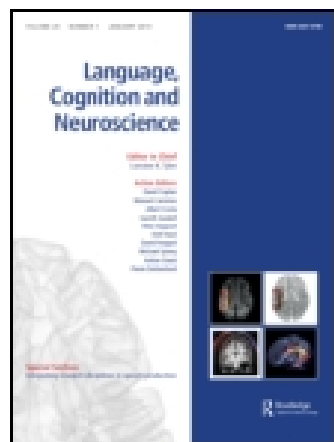


This article was downloaded by: [University of Rochester]

On: 29 January 2015, At: 13:21

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Language, Cognition and Neuroscience

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/plcp21>

### Interpreting prosodic cues in discourse context

Meredith Brown<sup>a</sup>, Anne Pier Salverda<sup>a</sup>, Christine Gunlogson<sup>ab</sup> & Michael K. Tanenhaus<sup>ab</sup>

<sup>a</sup> Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

<sup>b</sup> Department of Linguistics, University of Rochester, Rochester, NY, USA

Published online: 18 Dec 2013.



[Click for updates](#)

To cite this article: Meredith Brown, Anne Pier Salverda, Christine Gunlogson & Michael K. Tanenhaus (2015) Interpreting prosodic cues in discourse context, *Language, Cognition and Neuroscience*, 30:1-2, 149-166, DOI: [10.1080/01690965.2013.862285](https://doi.org/10.1080/01690965.2013.862285)

To link to this article: <http://dx.doi.org/10.1080/01690965.2013.862285>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## Interpreting prosodic cues in discourse context

Meredith Brown<sup>a\*</sup>, Anne Pier Salverda<sup>a</sup>, Christine Gunlogson<sup>a,b</sup> and Michael K. Tanenhaus<sup>a,b</sup>

<sup>a</sup>Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA; <sup>b</sup>Department of Linguistics, University of Rochester, Rochester, NY, USA

(Received 2 April 2012; accepted 21 October 2013)

Two visual-world experiments investigated whether and how quickly discourse-based expectations about the prosodic realisation of spoken words modulate interpretation of acoustic-prosodic cues. Experiment 1 replicated the effects of segmental lengthening on activation of onset-embedded words (e.g. *pumpkin*) using resynthetic manipulation of duration and fundamental frequency (F0). In Experiment 2, the same materials were preceded by instructions establishing information-structural differences between competing lexical alternatives (i.e. repeated vs. newly assigned thematic roles) in critical instructions. Eye movements generated upon hearing the critical target word revealed a significant interaction between information structure and target-word realisation: Segmental lengthening and pitch excursion elicited more fixations to the onset-embedded competitor when the target word remained in the same thematic role, but not when its thematic role changed. These results suggest that information structure modulates the interpretation of acoustic-prosodic cues by influencing expectations about fine-grained acoustic-phonetic properties of the unfolding utterance.

**Keywords:** information structure; prosodic structure; spoken-word recognition; eye movements

Prosodic phrasing organises spoken language into linear phrasal units based on prosodic properties such as pitch, rhythmic structure and syllable duration. The acoustic signatures of prosodic phrasing are therefore useful cues during spoken language comprehension (e.g. Frazier, Carlson, & Clifton, 2006; Salverda, Dahan, & McQueen, 2003). Recent work has shown that the context provided by realisation of prosodic phrasing in early stages of an utterance can shape the interpretation of downstream prosodic cues (e.g. Brown, Salverda, Dilley, & Tanenhaus, 2011; Dilley, Mattys, & Vinke, 2010; Dilley & McAuley, 2008; Frazier et al., 2006). Effects of context may also involve comparatively high-level representations known to influence prosodic realisation, such as the status of a referent with respect to preceding discourse material (its *information status*). When a referent is mentioned multiple times in the same thematic role across utterances (e.g. when it remains the theme), repeated references tend to be acoustically reduced compared to when a new referent is introduced or a previously mentioned referent is assigned a different thematic role (Terken & Hirschberg, 1994). Here, we capitalise on acoustic overlap between cues to information status and cues to prosodic phrasing to examine whether and how quickly the information status of competing lexical alternatives modulates the interpretation of acoustic cues to word boundaries conditioned by prosodic phrasing. We hypothesise that relative activation of competing lexical candidates is influenced by expectations about their acoustic-phonetic realisation that incorporate

information from multiple aspects of linguistic structure, including prosodic phrasing and information structure.

Previous work has demonstrated that listeners use acoustic correlates of word boundaries, such as segmental lengthening, to assist in spoken-word recognition. In a visual-world study, Salverda et al. (2003) used utterances containing words like *hamster*, which contains the onset-embedded word *ham*. Tokens of *hamster* containing a longer *ham-* resulted in more fixations to a visually displayed ham than tokens of *hamster* containing a shorter *ham-*. Salverda et al. concluded that this effect was related to prosodic phrasing: Because syllables are typically lengthened immediately before prosodic boundaries, and because prosodic boundaries are more likely to occur at word boundaries than word-medially, a relatively long *ham-* is more consistent with an immediately following prosodic boundary (i.e. the word *ham*) than a shorter *ham-*. The relative activation of monosyllabic and polysyllabic competitors is also dependent on position within an utterance (Salverda et al., 2007). When *cap* occurs in utterance-medial position (e.g. *Put the **cap** next to the square*), *captain* competes more strongly for recognition than *cat*. However, when *cap* is utterance-final and therefore followed by a stronger prosodic boundary (e.g. *Now click on the **cap***), *cat* is the stronger competitor. Taken together, these results suggest that segmental duration influences perceived prosodic phrasing and that aspects of this phrasing in turn modulate the relative activation of competing lexical alternatives.

\*Corresponding author. Email: [mbrown@bcs.rochester.edu](mailto:mbrown@bcs.rochester.edu)

Segmental duration, together with pitch accenting or lack thereof, also signals information structure: Repeated mentions of an expression in the same thematic role within a discourse are typically deaccented, and therefore relatively short in duration. In contrast, discourse-new expressions and previously mentioned expressions in newly assigned thematic roles are likely to be acoustically more prominent: longer in duration, containing a pitch accent and associated with greater pitch excursion. Dahan, Tanenhaus, and Chambers (2002) showed that listeners interpret the prosodic prominence of a spoken word as a cue to its information status. When participants encountered two successive instructions to move a candle within a visual display (1a), an acoustically prominent second mention of *candle* (which is infelicitous in this context) elicited stronger activation of a discourse-new cohort competitor (*candy*) than a reduced second mention (suggesting given information). Conversely, when the theme of the first instruction was *candy* rather than *candle* (1b), a prominent token of *candle* in the second instruction elicited weaker activation of the previously mentioned word *candy* than a reduced token. In another condition (1c), which is most relevant to the current study, *candle* was mentioned in both instructions but had different thematic roles in each instruction (goal in the first instruction; theme in the second). Contrastive pitch-accenting of the second mention of *candle* resulted in faster fixations to the correct target picture, compared to its repetition in the same thematic role in (1a). These findings suggest that listeners associate prosodic prominence with new information, and, crucially, that both discourse-new referents and previously mentioned referents in newly assigned thematic roles contribute to new information.

- (1a) Put the candle below the triangle ... Now put the candle above the square.
- (1b) Put the candy below the triangle ... Now put the candle above the square.
- (1c) Put the necklace below the candle ... Now put the candle above the square.

Segmental duration is thus related to both prosodic word boundaries and information status. We assume that whereas acoustic cues to prosodic boundaries and information structure may not be identical, they do intersect with respect to lengthening. In particular, listeners might attribute segmental lengthening to the occurrence of a monosyllabic word in phrase-final position or to it coinciding with prominence-lending pitch accenting typical of 'new' information status or both. Similarly, the relative brevity of a syllable may reflect its initial position within a polysyllabic word or may be associated with the 'given' information status of its containing expression or both. It follows that situations may occur in which segmental lengthening signals partially conflicting

information about an upcoming prosodic boundary vs. the information status of a referent. Such situations could shed light on how listeners interpret cues that could originate in diverse ways from diverse sources.

