a WS target word (e.g., *giraffe*), this bias was predicted to increase the initial proportions of fixations to the SW competitor (*jury*) relative to the WS target (*giraffe*).

WS-biasing prosody was predicted to make listeners more likely to perceive the initial syllable of the target word as unstressed. For stimuli containing an SW target word (e.g., *jury*), this bias was predicted to increase the initial proportions of fixations to the WS competitor (*giraffe*) on hearing the target word relative to the SW target (*jury*). For stimuli containing a WS target word (e.g., *giraffe*), this bias was predicted to decrease the proportions of fixations to the SW competitor (*jury*) relative to the WS target (*giraffe*).

Procedure. The procedure was the same as for Experiment 1. The stimulus lists used in Experiment 2 were the same as those used in Experiment 1 except for the addition of the context-prosody manipulation. Context prosody was crossed with target-word stress and splice conditions such that 12 of the 24 critical items had SW-biasing prosody (half of which had SW target words) and 12 had WS-biasing prosody (half of which had SW target words). The assignment of items to each condition was counterbalanced across the two lists of pseudorandomized stimuli for a total of 16 lists. Two participants were randomly assigned to each list.

Analyses. General analysis methods were the same as for Experiment 1, with two exceptions: (a) Context prosody and its interactions with other factors were added to the fixed and random effects structure of the initial regression models, and (b) splice condition and its interactions were removed from the models because this factor was no longer of primary interest and because splice condition did not contribute significantly to model fit in Experiment 1. Thus, the fixed effects of the initial model included target-word stress (SW vs. WS), context prosody (SW-biasing vs. WS-biasing), recall performance (both target and competitor correctly named vs. one or both incorrectly named), window of analysis (early vs. late), and their interactions. Random slopes included target-word stress, context prosody, window of analysis, and interactions. The procedure for determining the structure of the final model was the same as in the analysis of the data from Experiment 1.

Results and Discussion

As in Experiment 1, we excluded trials on which participants clicked on the incorrect picture (3.6% of the data). The proportion of fixations to target, competitor, and distractor pictures as a function of condition starting at the onset of the target word are shown in Figure 4. Between 300 and 400 ms after the onset of SW target words (see Figure 4, top), fixations to the target and competitor pictures started to diverge from fixations to distractor pictures. Crucially, SW-biasing distal prosody elicited an earlier rise in target fixations and lower proportions of competitor fixations than WS-biasing distal prosody. These effects persisted for approximately 800 ms. For WS words (see Figure 4, bottom), SW-biasing prosody had opposite effects, eliciting a somewhat later rise in target fixations and higher proportions of competitor fixations than WS-biasing prosody. However, the effects of context prosody on the interpretation of WS words appeared to be smaller and later occurring than the effects of context prosody on the interpretation of SW words.

As in Experiment 1, we performed multilevel linear regression analyses on logit-transformed target–competitor ratios (i.e., average target fixation proportions divided by the sum of averaged target and competitor fixation proportions) across the early and late windows of analysis (which were the same as in Experiment 1). Predictors in the model included target-word stress, context prosody, analysis window, and performance in the recall task. We predicted that distal prosody would interact with target-word stress. Specifically, we predicted that target–competitor ratios during the processing of SW target words would be higher in the SW-biasing condition than in the WS-biasing condition, whereas target–competitor ratios during the processing of WS target words would be lower in the SW-biasing condition than in the WS-biasing condition.

Analysis of target–competitor ratios yielded a significant interaction between distal prosody and target-word stress (see Table 3). This interaction was significant above and beyond significant effects of analysis window and performance in the recall task. Separate examination of fixations in response to SW and WS target words revealed differences in the effect of context prosody (see Table 4). For SW target words, target–competitor ratios were significantly higher in response to SW-biasing prosody than in response to WS-biasing prosody (untransformed ratios averaged across both analysis windows = 0.59 and 0.47, respectively). Conversely, prosody did not have a significant effect on fixations in response to WS target words, although target–competitor ratios were numerically higher in response to WS-biasing prosody than in response to SW-biasing prosody (untransformed mean ratios = 0.57 and 0.52, respectively). Other interactions between factors did not contribute significantly to model fit and were not included in the final model.

The three-way interaction between prosody, stress, and analysis window was not included in the final model because it did not contribute significantly to the proportion of variance explained. The interaction between prosody and stress thus appears to have had similar effects before and after the availability of segmental information in the speech signal distinguishing the target word from the competitor. Separate analyses of target–competitor ratios in the early and late analysis windows confirmed that the interaction between prosody and stress was significant in both analysis windows (see Table 5).

Thus, patterns in F0 and syllable timing across preceding material distal to the target word influenced the dynamics of competition between SW and WS alternatives. On hearing the target word, participants were more likely to fixate SW alternatives when preceding stressed syllables had suprasegmental acoustic characteristics similar to the initial syllable of the target word, but they were not more likely to fixate WS alternatives when preceding unstressed syllables had suprasegmental characteristics similar to the initial syllable of the target word. Taken together, the results suggest that pitch and timing patterns associated with stressed and unstressed syllables in distal preceding context affected listeners' interpretation of the initial syllable of a SW target word.

Although the interaction between stress and prosody conditions was consistent with our predictions, we did not expect the effect of context prosody on fixation patterns during the processing of SW- versus WS-stress target words to be asymmetrical. Preceding prosody seemed to have much stronger effects on fixation patterns for

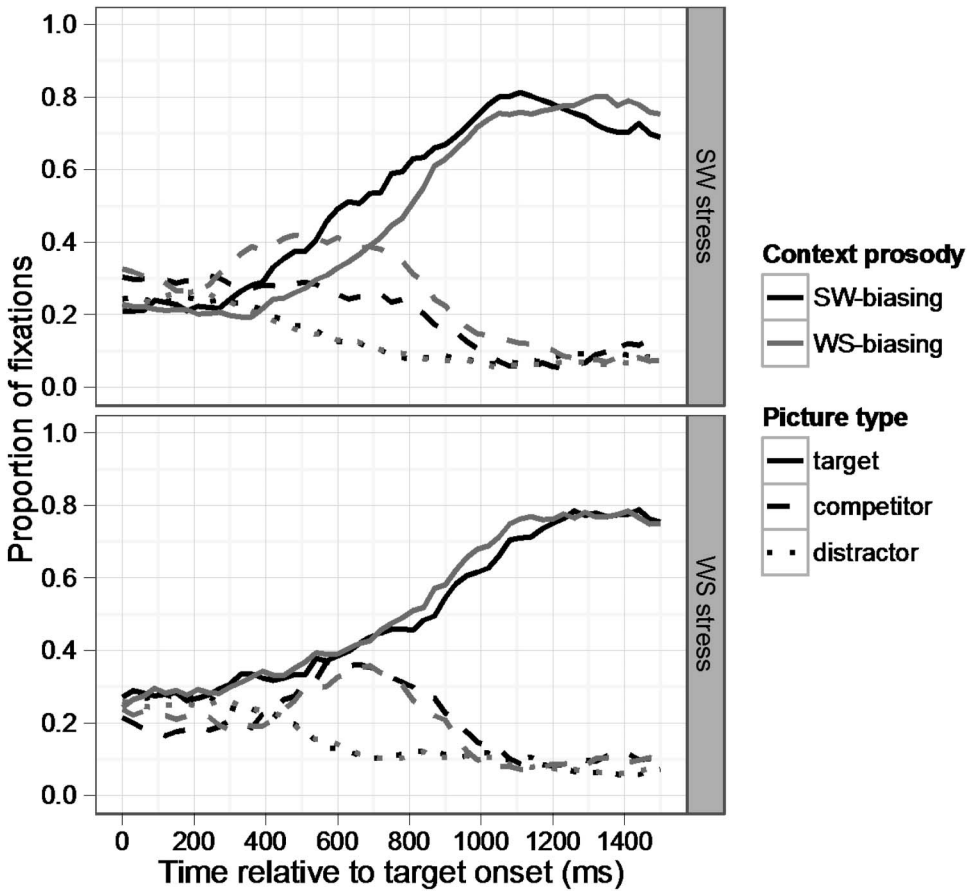

Figure 4. Proportions of fixations to the target, competitor, and averaged distractors in Experiment 2 for strong–weak (SW) items with SW-biasing versus WS-biasing prosody (top) and weak–strong (WS) items with SW-biasing versus WS-biasing prosody (bottom).

SW target words (e.g., *jury*) than for WS target words (e.g., *giraffe*).

We conducted a post hoc analysis of the data examining effects of the pitch contour across the target word as part of a response to a question from an anonymous reviewer about whether the relative pitch of stressed versus unstressed syllables influenced processing effects. The pitch contour across the target word varied across items such that half of the target words began with a low–high (LH) pitch contour (and were preceded by a monosyllabic word with high pitch), and half began with a high–low (HL) pitch contour (and were preceded by a monosyllabic word with low pitch). It was therefore possible to examine whether target-word pitch contour affected lexical competition or modulated the effects of preceding prosody on the processing of SW and WS words.

We conducted a multilevel linear regression analysis of logit-transformed target–competitor ratios following the same procedures as described for Experiment 1. Lexical stress (SW vs. WS), context prosody (SW-biasing vs. WS-biasing), analysis window (early vs. late), target-word pitch contour (HL vs. LH), and their interactions were included as fixed and random effects. The structure and parameters of the final model are shown in Table 6. The main finding of interest is a marginally significant three-way interaction between lexical stress, context prosody, and target-word pitch contour (see Figure 5). When the target word was associated with an LH pitch contour, the two-way interaction between lexical stress and context prosody was significant (see Table 7, bottom). Target–competitor ratios were higher for SW words in the SW-biasing prosody condition, whereas target–competitor ratios were higher for WS words in the WS-biasing prosody condition. However, when the target word began with an HL pitch contour, context prosody did not interact significantly with target-word stress (see Table 7, top). Thus, the asymmetrical effects of context prosody on the processing of SW versus WS words appear to have been mediated by differences in effects for words associated with LH versus HL pitch contours.

The observation that context prosody apparently had weaker effects on words associated with HL pitch contours may reflect an interaction between prior distributional knowledge and prosodic context. Because H^* is the most frequent pitch accent in American English (Dainora, 2001), stressed syllables are more often realized with high pitch than with low pitch. This, together with the tendency for word-initial syllables to be stressed (Cutler & Carter, 1987), may lead listeners to have a prior expectation for word-initial syllables with high pitch to be stressed. In comparison, listeners' prior expectations for word-initial syllables with low pitch are likely to be weaker or associated with more uncertainty,

Table 3
Parameters of Final Linear Regression Model of Logit-Transformed Target–Competitor Ratios in Experiment 2

Predictor	β	*SE*	*t*	*p*
Intercept	0.92	0.10	9.54	
Target-word stress = WS	0.12	0.14	0.83	*ns*
Prosody = WS-biasing	−0.09	0.09	−1.04	*ns*
Analysis window = late	0.18	0.06	2.77	<.01
Recall = successful	−0.14	0.06	−2.20	<.05
Stress × prosody	0.19	0.08	2.35	<.05

Note. The final model included target-word stress and prosody as predictors, even though these predictors were not significant, because the model included a significant interaction between target-word stress and prosody. The final model included random intercepts and random slopes as specified in the model formula. The kappa of the final model was 2.87. Formula: lmer[value ~ stress + prosody + window + recall + stress:prosody + (1 + stress + prosody + window + stress:prosody | subj) + (1 + stress + prosody + window + stress:prosody + stress:window | item), data = expt2data, weights = 1/wt]. WS = weak–strong initial stress pattern.

because word-initial stressed syllables with low pitch and word-initial unstressed syllables with low pitch are both less commonly encountered than word-initial syllables with high pitch.