Previous work suggests that listeners simultaneously consider different levels of linguistic representation to constrain the activation of competing lexical alternatives (e.g. Dahan & Tanenhaus, 2004; Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Magnuson, Tanenhaus, & Aslin, 2008; Revill, Tanenhaus, & Aslin, 2008). We hypothesise that listeners likewise consider the acoustic characteristics of a word simultaneously as cues to prosodic phrasing and information structure in order to rapidly establish the word's likely identity. This hypothesis, which we will refer to as the *immediate-interaction hypothesis*, entails two claims about spoken-word recognition: (1) Prosodic cues are evaluated in parallel with respect to multiple levels of linguistic representation (in this case, prosodic phrasing and information structure) and (2) this process has an immediate impact on lexical activation. That is, listeners should simultaneously interpret durational lengthening with respect to prosodic phrasing (i.e. as pre-boundary lengthening) and with respect to information status (i.e. as prominence-induced lengthening). Listeners' initial interpretation of a spoken word is thus predicted to incorporate relevant information about potential lexical alternatives with respect to both prosodic phrasing and information structure. The interpretation of duration as a prosodic phrasing cue should therefore interact with the interpretation of overlapping acoustic cues to information status (pitch-accenting and lengthening).

One possible mechanism for the interaction of these different levels of representation involves predictive processes that forecast aspects of likely acoustic realisation of different lexical candidates in context (Farmer, Brown, & Tanenhaus, 2013). According to this perspective, listeners develop expectations about the acoustic realisations of words by generating internal models, taking into account the words' prior realisation(s) and current context. These expectations incorporate information from different levels of linguistic representation, including information structure and prosodic phrasing. These sources of information predict aspects of the phonetic realisation of upcoming words, which can be used in the earliest moments of lexical processing. This expectation-based account has precedents in theories of other aspects of language processing (e.g. Altmann & Kamide, 1999; Levy, 2008) as well as in research in other cognitive and perceptual domains (Clark, 2013). Interaction in this framework occurs when a cue can be explained by hypotheses based on different subsystems and the interpretation of that cue simultaneously takes into account these hypotheses.

The phenomena we are investigating have not been explicitly addressed by existing models of spoken-word recognition. Therefore, we cannot directly map competing predictions onto specific models. Nonetheless, either or both of the assumptions underlying the immediate-interaction hypothesis might be incorrect, giving rise to at least two classes of alternative possibilities.

First, the interpretation of acoustic detail in spoken words might not depend simultaneously on both prosodic phrasing and information structure (the *non-interaction hypothesis*). Instead, listeners may interpret acoustic cues predominantly with respect to one level of linguistic representation, for instance, prosodic phrasing. This scenario might arise, for example, if listeners prefer to attribute proximal cues like segmental lengthening to proximal causes (i.e. an upcoming prosodic boundary) as opposed to distal causes (e.g. discourse history). In this case, listeners might retrieve information about whether and in what role a referent was previously mentioned only when acoustic variation cannot be attributed to more proximal causes.

Second, the cues associated with different types of representation (e.g. prosodic phrasing cues vs. prosodic cues to information status) might operate at different timescales, or initially be considered by non-interacting subsystems. According to the *delayed-interaction hypothesis*, the interpretation of acoustic cues to prosodic boundaries is initially unaffected by the information status of competing lexical candidates. Conversely, the initial interpretation of prosodic cues to information status might not be modulated by prosodic phrasing. Instead, prosodic cues to different types of structures may be interpreted independently and with distinct time courses. For example, under certain assumptions of traditional feedforward models of language processing, processing proceeds in a hierarchical bottom-up order: The perception and interpretation of speech sounds precede lexical access, which precedes situating a word within a syntactic representation, which in turn precedes retrieving semantic information and constructing information-structural representations (cf. Norris, 1994; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). Extending these assumptions to the construction of prosodic-phrasing representations and information-structural representations would predict that cues to relatively low-level and/or local aspects of prosodic phrasing would have earlier or faster effects compared to cues to information status. Segmental lengthening should thus initially be interpreted as a cue to an upcoming prosodic boundary, favouring a monosyllabic interpretation, whereas inferences relating perceived prominence to previous mention in discourse context should have slower effects on lexical activation, resulting in delayed effects and interactions.

We used the visual-world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995)

to assess effects of information structure on the interpretation of acoustic cues to prosodic phrasing. Experiment 1 evaluated the effect of segmental lengthening and pitch excursion on spoken-word recognition in isolated utterances (e.g. *Now put the **pumpkin** below the triangle*). In Experiment 2, an additional instruction at the beginning of each trial established a discourse context for each of the critical instructions used in Experiment 1, thereby manipulating the information status of the target word. Of interest was whether and how quickly the information status of the target word influenced the interpretation of segmental lengthening and pitch excursion.

## Experiment 1

The goal of Experiment 1 was to evaluate the effects of resynthetic manipulation of duration and F0 on spoken-word recognition.

## Methods

### Participants

Forty native English speakers from the University of Rochester and surrounding community participated in the experiment. All had normal hearing and normal or corrected-to-normal vision.

### Materials and design

The 20 experimental items were instructions to manipulate pictures within a visual display (Table 1). In addition to four shapes in the corners of the display, which remained the same across trials, five pictures were displayed on each trial. These pictures included a target picture associated with a polysyllabic word (e.g. pumpkin), a competitor picture associated with the corresponding onset-embedded word (e.g. pump) and three phonologically unrelated distractors (Figure 1). Distractor pictures were positioned such that all pictures were approximately evenly dispersed throughout the display, with minimal displacement of the centroid of picture locations relative to the centre of the grid.

All instructions were recorded by Christine Gunlogson using a Marantz PMD660 digital recorder sampling at 44.1 kHz. Sentences were produced with consistent speech rate and intonational contours. The pitch synchronous overlap-and-add algorithm (Moulines & Charpentier, 1990) was used in the speech-editing software Praat (Boersma & Weenink, 2010) to create two partially resynthesised versions of each critical instruction (e.g. *Now put the pumpkin below the triangle*). Both duration and F0 of the first syllable of the target word (*pumpkin*) were manipulated. We localised pitch excursion to the first syllable of the target word by flattening the F0 of the syllables immediately preceding and following it, setting their F0 level to match the initial F0 of the preposition

Table 1. Target and competitor pictures displayed in experimental trials.

Target	Competitor
Antlers	Ant
Beaker	Bee
Boulder	Bowl
Candy	Can
Captain	Cap
Carpet	Car
Dolphin	Doll
Hamster	Ham
Leaflet	Leaf
Nectarine	Neck
Panda	Pan
Pirate	Pie
Pumpkin	Pump
Reindeer	Rain
Soldier	Sole
Spider	Spy
Taxi	Tack
Toaster	Toe
Tractor	Track
Welder	Well

following the target word (e.g. *below*). Crucially, the acoustic properties of these proximal context syllables, and of the rest of each critical sentence, were the same across the two resynthesised sentences. Only the acoustic properties of the first syllable of the target word differed across conditions.

In *short-target* items,<sup>1</sup> the first syllable of the target word was shortened to 90% of its original duration, and its F0 was flattened to match that of the preceding and following syllables (Figure 2a). In *long-target* items, the first syllable of the target word was lengthened by 25%, and its F0 was altered to rise and fall approximately 45 Hz, beginning at vowel onset and peaking late in the rhyme (Figure 2b). These F0 manipulations were integral for the design of Experiment 2 because the prominence variations associated with information status affect both F0 and duration (e.g. Ladd, 2008). The goal of Experiment 1 was therefore to demonstrate that manipulating the duration plus F0 of the first syllable of the target word would still result in the effects of duration alone documented in previous studies, where it was shown to affect the interpretation of the syllable as corresponding to a monosyllabic vs. polysyllabic word (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda et al., 2003). As will become clear later, the logic of Experiment 2 required that these stimuli show a stronger monosyllabic competitor effect in the long compared to the short condition. The mean duration of the embedded word was 201 ms for short and 279 ms for long stimuli. The mean duration of

the entire target word was 410 ms for short and 488 ms for long stimuli.<sup>2</sup>

Forty-five filler sentences were constructed to mitigate prosodic and phonological contingencies in the stimuli. In all filler trials, the display contained pictures associated with two phonologically related words and three unrelated distractor words. To discourage expectations that the target word was the longer of two phonologically related words associated with displayed pictures, the target word was the shorter of two phonologically related words for 20 fillers and an unrelated word for the remaining 25. Target words in 10 of these 25 fillers underwent the same acoustic manipulations as target words in critical items, to further discourage expectations that acoustically manipulated target words corresponded to the longer member of a phonologically related pair.