Given these observations and assumptions, effects of preceding prosody on the processing of SW versus WS target words would be expected to be weaker for target words beginning with HL pitch contours than for words beginning with LH pitch contours. In the case of target words beginning with HL pitch, listeners have relatively strong distributional expectations that may make their interpretations of word-initial syllables with high pitch less dependent on preceding prosodic context than their interpretations of word-initial syllables with low pitch. This is an intriguing possibility that merits further investigation with more carefully controlled within-item manipulation of target-word prosody.

Table 4
Parameters of Final Linear Regression Model of Logit-transformed Target–Competitor Ratios in Response to Words With SW and WS Stress in Experiment 2

Predictor	β	*SE*	*t*	*p*
SW stress[a]				
Intercept	0.86	0.17	5.10	
Prosody = WS-biasing	−0.28	0.11	−2.50	<.05
Analysis window = late	0.22	0.08	2.77	<.01
Recall = successful	−0.19	0.09	−2.11	<.05
WS stress[b]				
Intercept	0.96	0.17	5.78	

Note. Random effects are specified in the model formulas. Kappa values were 1.16 for the SW-stress model and 1.00 for the WS-stress model. SW = strong–weak initial stress pattern; WS = weak–strong initial stress pattern.
[a] Formula: lmer[value ~ prosody + window + recall + (1 + prosody + window | subj) + (1 + prosody + window + prosody:window | item), data = exp2dataSW, weights = 1/wt]. [b] Formula: lmer[value ~ (1 + prosody + window + prosody:window | subj) + (1 + prosody + window + prosody:window | item), data = exp2dataWS, weights = 1/wt].

Table 5
Parameters of Final Linear Regression Model of Logit-Transformed Target–Competitor Ratios in the Early and Late Analysis Windows in Experiment 2

Predictor	β	*SE*	*t*	*p*
Early analysis window[a]				
Intercept	0.79	0.10	7.81	
Target-word stress = WS	0.13	0.14	0.92	*ns*
Prosody = WS-biasing	−0.10	0.08	−1.21	*ns*
Stress × prosody	0.19	0.09	2.02	<.05
Late analysis window[b]				
Intercept	1.10	0.12	9.12	
Target-word stress = WS	0.02	0.19	0.10	*ns*
Prosody = WS-biasing	−0.06	0.11	−0.55	*ns*
Recall = successful	−0.18	0.09	−2.04	<.05
Stress × prosody	0.23	0.09	2.45	<.01

Note. Random effects are specified in the model formulas. Kappa values were 1.03 for the early-window model and 1.16 for the late-window model. WS = weak–strong initial stress pattern.
[a] Formula: lmer[value ~ stress + prosody + stress:prosody + (1 + stress + prosody + stress:prosody | subj) + (1 + stress + prosody | item), data = expt2data.early, weights = 1/wt]. [b] Formula: lmer[value ~ stress + prosody + recall + stress:prosody + (1 + stress + prosody + stress:prosody | subj) + (1 + stress + prosody + stress:prosody | item), data = expt2data.late, weights = 1/wt].

The four-way interaction between lexical stress, context prosody, target-word pitch contour, and analysis window did not contribute significantly to model fit, indicating that the three-way interaction between stress, prosody, and pitch contour was statistically similar across the early and late analysis windows (see Table 6). Importantly, the two-way interaction between lexical stress and context prosody remained significant above and beyond the three-way interaction with target-word pitch contour, indicat-

Table 6
Parameters of Final Linear Regression Model of Logit-Transformed Target–Competitor Ratios in Experiment 2 With Target-Word Pitch Contour and Its Interactions Added as Fixed and Random Effects

Predictor	β	*SE*	*t*	*p*
Intercept	0.98	0.13	7.72	
Target-word stress = WS	0.13	0.12	1.11	*ns*
Prosody = WS-biasing	−0.06	0.08	−0.71	*ns*
Analysis window = late	0.27	0.06	4.68	<.0001
Recall = successful	−0.13	0.06	−2.09	<.05
Target-word pitch = LH	−0.04	0.11	−0.37	*ns*
Stress × prosody	0.20	0.07	2.70	<.01
Stress × target-word pitch	0.04	0.12	0.35	*ns*
Prosody × target-word pitch	0.05	0.09	0.58	*ns*
Stress × prosody × target-word pitch	0.12	0.07	1.65	<.10

Note. The final model included random intercepts and random slopes as specified in the model formula. The kappa of the final model was 1.26. LH = low-high. Formula: lmer[value ~ stress + prosody + window + recall + pitch + stress:prosody + stress:pitch + prosody:pitch + stress:prosody:pitch + (1 + stress + prosody + window + pitch + stress:prosody + stress:window + stress:pitch + prosody:pitch + window:pitch | subj) + (1 + stress + prosody + window | item), data = expt2data, weights = 1/wt]. WS = weak–strong initial stress pattern.

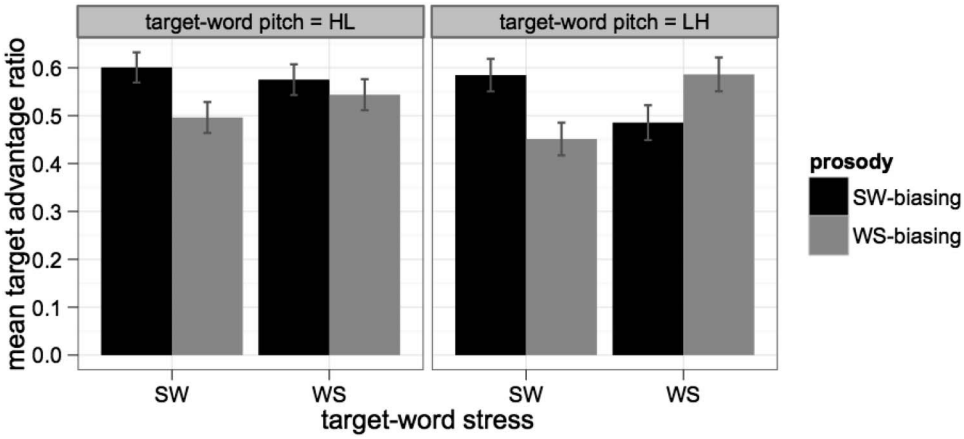

Figure 5. Mean untransformed target–competitor ratios by target-word stress and context prosody condition, averaged between 200 ms after the onset of the target word and 200 ms after the offset of the target word, for target words beginning with high–low (HL) pitch contours and for target words beginning with low–high (LH) pitch contours. Error bars indicate bootstrapped 95% confidence intervals.

ing that the predicted effect of primary interest cannot be fully explained away on the basis of target-word pitch contour effects.

General Discussion

Results from two visual-world experiments support the hypothesis that spoken-language processing involves expectations about metrical organization and the acoustic manifestation of lexical stress and that these expectations constrain the initial interpretation of suprasegmental stress cues during spoken word recognition. Thus, cues to lexical stress in English are not restricted to proximal cues such as a syllable's F0, duration, and amplitude. Rather, cues to lexical stress can also include sentence-level patterning associated with metrically strong and weak events in context.

The distal locus of the effects reported here demonstrates that the interpretation of prosodic cues to stress is rapidly tuned by recently encountered speech. This finding is consistent with observations that many prosodic cues are interpreted not in absolute terms but, rather, with respect to surrounding context. For example, the interpretation of cues to a prosodic boundary is dependent on the magnitude of cues to prosodic boundaries in preceding utterance context (Clifton, Carlson, & Frazier, 2006). Further, the perception of Mandarin Chinese and Cantonese tones is influenced by the overall pitch range of a carrier utterance (Fox & Qi, 1990; Wong & Diehl, 2003). Listeners' interpretation of rising intonational contours in Korean is likewise sensitive to both the position of the rising contour within the utterance and to the tonal properties of surrounding context (Kim, 2004).

The resynthetic manipulations of distal prosody and the ambiguous segmental and suprasegmental stress cues used in our experiments raise at least two questions about the source and generalizability of the effects reported here. First, the distal prosody manipulations that we used involved simultaneous alterations of both the F0 contour and temporal characteristics of preceding context. It is therefore an open question what the relative contribution is of each source of patterning to the interpretation of proximal stress cues. Previous work has demonstrated effects of distal prosody on lexical segmentation that are carried by any of several combinations of cues: pitch alone, duration alone, or a combination of pitch and duration (Dilley & McAuley, 2008; Morrill, Dilley, & McAuley, 2014). Thus, at least within the domain of lexical segmentation, effects are not dependent on a particular combination of distal prosodic characteristics. The extent to which distal effects on the interpretation of lexical stress are attributable to pitch patterning versus temporal patterning in preceding speech remains a topic for further investigation.

Second, in light of our findings, it is important to consider whether and how distal sources of information might contribute to

Table 7
Parameters of Final Linear Regression Model of Logit-Transformed Target–Competitor Ratios in Response to Words With HL and LH Target-Word Pitch Contours in Experiment 2

Predictor	β	*SE*	*t*	*p*
Target words with HL pitch[a]				
Intercept	1.02	0.13	8.06	
Analysis window = late	0.38	0.10	3.91	<.0005
Target words with LH pitch[b]				
Intercept	0.93	0.21	4.45	
Target-word stress = WS	0.18	0.20	0.92	*ns*
Prosody = WS-biasing	0.01	0.13	0.10	*ns*
Analysis window = late	0.17	0.9	2.04	<.05
Recall = successful	−0.19	0.09	−2.12	<.05
Stress × prosody	0.30	0.13	2.29	<.05

Note. Random effects are specified in the model formulas. Kappa values were 1.00 for the HL-pitch model and 1.19 for the LH-pitch model. HL = high–low; LH = low–high; WS = weak–strong initial stress pattern.
[a] Formula: lmer[value ~ window + (1 + stress + prosody + window + stress:prosody | subj) + (1 + stress + prosody + window + stress:prosody | item), data = expt2data.HL, weights = 1/wt]. [b] Formula: lmer[value ~ stress + prosody + window + recall + stress:prosody + (1 + stress + prosody + window + stress:prosody | subj) + (1 + stress + prosody + window + stress:window + stress:prosody | item), data = expt2data.LH, weights = 1/wt].

the use of stress cues in the perception of natural (i.e., nonlaboratory) speech. In English, vowel quality and suprasegmental cues are generally stronger cues to lexical stress than they were in our materials. However, these cues are probabilistic, and their realization is closely tied to contextual factors such as speaker identity and speech register (e.g., Sumner, Gafter, Kurumada, & Tice, 2014), giving rise to considerable acoustic variation in natural speech. Further, as previously noted, several studies have indicated that suprasegmental proximal cues to stress in English are quite confusable. For example, Mattys (2000) showed that primary and secondary stress are highly confusable when syllables are presented in isolation (e.g., [pra] from *prosecutor* vs. *prosecution*): Listeners distinguish the two just above chance level (54% correct). In addition, replacing the vowel in a syllable with primary or secondary stress with a vowel from an unstressed unreduced syllable via cross-splicing generally does not affect acceptability ratings for the spliced words, whereas replacing the same vowel with a vowel from an unstressed reduced syllable makes the cross-spliced words significantly less acceptable (Fear et al., 1995). The interpretation of proximal stress cues in natural speech situations is further complicated by the comparably noisy conditions in which natural speech is often encountered (Mattys, White, & Melhorn, 2005). The integration of multiple sources of distal and proximal information to interpret stress cues and other speech cues in context is therefore potentially quite valuable for robust language comprehension and communication.