Five stimulus-lists were constructed by pseudo-randomising picture positions within each trial (within the set of positions licensed by the spoken instruction and the even-dispersion constraint) and pseudo-randomising trial order such that experimental trials were non-adjacent. Condition (short-target vs. long-target) was counterbalanced in a Latin square design, resulting in 10 lists completed by an equal number of participants. Each list began with five practice trials to familiarise participants with the task.

### Procedure

Participants' eye movements were recorded using a head-mounted SR Research Eyelink II eye-tracker sampling at 250 Hz. Drift-correction procedures were performed every five trials. Auditory stimuli were presented through headphones.

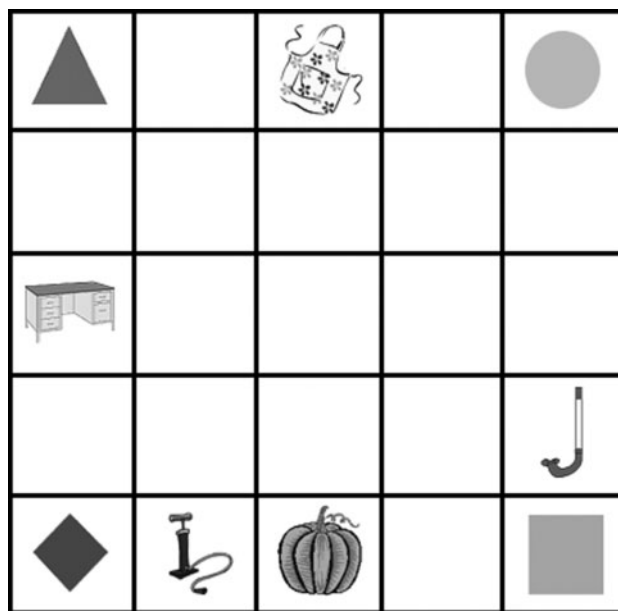


Figure 1. Example of visual display at the start of a trial in Experiment 1.



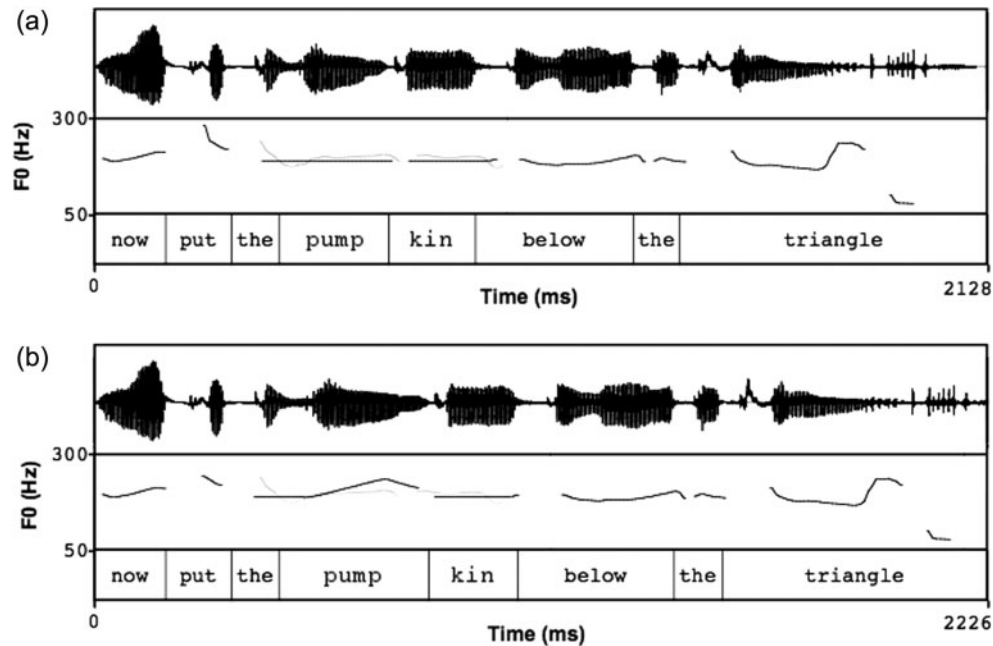


Figure 2. Waveforms and F0 of a resynthesized critical instruction containing (a) short- and (b) long-target words. The F0 of the original instruction is depicted in grey across the acoustically altered syllables.

Each trial began with the visual display appearing in the centre of the screen. After one second, the spoken instruction was presented. One second after the participant successfully executed the instruction, the next trial was initiated. The experiment lasted approximately 10 minutes.

### Results and discussion

Trials on which participants required multiple attempts to select the correct object and drag it to the correct location were excluded from analysis (3.2%). Proportions of errors did not differ significantly across conditions.

Figure 3 shows mean proportions of fixations to target, competitor and distractor pictures over time by experimental condition. At the onset of the target word, proportions of fixations to target and competitor pictures did not differ significantly between conditions. This was confirmed statistically by first transforming the proportions of fixations to each picture at target-word onset using the empirical logit function (Barr, 2008; Cox, 1970) and then conducting one-tailed paired *t* tests on the proportions of competitor fixations (mean untransformed percentage of fixations = 14% for short and 17% for long items;  $t_1(39) = 1.23$ ,  $p > .1$ ;  $t_2(19) = 1.12$ ,  $p > .1$ ) and target fixations (mean = 16% for both short and long items;  $t_1(39) = 0.12$ ,  $p > .1$ ;  $t_2(19) = 0.13$ ,  $p > .1$ ).

Following Salverda et al. (2003), we hypothesised that the proportion of fixations to the picture depicting the embedded word (e.g. *pump*) as participants processed the target word would be higher for long than for short-target words. The data were consistent with this prediction.

Around 200 ms after target-word onset, the proportion of fixations to the target and competitor increased, whereas the proportion of fixations to the distractors decreased. Competitor fixations started to rise at the same rate in both conditions, but ultimately reached a higher peak value and decreased more slowly in the long-target condition than in the short-target condition. Around 1000 ms after target-word onset, competitor fixations merged with distractor fixations. One-tailed paired *t* tests on logit-transformed averaged proportions of competitor fixations between 200 and 1000 ms after the onset of the target word confirmed that there were more fixations to the competitor in the long-target condition than in the short-target condition ( $t_1(39) = 3.14$ ,  $p < .005$ ;  $t_2(19) = 3.02$ ,  $p < .005$ ). Complementary effects were observed in the proportions of target fixations, which exhibited a faster increase in the short-target condition than in the long-target condition, starting at approximately 550 ms after target-word onset ( $t_1(39) = -1.88$ ,  $p < .05$ ;  $t_2(19) = -1.70$ ,  $p = .053$ ).

The results demonstrate that resynthetic manipulation of duration and F0 across the initial syllable of the target word elicits effects similar to those obtained by Salverda et al. (2003) using materials with natural prosodically conditioned acoustic variation. Target words with lengthened duration and F0 excursion were associated with stronger activation of embedded competitors, compared to target words that were compressed and had flat F0. This result sets the stage for Experiment 2 in which listeners heard the same sentences while we manipulated the information status of target and competitor words to adjudicate between the immediate-, delayed- and non-interaction hypotheses. (In

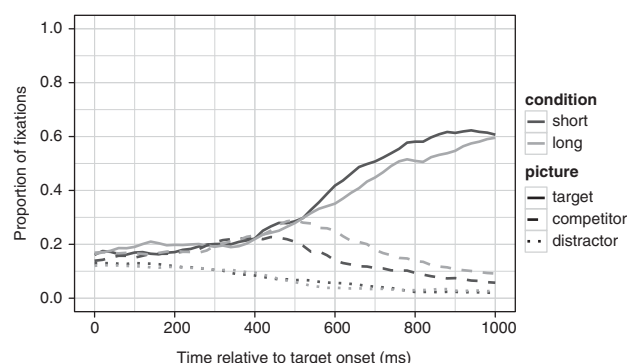


Figure 3. Proportions of fixations to the target, competitor and averaged distractors for short- and long-target items in Experiment 1, relative to the onset of the target word.

the general discussion and [Appendix 1](#), we address concerns raised by one of the editors, Michael Wagner, that the competitor bias observed in response to long stimuli can be explained without appeal to prosodic phrasing because the longer initial syllable of the target word is consistent with both target and competitor interpretations for a longer period of time).