Interpretation of Prosody as Probabilistic Inference

Most generally, our results are compatible with emerging approaches recasting aspects of perception and cognition as probabilistic inference about the sources of observed sensory input. Such approaches have been proposed for domains ranging from visual perception (e.g., Rao & Ballard, 1999) to causal reasoning (e.g., Tenenbaum, Kemp, Griffiths, & Goodman, 2011). So-called Bayesian generative models operate to predict the incoming sensory signal on the basis of the likelihoods of different possible causes and their prior probabilities and to explain the data on the basis of these predictions. For example, when inferring the cause of a cough, observers are more likely to infer that it is caused by a cold than by lung cancer (because of the higher prior probability of colds) or by heartburn (because of the higher likelihood that a cold would result in coughing; Tenenbaum et al., 2011). To the extent that the coughing is considered evidence of a cold, the estimated probability of other possible causes decreases (i.e., they are "explained away"). Further, the inference that the coughing results from a cold permits observers to predict other consequences of colds (e.g., sniffles) that may occur in the future.

Such data-explanation approaches provide a potential explanatory framework for the integration of disparate and temporally distributed cues in spoken-word recognition (and in language processing more generally; Farmer, Brown, & Tanenhaus, 2013). According to this view, listeners continuously update their inferences about the hidden causes underlying aspects of an utterance as the signal unfolds. These inferences inform fine-grained probabilistic expectations about how (aspects of) lexical alternatives are likely to be realized in context. These expectations percolate down through the levels of a hierarchically organized generative model. The model's levels of representation do not necessarily correspond to traditional formal levels of linguistic representation—such as semantics, syntax, and phonology—but they should generally represent progressively less abstract and/or more fine-grained perceptual information. Provisional hypotheses (e.g., lexical candidates) compete to explain the acoustic data, with the predicted acoustic realization of each alternative being evaluated against the actual acoustic properties of the input. At each point in time, as the signal unfolds, then, support for a particular lexical candidate is proportional to the degree of similarity between the predicted acoustic realization of that candidate (in the utterance context) and the actual acoustic signal, favoring lexical candidates whose predicted realizations are the most congruent with the acoustic signal.

This data-explanation framework provides straightforward accounts for effects of higher level or distal context on the interpretation of prosodic and acoustic cues during spoken word recognition. Consider, for instance, a recent study by Brown, Salverda, Gunlogson, and Tanenhaus (2015) examining effects of preceding discourse context on the interpretation of local acoustic cues to a prosodic boundary. Typically, segmental lengthening across the initial syllable of words like *hamster* increases lexical competition from the onset-embedded cohort competitor *ham*, because the segments of a word immediately preceding a prosodic word boundary (e.g., the rime of the word *ham*) tend to be lengthened (Salverda, Dahan, & McQueen, 2003). Indeed, the extent to which monosyllabic words like *cap* elicit competition from other monosyllabic (e.g., *cat*) versus polysyllabic (e.g., *captain*) cohort competitors depends on the size of the prosodic boundary following *cap*: When *cap* occurs at the end of a carrier utterance, *cat* is a stronger competitor than *captain*; when *cap* occurs in utterance-medial position, *captain* competes more strongly (Salverda et al., 2007).

Data-explanation approaches predict that listeners' use of proximal cues to a prosodic boundary should depend on aspects of the context in which the syllable is processed. When higher level inferences can provide a potential explanation for observed cues, such as segmental lengthening, data-explanation approaches predict a corresponding reduction in the estimated probability that segmental lengthening results from an upcoming prosodic boundary. Indeed, segmental duration is dependent on several factors other than preboundary lengthening. For example, whether and how a referent has been mentioned in prior discourse (its *information status*) is associated with its acoustic prominence: Referents that have not been mentioned previously or are referred to in new ways tend to have more acoustic prominence (and longer durations) than previously mentioned referents. Brown et al. (2015) demonstrated that information status modulates the extent to which segmental lengthening is interpreted as signaling an upcoming prosodic boundary on hearing instructions like "Now put the captain next to the circle." When *captain* had just been mentioned as the theme of a prior instruction (e.g., *Put the captain between the cap and the circle*), segmental lengthening elicited more looks to *cap* (because lengthening is consistent with both preboundary lengthening and prosodic prominence because of a change in in thematic role). However, when *captain* was instead the goal of the prior instruction (e.g., *Put the cap between the captain and the circle*), segmental lengthening had little effect on *cap* fixations: Although lengthening remained consistent with preboundary lengthening (favoring a *cap* interpretation), it was in-

consistent with repeated mention of *captain* in the same thematic role (disfavoring a *cap* interpretation). Therefore, inferences about the information structure of an utterance influence expectations about the acoustic characteristics of certain words within that utterance.