## Experiment 2

The goal of Experiment 2 was to characterise the effects of information structure on the interpretation of acoustic cues to prosodic phrasing. An additional instruction preceding each of the critical sentences from Experiment 1 established a discourse context for each critical instruction. Thematic roles of the target and competitor words were varied within the context instruction, thereby manipulating the information status of competing lexical alternatives associated with the critical second instruction.

## Methods

### Participants

Eighty-four native English speakers from the University of Rochester and surrounding community took part in the experiment. They reported normal hearing and normal or corrected-to-normal vision.

### Materials and design

The critical instructions in Experiment 2 were the same manipulated utterances used in Experiment 1. Context instructions were recorded during the same session as critical instructions. The display characteristics at the start of each critical instruction were identical to those of Experiment 1.

In each trial, participants heard one of two versions of the first instruction, as in (2), after one second of display preview. In same-role (SR) context instructions, the target word (e.g. *pumpkin*) was mentioned in the theme position, with the monosyllabic embedded word (e.g. *pump*) mentioned as part of the composite goal. In

different-role (DR) context instructions, the monosyllabic embedded word was instead mentioned as the theme, and the target word was mentioned as part of the composite goal. Because both words were mentioned in both types of context sentences, the critical information-structural manipulation was whether the thematic role of the target word in the critical instruction remained the same as in the context instruction. In SR-context conditions, the target word had the same thematic role in both sentences, thus corresponding to *given* information in the critical instruction. In DR-context conditions, the critical target word had a newly assigned thematic role in the second instruction, thus contributing to *new* information (compared to SR-context conditions).<sup>3</sup> Further, in both types of context sentences, coordinate structures were used for the goal (e.g. *the pump and the circle*) to avoid differential effects of utterance-final lengthening on the target word and embedded competitor word across the two context-types (cf. Salverda et al., 2007). A continuation rise (i.e. L–H% boundary tone; Pierrehumbert & Hirschberg, 1990) was used at the end of the first instruction to indicate prosodically that the second instruction should be interpreted with respect to the context established by the first instruction.

(2) SR-context instruction:

Put the [pump]<sub>H\*kin</sub> between the [pump]<sub>!H\*</sub> and the [cir]<sub>H\*cle<sub>L</sub>-H%</sub> ...

DR-context instruction:

Put the [pump]<sub>H\*</sub> between the [pump]<sub>!H\*kin</sub> and the [cir]<sub>H\*cle<sub>L</sub>-H%</sub> ...

Short critical instruction:

[Now]<sub>H\*</sub> put the [pump]<sub>!H\*kin</sub> below the [tri]<sub>!H\*angle<sub>L</sub>-L%</sub>.

Long critical instruction:

[Now]<sub>H\*</sub> put the [pump]<sub>L+H\*kin</sub> below the [tri]<sub>!H\*angle<sub>L</sub>-L%</sub>.

The 45 filler items used in Experiment 1 were also paired with context instructions. The resulting set included 8 fillers in which the context instruction contained both members of a phonologically related pair (like the context instructions for the experimental items) followed by an instruction whose theme was either the shorter member of the pair or an unrelated word. In addition, because the theme of each critical instruction was always mentioned in the prior instruction, the second theme of 17 fillers was discourse-new.

Experiment 2 used the same five pseudo-randomised trial lists as Experiment 1. Context-type (SR vs. DR) was crossed with target-word realisation (short vs. long) and counterbalanced in a Latin square design, resulting in 20 lists.

### Procedure

The procedure was similar to that of Experiment 1. Participants completed the initial instruction by moving

the appropriate picture (e.g. pumpkin in SR-context conditions and pump in DR-context conditions) to its appropriate location. The successful completion of this action triggered the appearance of a central dot, which played the second instruction when clicked. This aspect of the procedure brought participants' gaze back to the centre of the display prior to the critical instruction, which minimised baseline differences in fixations to target and competitor pictures at the onset of the critical instruction (e.g. continuing to fixate the most recently manipulated picture or the current location of the cursor). The successful completion of the critical instruction was required to proceed to the next trial. The experiment lasted approximately 15 minutes.

### Predictions

The immediate-interaction hypothesis predicts that listeners simultaneously evaluate the acoustic realisation of a word with respect to multiple dimensions of linguistic structure, including prosodic phrasing and information structure, resulting in an interaction between context-type and target-word realisation in the earliest moments of lexical activation.

In SR-contexts, the polysyllabic target word appears in the role of theme in the initial context-setting instruction, with the monosyllabic competitor as part of the goal phrase (e.g. *Put the pumpkin between the pump and the triangle*). If the theme of the critical instruction (e.g. *Now put the pumpkin below the triangle*) is acoustically reduced, listeners should be biased to interpret it as referring to *given* information (i.e. coreferential with the theme of the context instruction), favouring a polysyllabic interpretation. Conversely, if the target word is acoustically prominent, listeners should be biased to interpret it as referring to comparatively *new* information, favouring the monosyllabic competitor. These expectations are congruent with the durational cues to prosodic phrasing: Longer duration is a cue to an upcoming prosodic boundary, which should favour the competitor over the target. Both prosodic-boundary expectations and information-structural expectations should thus bias listeners towards an initial polysyllabic interpretation of short-target words and an initial monosyllabic interpretation of long-target words, as in Experiment 1.

In DR-contexts, the immediate-interaction hypothesis predicts a different pattern of results. Because DR-contexts introduce the polysyllabic target word as part of the goal and the monosyllabic competitor word as the theme (e.g. *Put the pump between the pumpkin and the triangle*), expectations based on information structure should favour the initial interpretation of an acoustically reduced word in the theme position of the second instruction as a less-prominent instance of the monosyllabic word (consistent with its previous mention in the

same thematic role), and of a lengthened target word as a contrastively accented reference to the polysyllabic target (in its new role as theme). The information-structural expectations conflict with prosodic-phrasing expectations regarding duration: Longer duration is a cue to an upcoming prosodic boundary, as in the SR condition, thus favouring the monosyllabic competitor over the target word. The immediate-interaction hypothesis therefore predicts that the effects of duration on fixations to competitor and target pictures should be reduced in DR-context conditions, compared to SR-context conditions. In fact, proportions of fixations to the competitor picture might initially be higher for short-target words than for long ones.

In contrast, the main prediction of the non-interaction hypothesis is that listeners consider acoustic cues primarily with respect to proximal aspects of linguistic structure (i.e. an immediately upcoming prosodic boundary), such that the effect of target-word realisation on lexical activation is not affected by information structure. That is, listeners may attribute acoustic cues like segmental lengthening primarily to proximal prosodic phrasing, considering information from prior discourse history only when there is no obvious proximal cause for segmental lengthening. According to this hypothesis, effects of lengthening and pitch excursion on early fixations to target and competitor pictures should be similar to those observed in Experiment 1. When the initial syllable of the target word is short, listeners should not expect an upcoming prosodic boundary. When the initial syllable of the target word is long, listeners should expect an upcoming prosodic boundary, favouring the monosyllabic competitor. Thus, there should be more fixations to the monosyllabic competitor (and slower convergence on the target) for long-target words relative to short-target words.

Finally, the delayed-interaction hypothesis predicts that listeners evaluate the acoustic realisation of a word with respect to different aspects of linguistic structure, but with distinct time courses. According to this hypothesis, listeners evaluate the acoustic realisation of a word first with respect to local prosodic phrasing and then with respect to information status. Thus, target words with long initial syllables should initially be associated with a higher proportion of competitor fixations regardless of context-type, followed by delayed effects of information status on fixation patterns (i.e. a delayed attenuation or reversal of duration effects in DR-context conditions, but not SR-context conditions).

### Results and discussion

Four participants were excluded from the analysis because they did not complete the experiment as instructed. Errors occurred in response to either or both instructions on 5.1%



of experimental trials, which were excluded from analysis. Proportions of errors did not differ significantly across conditions.