Likewise, results from the current study suggest that inferences about the association between recurring prosodic patterning and the metrical structure of an utterance can inform listeners' expectations about the prosodic characteristics of metrically prominent syllables downstream. The inference that the metrical structure of an utterance is associated with recurring prosodic patterning enables listeners to predict prosodic aspects of the acoustic realization of upcoming metrically prominent syllables. Because stressed syllables are much more likely to be metrically prominent than unstressed syllables, syllables that have prosodic characteristics expected to be associated with metrical prominence are more likely to be perceived as lexically stressed than syllables that do not have such prosodic characteristics. Most other models of spoken-word recognition have focused on how local acoustic features of speech activate lexical representations (e.g., McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008). Although these models have generally not focused on the contributions of prosody to this process, effects of proximal suprasegmental stress cues on lexical access are readily explained via feature-detection mechanisms similar to those proposed to detect segmental features. However, effects of distal prosodic context are hard to incorporate within these models, because they do not affect the local acoustic features associated with a currently processed word or sound. These effects are also difficult to explain within the framework of exemplar models, in which specific episodes or exemplars of spoken words are represented in memory in terms of their surface acoustic characteristics, and newly encountered exemplars are then classified by comparison with the set of stored exemplars (e.g., Goldinger, 1998; Hawkins, 2003). To explain the distal prosody effects obtained in this study and in previous studies on distal prosody (e.g., Dilley et al., 2010; Dilley & McAuley, 2008), listeners would not only have to maintain exemplar representations of the target words that include detailed contextual information spanning several syllables, they would also (implausibly) have previously encountered target words in prosodic contexts similar to the experimental materials. Instead, distal prosody effects, and their apparent integration with prior distributional knowledge, suggest a more flexible, dynamic processing mechanism that can simultaneously take into account multiple aspects of surrounding context.

Role of Prior Distributional Knowledge

In light of the data-explanation perspective, we return again to the null effect of target-word stress in Experiments 1 and 2. Given listeners' prior knowledge about the predominance of SW nouns in the English lexicon, one might expect at least a slight initial bias toward SW interpretations of an acoustic signal that is (to some extent) ambiguous between the onset of an SW and WS word. However, the time course of competition effects was similar across stress and splice conditions. Part of this may be attributable to differences in the way that knowledge about the distributional characteristics of stressed syllables within the English lexicon might affect segmentation versus lexical access. Listeners' knowledge that most English content words have initial stress confers clear advantages for lexical segmentation (Cutler & Norris, 1988). However, it may not be the case that listeners' prior knowledge about the distribution of stressed syllables in English affects lexical access in the same way. For example, consider a listener processing the initial sounds of a token of *jury*. In the absence of segmental or suprasegmental stress cues, the set of lexical candidates considered in response to these sounds will likely include *jury* and *giraffe* as well as other phonemically overlapping words (e.g., *jersey*, *journal*, *geranium*). It is likely that the set of words under consideration will include more SW candidates than WS candidates, given the predominance of initial lexical stress in English content words. However, this property of the *set* of words under consideration should not necessarily affect how strongly *individual* lexical candidates within this set are considered (e.g., *jury* vs. *giraffe*). Listeners' knowledge about the predominance of SW words would, thus, not necessarily be expected to contribute to differences in fixations to pictures representing SW and WS alternatives in our experiments.

In addition, at least two aspects of our design may have contributed to the null effect of target-word stress. First, as mentioned earlier, effort was taken to produce SW and WS alternatives with vowels of similar quality. The vowels in the initial syllables of SW target words were, thus, likely more centralized and reduced than listeners would expect them to be, particularly in light of the manner in which surrounding context material was produced (i.e., "careful" laboratory speech as opposed to casual speech; cf. Sumner et al., 2014). Even if listeners would normally have had a prior bias to interpret the onset of the noun as stressed, vowel reduction might have neutralized effects of such a bias. An additional intriguing possibility is that the relative frequency of WS target words among the set of critical items and fillers may have counteracted a prior bias toward SW interpretations over the course of the experiment. Although WS words are encountered much less frequently than SW words in natural language use (Cutler & Carter, 1987), they made up half of the critical target words and a third of the filler target words. In light of recent findings on rapid online adaptation to novel statistical distributions across multiple dimensions of language comprehension (e.g., Fine & Jaeger, 2013; Kleinschmidt & Jaeger, 2014; Kurumada, Brown, & Tanenhaus, 2014), it is possible that listeners' behavior in the present experiments reflected, in part, the statistics of the experimental environment. The effects of prior lexical knowledge, exposure to novel distributions of SW and WS words, and the online interpretation of prosodic cues to stress constitute an interesting avenue for further research.

Although prior knowledge about the relative prevalence of SW words in English did not manifest in an overall bias toward SW interpretations, a post hoc analysis of effects of target-word pitch contour on the interaction between lexical stress and context prosody in Experiment 2 suggested a complex interaction between prior distributional knowledge and context prosody. Specifically, context prosody had stronger effects on the interpretation of lexical stress when target words began with an LH pitch contour than when they began with an HL pitch. We interpreted this pattern of results as reflecting differences in the strength of prior stress expectations for word-initial syllables with high pitch versus low pitch. When listeners have relatively strong prior stress expectations (as for word-initial syllables with high pitch, which are likely

to be stressed), they may rely less on context prosody than when they have relatively weak prior stress expectations (as for word-initial syllables with low pitch). However, the manipulation of target-word pitch contour in our materials was incidental, and because the pitch contour of the target words did not vary between participants or conditions, it is also possible that the apparent interaction between stress, context prosody, and target-word pitch contour was an artifact of some other aspect of the materials. The interplay between prior distributional knowledge and effects of preceding prosody should therefore be investigated more explicitly in future work with more carefully controlled materials.

Neural Processing of Speech Rhythm

Another interesting set of questions concerns the neurophysiological basis of these distal prosody effects and, in particular, the role of dynamic neural oscillators in generating and maintaining rhythmic expectations during spoken-language processing. It is increasingly evident that synchronized cycles of excitation and inhibition within and between local populations of neurons are modulated by exposure to external rhythmic phenomena (including but not limited to speech and music), resulting in alignments between the phase of temporally regular stimuli and the phase of internal neural rhythms (e.g., Lakatos et al., 2005; Peelle & Davis, 2012; Poeppel, 2003). This phase alignment causes neural excitability peaks to become aligned in time with the most likely time of occurrence of upcoming stimuli. Synchronization of endogenous neural oscillations to temporal aspects of external stimuli may facilitate the processing of temporally predictable stimuli by increasing the responsiveness and sensitivity of sensory cortex at points in time at which future stimuli are likely to occur (e.g., Large & Jones, 1999; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010). Much ongoing research concerns the types of acoustic cues that may drive entrainment of neural oscillators to the speech signal (e.g., Doelling, Arnal, Ghitza, & Poeppel, 2014) and how such entrainment influences the perceptual parsing and processing of the speech signal at different time scales (e.g., Ghitza & Greenberg, 2009; Giraud et al., 2007; Poeppel, 2003). Distal prosody effects may prove to be a useful domain for future work in this area.

Conclusions

Expectations about the prosodic characteristics of metrically prominent syllables within an utterance constrain the interpretation of suprasegmental cues to lexical stress during spoken-word recognition. Taken together with other effects of distal prosody, these results suggest that spoken-word recognition involves probabilistic expectations about the acoustic–phonetic realization of lexical alternatives on the basis of prior knowledge and preceding context information, which enable spoken-word recognition to proceed efficiently and robustly in real-world (i.e., noisy) conditions. Distal prosody effects therefore provide a useful domain for testing predictions of probabilistic inference approaches. For example, the probabilistic nature of proximal and distal stress cues raises the possibility that listeners flexibly adapt to speaker-specific rhythmicity or use of suprasegmental stress cues, in much the same way as they adapt to novel distributions of phonetic cues (e.g., Bradlow & Bent, 2008; Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2014). Further, expectations based on speech rhythm may play a particularly important role when noisy conditions reduce the availability of proximal stress cues. Whether listeners' use of distal prosody adapts to the relative reliability of various cues to intended meaning is an important domain for future investigation.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264. http://dx.doi.org/10.1016/S0010-0277(99)00059-1

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457–474. http://dx.doi.org/10.1016/j.jml.2007.09.002

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (R Package, Version 0.999375–42) [Computer software]. Retrieved from http://lme4.r-forge.r-project.org

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

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106,* 707–729. http://dx.doi.org/10.1016/j.cognition.2007.04.005

Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience, 29,* 1132–1146. http://dx.doi.org/10.1080/23273798.2014.894642

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review, 18,* 1189–1196. http://dx.doi.org/10.3758/s13423-011-0167-9

Brown, M., Salverda, A. P., Gunlogson, C., & Tanenhaus, M. K. (2015). Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience, 30,* 149–166. http://dx.doi.org/10.1080/01690965.2013.862285

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108,* 804–809. http://dx.doi.org/10.1016/j.cognition.2008.04.004

Clifton, C., Jr., Carlson, K., & Frazier, L. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review, 13,* 854–861. http://dx.doi.org/10.3758/BF03194009

Cole, J., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics, 38,* 167–184. http://dx.doi.org/10.1016/j.wocn.2009.08.004

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45,* 207–228. http://dx.doi.org/10.1177/00238309020450030101

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107. http://dx.doi.org/10.1016/0010-0285(74)90005-X

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction.* Amsterdam: John Benjamins. http://dx.doi.org/10.1075/pbns.25

Cox, D. R. (1970). *The analysis of binary data*. London: Chapman and Hall.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20,* 55–60. http://dx.doi.org/10.3758/BF03198706

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

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236. http://dx.doi.org/10.1016/0749-596X(92)90012-M

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2,* 133–142. http://dx.doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141–201.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121. http://dx.doi.org/10.1037/0096-1523.14.1.113

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

Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990–present*. Retrieved from http://corpus.byu.edu/coca/

Dilley, L., 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. http://dx.doi.org/10.1016/j.jml.2010.06.003

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

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664–1670. http://dx.doi.org/10.1177/0956797610384743

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85,* 761–768. http://dx.doi.org/10.1016/j.neuroimage.2013.06.035

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing [Commentary]. *Behavioral and Brain Sciences, 36,* 211–212. http://dx.doi.org/10.1017/S0140525X12002312

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America, 97,* 1893–1904. http://dx.doi.org/10.1121/1.412063

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science, 37,* 578–591. http://dx.doi.org/10.1111/cogs.12022

Fox, R. A., & Qi, Y. Y. (1990). Context effects in the perception of lexical tone. *Journal of Chinese Linguistics, 18,* 261–264.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica, 66,* 113–126. http://dx.doi.org/10.1159/000208934

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron, 56,* 1127–1134. http://dx.doi.org/10.1016/j.neuron.2007.09.038

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. http://dx.doi.org/10.1037/0033-295X.105.2.251

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31,* 373–405. http://dx.doi.org/10.1016/j.wocn.2003.09.006

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13,* 313–319. http://dx.doi.org/10.1111/1467-9280.00458

Kim, S. (2004). *The role of prosodic phrasing in Korean word segmentation* (Unpublished doctoral dissertation). University of California, Los Angeles.