### Baseline analyses of fixation proportions

Figure 4 shows the proportions of fixations to each picture in SR-context conditions, relative to the onset of the target word. Figure 5 shows the proportions of fixations in DR-context conditions. In both conditions, the proportion of fixations to phonologically unrelated distractor pictures began to drop between 100 and 200 ms following target-word onset. Assuming that it takes approximately 200 ms to programme an eye movement in multi-object displays in response to spoken language (Salverda et al., *in press*), this early effect suggests that participants anticipated repeated mention of pictures referred to in the context sentence. Further analyses of logit-transformed fixation proportions at the onset of the target word revealed that participants were more likely to be looking at the competitor picture in the SR-context conditions (mean untransformed percentage = 10%) than in the DR-context conditions (mean = 6%;  $F_1(1,79) = 8.445$ ,  $p < .005$ ;  $F_2(1,19) = 6.425$ ,  $p < .05$ ), and more likely to be fixating the target picture in the DR-context conditions (mean = 15%) than in the SR-context conditions (mean = 7%;  $F_1(1,79) = 17.174$ ,  $p < .0001$ ;  $F_2(1,19) = 14.466$ ,  $p < .0005$ ). Neither target-word realisation nor the interaction between context-type and target-word realisation significantly influenced competitor or target fixations at word onset.

The baseline differences in fixation proportions between context-types suggest that some aspect of our materials or task structure biased listeners to have a modest expectation for the theme of the second instruction to correspond to the goal of the first instruction rather than the theme of the first instruction. Similar baseline tendencies to fixate pictures corresponding to new information were observed by Dahan et al. (2002). Although listeners generally expect repeated mention across syntactically parallel sentence positions (Chambers & Smyth, 1998), several factors could have reversed this expectation. For example, it is possible that the intonation of the first words of the critical instruction (*Now put the...*) led listeners to expect contrastive accenting on the theme (cf. Braun & Chen, 2010), resulting in an expectation for the upcoming theme to have a new thematic role. However, the intonation of the carrier phrase cannot entirely account for baseline differences in fixations to the theme and goal of the context instruction because they were present as early as 200 ms following onset of the carrier phrase (2.09% of fixations for themes vs. 4.86% for goals;  $t_1(79) = 2.890$ ,  $p < .005$ ;  $t_2(19) = 3.781$ ,  $p < .005$ ). A more likely explanation is that participants encoded more detailed visual and spatial information about the referent of the

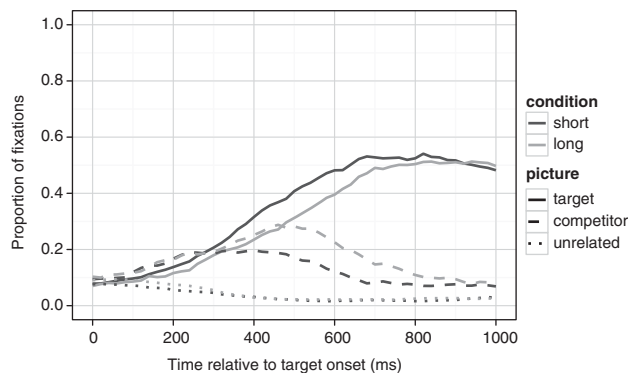


Figure 4. Proportions of fixations to the target, competitor and averaged distractors for short- and long-target items in same-role contexts in Experiment 2, relative to the onset of the target word.

theme of the context instruction while moving it, resulting in a decrease in exploratory fixations to the previously mentioned theme relative to the goal during initial portions of the critical instruction (cf. Salverda, Brown, & Tanenhaus, 2011). Listeners might also have been biased to fixate the most recently mentioned item. Crucially, however, baseline differences in fixation proportions did not prevent examination of effects of target-word realisation on fixations to competitor and target pictures, which are described in the following sections.

### Analysis of competitor fixations between 200 and 1000 ms

For statistical analysis, average proportions of competitor fixations were computed for the interval ranging from 200 to 1000 ms after target-word onset (as in Experiment 1), and transformed using the empirical logit function. Of primary interest was the interaction between context-type and target-word realisation. The immediate-interaction hypothesis predicts that effects of acoustic manipulation on interpretation of the target word should be attenuated in DR-contexts, compared to SR-contexts. In contrast, the non-interaction hypothesis predicts that these two factors

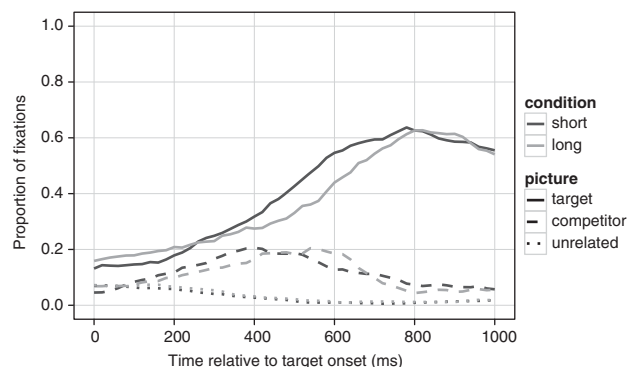


Figure 5. Proportions of fixations to the target, competitor and averaged distractors for short- and long-target items in different-role contexts in Experiment 2, relative to the onset of the target word.

should not interact in the earliest moments of spoken-word recognition, whereas the delayed-interaction hypothesis predicts slower effects of information structure relative to prosodic boundaries. A comparison of competitor fixations in Figures 4 and 5 supports the predictions of the immediate-interaction hypothesis. In SR-contexts, long-target words elicited a higher proportion of competitor fixations than short ones. In DR-contexts, however, proportions of competitor fixations followed a similar trajectory in short-target and long-target conditions.

A two-way analysis of variance (ANOVA) on logit-transformed proportions of competitor fixations confirmed a significant interaction between context-type and target-word realisation ( $F_1(1,79) = 6.610, p < .05$ ;  $F_2(1,19) = 4.051, p < .05$ ). This interaction indicates that the effects of acoustic manipulation on competitor activation differed as a function of context. In SR-contexts, a planned two-tailed paired  $t$  test revealed that competitor fixations were significantly higher for long-target items (mean = 18%) than for short-target items (mean = 13%;  $t_1(79) = 3.291, p < .005$ ;  $t_2(19) = 2.952, p < .01$ ). This result is consistent with all three hypotheses. In DR-contexts, however, competitor fixations were not significantly higher for long-target items (mean = 12%) than for short-target items (mean = 13%;  $t_1(79) = -0.686, p > .1$ ;  $t_2(19) = -0.947, p > .1$ ). In fact, competitor fixations were numerically slightly lower for long-target items than for short-target items.

#### *Analysis of target fixations between 200 and 1000 ms*

Although our hypotheses primarily focused on activation of the competitor word as a function of context-type and target-word realisation, we expected the proportions of target fixations to show a complementary pattern to the proportions of competitor fixations (though see Footnote 4). We expected processing of short-target words to be facilitated in SR-contexts, but not in DR-contexts. However, participants fixated the target less upon hearing long-target items, regardless of context-type. Mean untransformed percentages of fixations in SR-context conditions were 41% for short-target items and 37% for long-target items, whereas mean percentages in DR-context conditions were 47% for short-target items and 43% for long-target items. A two-way ANOVA on logit-transformed proportions of target fixations revealed significant main effects of context-type ( $F_1(1,79) = 8.000, p < .01$ ;  $F_2(1,19) = 11.765, p < .005$ ) and target-word realisation ( $F_1(1,79) = 4.394, p < .05$ ;  $F_2(1,19) = 4.203, p < .05$ ). However, the interaction between context-type and target-word realisation was not significant ( $F_1(1,79) = 0.038, p > .1$ ;  $F_2(1,19) = 0.167, p > .1$ ).

#### *Adaptation analyses*

One explanation for this unexpected result is suggested by an expectation-based account of the immediate-interaction

hypothesis: Listeners may adapt to the prosodic characteristics of speakers, including their speech rate, pitch range and consistency with which they signal information structure intonationally (see Kurumada, Brown, & Tanenhaus, 2012, for supporting evidence). Because Experiment 2 used a counterbalanced crossed-factors design, the prosodic characteristics of half of the critical items were infelicitous or unexpected (e.g. SR-long items, in which a target word with a repeated thematic role was acoustically prominent). Therefore, interactions between context-type and target-word realisation may have attenuated over the course of the experiment, due to increasing exposure to inconsistent prosodic usage.