Kleinschmidt, D., & Jaeger, T. F. (2014). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, (In press).*

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2014). *Probabilistic inferences in pragmatic interpretation of English contrastive prosody*. Manuscript submitted for publication.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104. http://dx.doi.org/10.1121/1.1908694

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology, 94,* 1904–1911. http://dx.doi.org/10.1152/jn.00263.2005

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106,* 119–159. http://dx.doi.org/10.1037/0033-295X.106.1.119

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

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

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62,* 253–265. http://dx.doi.org/10.3758/BF03205547

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

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86. http://dx.doi.org/10.1016/0010-0285(86)90015-0

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118,* 219–246. http://dx.doi.org/10.1037/a0022325

Morrill, T., Dilley, L. C., & McAuley, J. D. (2014). Prosodic patterning in distal speech context: Effects of list intonation and f0 downtrend on perception of proximal prosodic structure. *Journal of Phonetics, 46,* 68–85. http://dx.doi.org/10.1016/j.wocn.2014.06.001

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9,* 453–467. http://dx.doi.org/10.1016/0167-6393(90)90021-Z

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics, 58,* 540–560. http://dx.doi.org/10.3758/BF03213089

Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica, 66,* 95–112. http://dx.doi.org/10.1159/000208933

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234. http://dx.doi.org/10.1016/0010-0277(94)90043-4

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395. http://dx.doi.org/10.1037/0033-295X.115.2.357

Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1299–1313. http://dx.doi.org/10.1037/0278-7393.27.5.1299

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3,* 320. http://dx.doi.org/10.3389/fpsyg.2012.00320

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 11–36). Dordrecht, the Netherlands: Kluwer Academic. http://dx.doi.org/10.1007/978-94-015-9413-4_2

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 564–573. http://dx.doi.org/10.1037/0096-1523.16.3.564

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication, 41,* 245–255. http://dx.doi.org/10.1016/S0167-6393(02)00107-3

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87. http://dx.doi.org/10.1038/4580

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology, 63,* 772–783. http://dx.doi.org/10.1080/17470210903104412

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech, 54,* 147–165. http://dx.doi.org/10.1177/0023830910397489

Rolke, B., & Hofmann, P. (2007). Temporal uncertainty degrades perceptual processing. *Psychonomic Bulletin & Review, 14,* 522–526. http://dx.doi.org/10.3758/BF03194101

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.cognition.2006.10.008

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20,* 172–176. http://dx.doi.org/10.1016/j.conb.2010.02.010

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech, 33,* 47–68.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412–432. http://dx.doi.org/10.1006/jmla.2000.2783

Sumner, M., Gafter, R., Kurumada, C., & Tice, M. (2014). *Integrative effects of speech mode, phonological variants, and frequency in speech perception*. Manuscript submitted for publication.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995, June 16). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634. http://dx.doi.org/10.1126/science.7777863

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011, March 11). How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279–1285. http://dx.doi.org/10.1126/science.1192788

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58,* 251–273. http://dx.doi.org/10.1080/02724980343000927

Wong, P. C., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46,* 413–421. http://dx.doi.org/10.1044/1092-4388(2003/034)

Appendix

Test Items

The temporal-manipulation syllable in each stimulus context is italicized. The words beginning with strong–weak (SW) and weak–strong (WS) stress patterns are in bold.

Many city people *gave* those two . . .	
SW: **alligators** a wide berth.	WS: **alpacas** carrots and hay.
Jenny's helpful neighbor *saw* her two . . .	
SW: **bunnies** escape from their cage.	WS: **bananas** in her driveway.
The teacher's brother *took* that . . .	
SW: **canteloupe** out of the fridge.	WS: **canteen** into the desert.
Mr. Johnson *gave* his . . .	
SW: **cashews** to his students.	WS: **cashier** the wrong credit card.
The seven students rarely *used* their . . .	
SW: **compasses** to find their camp.	WS: **computers** during their lectures.
Betsy never *found* that . . .	
SW: **corset** very comfortable.	WS: **corsage** she was looking for.
Many people *saw* that . . .	
SW: **eagle** around campus.	WS: **eclipse** of the sun.

(*Appendix continues*)

Appendix (*continued*)

Maybe Carlos *got* his . . .	
SW: **gokart** from a local store.	WS: **goatee** to impress a girl.
My doctor often *says* that those . . .	
SW: **harpists** get wrist injuries.	WS: **harpoons** are dangerous.
Heidi sometimes *saw* that . . .	
SW: **jury** leaving the courthouse.	WS: **giraffe** in the city zoo.
The mayor's youngest daughter *said* that her . . .	
SW: **lizard** has escaped before.	WS: **lasagna** has won state prizes.
Jamie's family *bought* those two . . .	
SW: **magnets** for their car.	WS: **magnolia** trees for their yard.
Marcus really *wants* that new . . .	
SW: **mermaid** figurine.	WS: **marina** to be shut down.
Maybe Emma *bought* this . . .	
SW: **mushroom** casserole.	WS: **machine** for her art project.
The tourist never *got* some . . .	
SW: **mustard** for his ham sandwich.	WS: **masseuse** to rub his shoulders.
My sister Marsha really *hates* when her . . .	
SW: **peanut** supply runs out.	WS: **piano** teacher is sick.
The student's final project *used* some old . . .	
SW: **pirate** hats and feathers.	WS: **pipettes** and centrifuges.
The doorman really never *liked* when that . . .	
SW: **pulley** was used unsafely.	WS: **police** man caught him speeding.
Soon Mr. Morris *wants* one . . .	
SW: **racket** for his daughter.	WS: **raccoon** pelt for his cabin.
Cara never *found* that old . . .	
SW: **Reese's** in her purse.	WS: **receipt** in her purse.
On Sunday Sarah's mother *ran* from that . . .	
SW: **tortoise** on the highway.	WS: **tornado** on the highway.
Timmy Mitchell sometimes *sees* that one . . .	
SW: **trafficlight** blinking on and off.	WS: **trapeze** dancer at the circus.
My roommate's father painted *her* with her . . .	
SW: **umpire** and her coach.	WS: **umbrella** and her raincoat.
The council meeting's chairman *asked* for that . . .	
SW: **volume** setting to be changed.	WS: **volcano** to be monitored.

Received October 22, 2013
Revision received September 22, 2014
Accepted October 28, 2014 ■