This possibility was examined by analysing logit-transformed competitor and target fixation proportions, using multi-level linear regression models, with random intercepts and slopes for participants and items, using the *lmer* function within the *lme4* package (Bates, Maechler, & Bolker, 2011) in *R* (version 2.15.0, R Development Core Team, 2012). Fixed effects and random slopes in the regression models included context-type, target-word realisation, trial number (the position of the item in the sequence of trials encountered by the participant) and interactions between these factors. Context-type and target-word realisation were sum-coded. Trial number was standardised by subtracting the mean value and dividing by the standard deviation. Final models were chosen by removing the fixed effects stepwise and comparing each reduced model to the more complex model, using a likelihood ratio test (Baayen, Davidson, & Bates, 2008). Models included maximal random effect structure except as noted (Barr, Levy, Scheepers, & Tily, 2013). We used the *kappa* function in *R* (R Development Core Team, 2012) to verify that the condition number was below 7 for all models. Condition numbers below 7 indicate a low degree of collinearity between predictors, which would not be expected to jeopardise the reliability of model estimates.

The final model predicting logit-transformed proportions of competitor fixations revealed a significant interaction between context-type and target-word realisation (Table 2), confirming the results obtained by the two-way ANOVA. In addition, a main effect of trial number indicated an overall decrease in proportions of competitor fixations as the experiment progressed. The three-way interaction between context-type, target-word realisation and trial number was not included in the final model because it did not contribute significantly to explained variance.

For targets, the final model confirmed that the interaction between context-type and target-word realisation was not significant (Table 3). However, this model revealed a significant three-way interaction between context-type, target-word realisation and trial number, indicating that the interaction between context-type and

Table 2. Parameters and formula of final linear regression model of logit-transformed competitor fixations averaged between 200 and 1000 ms after target-word onset.

	$\beta$	SE	$t$	$p$
Intercept	-0.781	0.023	-33.53	< .0001
Context = DR	-0.062	0.040	-1.53	n.s.
Target = long	0.043	0.022	1.90	< .10
Trialnumber	-0.035	0.011	-3.11	< .005
Context:target	-0.129	0.044	-2.93	< .005

Formula:  $\text{lmer}(\text{CFIX} \sim \text{context} + \text{target} + \text{context}:\text{target} + \text{trialnumber} + (1 + \text{context} + \text{target} + \text{context}:\text{target} | \text{subj})) + (1 + \text{context} + \text{target} + \text{context}:\text{target} | \text{item}), \text{data} = \text{data})$

target-word realisation changed as the experiment progressed. The effects of trial number on the interaction between context-type and target-word realisation are illustrated in Figure 6, which shows fixation proportions averaged across the first and last third of experimental trials in SR-context and DR-context conditions, respectively. Early in the experiment, mean proportions of target fixations in SR-contexts were higher for short-target words than for long ones (and, importantly, this effect originated early in the processing of the target word), whereas differences in the proportions of target fixations by condition were not apparent in DR-contexts. Later in the experiment, however, mean proportions of target fixations were numerically higher in the DR-short condition than in the DR-long condition.

These patterns of target fixations were further evaluated statistically by comparing the regression lines for the two-way interaction between context-type and target-word realisation when the trial number was set to different values (cf. Aiken & West, 1991). For this analysis, we fitted two additional models to the data, with the trial number centred as one standard deviation either below or above the mean, and examined the interaction between context-type and target-word realisation within each model. The critical interaction was significant at one standard deviation below the trial-number mean (i.e. early in the experiment; Table 4), but not at one standard deviation above the mean (i.e. late in the experiment; Table 5). The significant interaction in the early trial-number model was driven by a significant difference in the interpretation of short vs. long items between SR and DR-contexts. Although fixations to the target picture rose significantly faster in the SR-short condition than in the SR-long condition early in the experiment ( $\beta = -0.152$ ,  $SE = 0.066$ ,  $t = -2.282$ ,  $p < .05$ ), target-word realisation did not elicit significant differences in proportions of target fixations in DR-contexts early in the experiment ( $\beta = 0.041$ ,  $SE = 0.063$ ,  $t = 0.658$ ,  $p > .1$ ). That is, the proportions of target fixations showed patterns that mirrored those observed for competitor fixations early in the experiment, and the interaction of context-type and target-word realisation attenuated significantly across the

experiment for target fixations but not for competitor fixations. Note, however, that the numerical size of the interaction for competitor fixations was smaller in the second half of the experiment,<sup>5</sup> suggesting similar adaptation effects for fixations to competitors.

To summarise, the effects of segmental lengthening and F0 excursion on fixations to potential referents during the initial moments of processing the target word were modulated by context-type. In SR-contexts, monosyllabic embedded words competed for recognition more strongly when the initial syllables of target words were lengthened and associated with pitch excursion. However, this pattern was attenuated due to perceived prominence in DR-contexts, consistent with predictions of the immediate-interaction hypothesis. Similar patterns of results were observed for target fixations, but only early in the experiment.

## General discussion

Our results suggest that the information status of competing referents affects interpretation of prosodically conditioned detail in the speech signal. Segmental lengthening that is interpreted as signalling a following prosodic boundary when no discourse context is available (Experiment 1) can be interpreted as a consequence of pitch-accenting, signalling the information status of a referent when an appropriate discourse context is available (Experiment 2). Further, listeners appear to evaluate the acoustic realisation of a target word with respect to both prosodic phrasing and information structure upon processing its initial sounds (Experiment 2).

These results contribute to a growing body of work demonstrating that the effects of prosodically conditioned detail on lexical activation depend on the context in which a word is processed (e.g. Brown et al., 2011; Dilley & McAuley, 2008). For example, manipulations of prosodic information distal to the locus of processing (e.g. F0 and duration of preceding prosodic constituents or metrically prominent vs. non-prominent syllables) influence the extent to which phonemically overlapping words compete for recognition during spoken-word recognition (Brown,

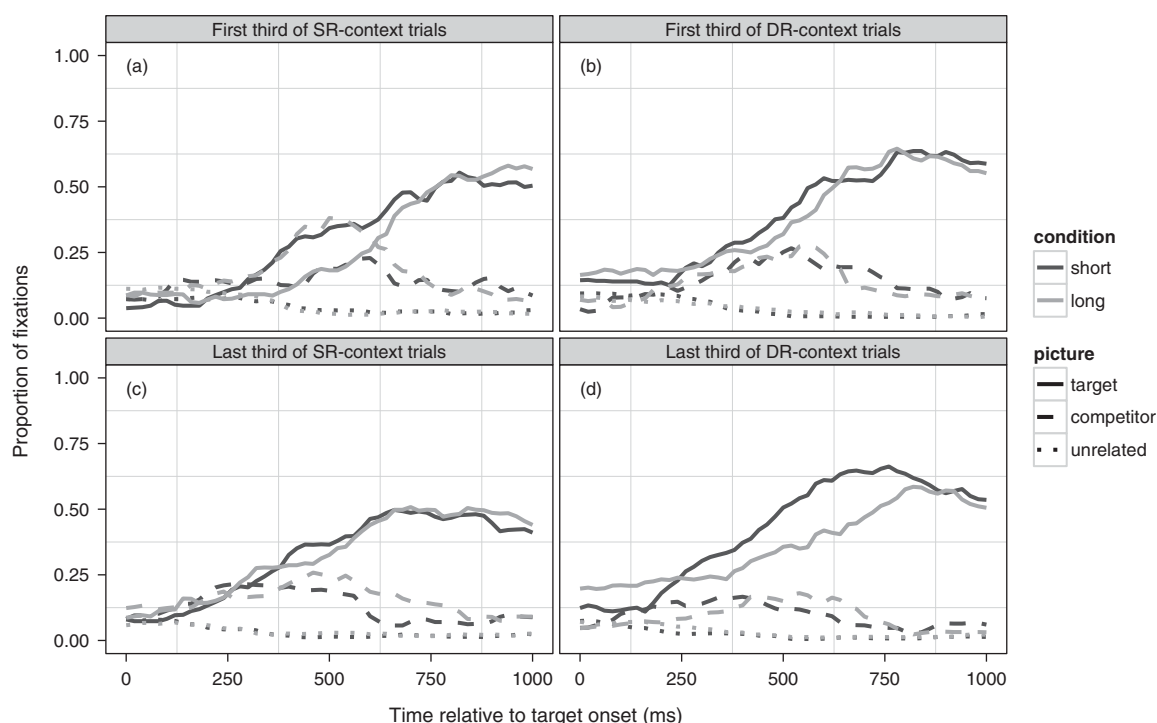


Figure 6. Proportions of fixations to target, competitor and distractor pictures for short- and long-target items in Experiment 2 for (a) the first third of the trials in same-role contexts, (b) the first third of the trials in different-role contexts, (c) the last third of the trials in same-role contexts and (d) the last third of the trials in different-role contexts.

Dilley, & Tanenhaus, 2012; Brown et al., 2011; Brown, Salverda, Dilley, & Tanenhaus, 2012). They are also consistent with work in speech perception, showing that listeners evaluate phonetic cues with respect to speech context (e.g. Cole, Linebaugh, Munson, & McMurray, 2010; Ladefoged & Broadbent, 1957; McMurray & Jongman, 2011). Taken together with the results of the present study, these context effects suggest that proximal prosodically conditioned detail is interpreted dynamically with respect to a variety of broader contextual factors.

One potential concern about interpreting the present data involves the relative timing with which different aspects of linguistic structure are available to the language processing system. For example, if information status is available at an earlier point in time than prosodic-boundary information, then delayed effects of information status in conjunction with immediate (but later-occurring) effects of prosodic-boundary information could result in what might appear to be an immediate interaction. Indeed, the presence of baseline differences between fixations and referents previously mentioned in a prominent (theme) vs. non-prominent (goal) position in Experiment 2 suggests the possibility that the discourse history influenced fixations prior to the onset of the target word (although, as previously discussed, baseline differences may also have resulted from task-based factors). Regardless, the information structure of previous utterances does not unambiguously specify the information structure of

upcoming material. Listeners may have expectations about the relative likelihood of different referents being mentioned as the theme of the critical instruction – just as they may have expectations about the prosodic constituency of upcoming material (Brown et al., 2011) – but they cannot be certain of the identity of the theme during the processing of the first syllable of the target word. The crucial factor in Experiment 2 is the acoustic manipulation of the initial syllable of the target word. This manipulation results in patterns of fixations that are distinct from prior baseline effects, suggesting that listeners interpret the acoustic properties of this syllable as evidence for one referent or the other with respect to prosodic phrasing and information structure. The onset of the target word is the earliest point at which acoustic cues to prosodic boundaries and information status are available to the listener, and the effects that we observed originated shortly after target-word onset.

As mentioned previously, another potential concern involves whether the patterns of competitor fixations observed in Experiment 1 reflect the interpretation of segmental lengthening as a cue to prosodic phrasing, or simply a delayed point-of-disambiguation effect. We address this concern in more detail in Appendix 1. However, even if one assumes that the competitor bias observed in Experiment 1 does not reflect the interpretation of lengthening as signalling an upcoming prosodic boundary, the general interpretation and theoretical



Table 3. Parameters and formula of final linear regression model of logit-transformed target fixations averaged between 200-1000 ms after target-word onset.

	$\beta$	SE	$t$	$p$
Intercept	-0.173	0.040	-4.30	< .0001
Context=DR	0.135	0.051	2.64	< .01
Target=long	-0.081	0.033	-2.46	< .05
Trialnumber	0.019	0.020	0.93	n.s.
Context:target	0.027	0.068	0.40	n.s.
Context:trialnumber	-0.023	0.042	-0.55	n.s.
target:trialnumber	-0.037	0.032	-1.15	n.s.
Context:target:trialnumber	-0.169	0.062	-2.72	< .01

Formula:  $\text{lmer}(\text{TFIX} \sim \text{context} + \text{target} + \text{context}:\text{target} + \text{trialnumber} + \text{context}:\text{trialnumber} + \text{target}:\text{trialnumber} + \text{target}:\text{trialnumber}:\text{context} + (1 + \text{context} + \text{target} + \text{trialnumber} | \text{subj}) + (1 + \text{context} + \text{target} + \text{trialnumber} | \text{item}), \text{data} = \text{data})$

conclusions from the data in Experiment 2 would be essentially the same. The data would still demonstrate that a higher-level factor (information structure established by preceding discourse context) has an immediate constraining effect on the interpretation of different acoustic realisations of the target word.

The finding that various and disparate aspects of preceding context influence the initial interpretation of segmentally identical material suggests that these effects arise from perceptual expectations. That is, listeners may evaluate incoming prosodic information with respect to expectations about aspects of the acoustic-prosodic realisation of spoken words (based on, e.g. prior linguistic knowledge, prior experience and the acoustic realisation of words in preceding context). For example, listeners may expect a word to have a relatively short duration and small F0 excursion when it is mentioned twice in the theme position of sequential utterances (as in SR-context conditions in Experiment 2). The acoustic realisation of the initial syllable of the target word in the SR-long condition is incongruent with this expectation. Likewise, following a DR-context sentence, listeners should expect a second mention of the target word in a newly assigned thematic role to be associated with lengthening and F0 excursion, making the acoustic realisation of the target word in the DR-short condition infelicitous within this context.

As in other cognitive domains, expectations about certain perceptual attributes of upcoming speech (e.g. prosodic characteristics) may provide a compelling explanation for the rapidity with which listeners are capable of evaluating acoustic information with respect to both prosodic phrasing and discourse-based information structure. According to this perspective, listeners develop provisional hypotheses (internal models) about the expected realisation of different lexical alternatives (e.g. recently mentioned words). These hypotheses compete to explain the incoming linguistic input. The strength of support for competing lexical alternatives depends on the degree of congruence between the signal and the expected

realisation of each competing alternative in context. The primary determinant of whether different aspects of language processing interact, then, is the degree to which they contribute to alternative hypotheses about the realisation of the same acoustic-phonetic cues, rather than their relative positions within a hierarchy of representations in the language-processing architecture. Because prosody is influenced by a wide range of linguistic factors, listeners' prosodic expectations may incorporate various sources of information from the surrounding context, including preceding prosodic patterning, information structure and prosodic phrasing. The interaction between context-type and target-word realisation observed in Experiment 2 may have reflected expectations incorporating projected effects of both information status and upcoming prosodic boundaries on the acoustic realisation of competing lexical alternatives.

The present results are incompatible with models of spoken language processing that assume that cues to relatively low-level or local aspects of linguistic structure are interpreted prior to cues to higher-level or more global structure. More specifically, such models would predict that the spoken-word recognition system does not rapidly take into account how the interpretation of a prosodic phrasing cue that is relevant to word recognition would be modulated by information structure. Instead, the findings of the present study (and other results demonstrating effects of different types of contextual constraint on spoken-word recognition) suggest that spoken-word recognition involves rapid integration of different types of linguistic information.

More generally, existing models of spoken-word recognition cannot straightforwardly account for our results. Simple effects of acoustic cues such as lengthening and F0 excursion on recognition of words in null or neutral contexts can be explained within the existing framework of several models of spoken-word recognition, such as localist connectionist models, such as TRACE and Shortlist (McClelland & Elman, 1986; Norris, 1994), by incorporating the effects of such cues into the dynamics of



Table 4. Parameters and formula of linear regression model of averaged logit-transformed target fixations with trial number centred one standard deviation below the mean.

	$\beta$	SE	$t$	$p$
Intercept	-0.192	0.042	-4.58	< .0001
Context = DR	0.162	0.054	2.98	< .005
Target = long	-0.044	0.045	-0.99	n.s.
Trialnumber	0.020	0.020	1.00	n.s.
Context:target	0.197	0.091	2.16	< .05
Context:trialnumber	-0.028	0.041	-0.68	n.s.
Target:trialnumber	-0.038	0.032	1.18	n.s.
Context:target:trialnumber	-0.175	0.065	-2.69	< .01

Formula: `lmer(TFIX~context*target*lowtrialnumber+(1+context+target+lowtrialnumber+context:target+context:lowtrialnumber+target:lowtrialnumber+context:target:lowtrialnumber|subj)+(1+context+target+lowtrialnumber|item),data=data)`

lexical activation or into the lexical selection criteria (see Salverda et al., 2003, for more discussion). However, distal-context effects observed in this and previous studies suggest that the interpretation of segmental duration as a proximal cue to spoken-word recognition cannot be attributed solely to dynamics of lexical competition elicited by the currently processed word. Rather, segmental duration is interpreted dynamically with respect to a variety of broader contextual factors.

#### *Effects of prosodic adaptation*

The expectation-based perspective is further supported by the adaptation effects suggested by the significant three-way interaction between trial number, context-type and target-word realisation in the regression model predicting target fixation proportions. Numerous studies of phonetic processing have demonstrated that listeners' phonemic categories are continuously updated and adjusted in accordance with distributional properties of recent input (e.g. Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, under review; Kraljic & Samuel, 2007, 2011; Kraljic, Samuel, & Brennan, 2008; Norris, McQueen, & Cutler, 2003). The question of whether and how listeners adapt to prosodic properties of speech has

received comparatively little attention. Nonetheless, observations that speakers vary considerably in their realisation of intonational categories and in their use of intonational cues to signal linguistic or paralinguistic contrasts suggest a role for prosodic adaptation in language comprehension. Indeed, pragmatic interpretation of sentence-level prosodic contours can be modulated by exposure to a relatively small set of utterances sampled from novel distributions of prosodic contours (Kurumada et al., 2012).

Within-experiment adaptation is a methodological issue that has not received much attention within psycholinguistic domains in which effect sizes tend to be large and stable, such as syntactic processing (though see Farmer, Monaghan, Misyak, & Christiansen, 2011; Fine & Jaeger, 2011). For studies of prosody, however, adaptation may present methodological challenges, particularly for standard experimental designs involving counterbalanced crossed factors (see also Bibyk, Heeren, Gunlogson & Tanenhaus, 2011; Carbary et al., under review). When participants' expectations are incongruent with the prosody of some stimuli, they may quickly learn that prosody is an unreliable cue within the experimental stimuli, given their expectations. As a result, participants may subsequently discount prosodic factors or interpret them differently when evaluating the acoustic realisation

Table 5. Parameters and formula of linear regression model of averaged logit-transformed target fixations with trial number centred one standard deviation above the mean.<sup>6</sup>

	$\beta$	SE	$t$	$p$
Intercept	-0.170	0.040	-4.27	< .0001
Context = DR	0.112	0.053	2.12	< .05
Target =long	-0.112	0.046	-2.44	< .5
Trial number	0.006	0.019	0.299	n.s.
Context:target	-0.167	0.095	-1.75	< .10
Context:trial number	-0.025	0.038	-0.66	n.s.
Target:trial number	-0.028	0.032	-0.86	n.s.
Context:target:trial number	-0.186	0.067	-2.76	< .01

Formula: `lmer(TFIX~context*target*hightrialnumber+(1+context+target+hightrialnumber+context:target+context:hightrialnumber+target:hightrialnumber+context:target:hightrialnumber|subj),data=data)`

of incoming words or phrases (Kurumada et al., 2012). Such learning effects are predicted by the expectation-based account: The perceived reliability of the association between information status and acoustic prominence, for example, should be related to the degree to which information structure contributes to listeners' expectations about the acoustic realisation of competing lexical alternatives. Generally speaking, when perceptual expectations are incongruent with the actual realisation of a word, listeners should update their beliefs about the cues that condition their perceptual expectations, resulting in adaptation that more closely aligns listeners' expectations with the characteristics of the signal in context.

Our results suggest that participants adapted to the reliability of perceived prominence as a referential cue over the course of Experiment 2 in which acoustic characteristics of our stimuli sometimes violated information-structural expectations. This apparently encouraged listeners to discount perceived prominence as a cue to information status over the course of the experiment. Acoustic cues to prosodic boundaries were also occasionally at odds with prosodic expectations in our materials, but the overall reliability of prosodic phrasing cues likely remained relatively high due to the comparatively large number of naturally produced prosodic boundaries across the experimental materials. Prosodic adaptation would therefore be predicted to affect information-structural effects to a larger degree than prosodic phrasing effects in Experiment 2.

The effects of the information structure manipulation in Experiment 2 are reflected more strongly in the patterns of looks to competitors than to targets. This in itself is unsurprising: Although it is rarely commented upon in the literature, most studies using cohort competitors find stronger effects of experimental manipulations on looks to cohorts compared to looks to targets (e.g. McMurray, Tanenhaus, & Aslin, 2002, 2009). Although a detailed explanation of the comparative sensitivity of competitor fixations is beyond the scope of this paper, we present a possible explanation in Footnote 4.

## Conclusions

Our results provide evidence that listeners evaluate the acoustic realisation of spoken words with respect to both prosodic phrasing and discourse-derived information structure, and that both sources of information have immediate effects on spoken-word recognition. This finding adds to a growing body of work on spoken-word recognition, suggesting that information from multiple levels of linguistic representation simultaneously constrains the activation of competing lexical alternatives. The current study extends this line of inquiry to the interpretation of prosodic information at different levels of linguistic representation. Most importantly, these results suggest a crucial role for expectations about aspects of the

acoustic-phonetic properties of unfolding utterances in language comprehension.

## Acknowledgements

We thank Dana Subik for assistance with participant recruitment and testing; Sarah Bibyk, T. Florian Jaeger, and the audiences at Experimental and Theoretical Advances in Prosody II and CUNY 2011 for constructive discussions; and Delphine Dahan, Joe Toscano and the editors, Duane Watson and Michael Wagner, for cogent reviews which resulted in a much-improved paper.

## Funding

This research was supported by a National Science Foundation graduate research fellowship to MB and National Institutes of Health grants [HD27206] and [DC0005071/HD073890] to MKT.

## Notes

1. We use the terms *short-target* and *long-target* for ease of exposition. Each condition also involved F0 manipulations.
2. We elicited spoken instructions from five native speakers in which the theme was either the polysyllabic target word or the monosyllabic competitor, intermixed with 96 syntactically varied sentences. The mean difference in duration between the monosyllabic competitor word and the initial syllable of the polysyllabic target word in these recordings was 78 ms (319 vs. 241 ms) – the same magnitude as the durational difference in our resynthetically manipulated stimuli.
3. We obtained recordings of the stimulus pairs and additional utterance pairs in which the critical themes were monosyllabic embedded words (e.g. *pump*), from the five speakers who provided supplementary recordings for Experiment 1. Following a context sentence in which a polysyllabic target word was mentioned as the theme and the embedded competitor word was part of the composite goal (i.e. a SR-context sentence), the mean duration of the first syllable of the polysyllabic target word in the critical instruction was 217 ms, whereas the duration of the monosyllabic word was 311 ms, a difference of 94 ms. When the roles of the polysyllabic and monosyllabic words were switched in the context sentence (i.e. a DR-context sentence), difference in durations between the syllables in the critical instruction was substantially reduced (248 ms for polysyllabic vs. 291 ms for monosyllabic words, a difference of 43 ms).
4. Effects in visual-world experiments examining spoken-word recognition are often statistically more robust for competitors than for targets. A possible explanation comes from in-progress work on an event-based analysis framework that predicts both (1) the probability that a saccade will occur to a target, competitor and unrelated distractor and (2) the duration of the resulting fixation (Frank, Salverda, Pontillo, Jaeger, & Tanenhaus, in preparation). In examining a data-set (Salverda & Tanenhaus, 2010) we noticed the following pattern. Differences in fixation proportions to a strong and weak competitor were primarily due to fixations on the stronger competitor having longer durations than fixations on the weaker competitor. Thus the size of the competitor effect is due to participants being less likely to leave the stronger competitor within a specified time window. These 'competitor-to-target' shifts are the fixations that are most likely to be sensitive to the relevant manipulations because participants nearly always shift from

the competitor to the target by the end of the analysis window (which is typically chosen to coincide with the point in time at which competitor fixations merge with distractor fixations). The same pattern is observed in studies manipulating voice onset time (McMurray, p.c.). Dahan and Gaskell (2007) make a similar argument for frequency effects. Competitor-to-target shifts also contribute to different patterns of looks to targets across conditions. However, target fixations within that same window also include fixations that are likely to be much less sensitive to experimental manipulations. These include initial fixations to the target, where participants never receive inconsistent segmental information. They also include distractor-to-target shifts, which may be less sensitive to experimental manipulations because distractor fixations are less likely to have been launched as a result of information from the target word.

If we assume that competitor-to-target fixations provide most of the signals with respect to the manipulation, competitor fixations will contain a higher signal-to-noise ratio than target fixations. If we further assume that noise-fixations are distributed similarly across conditions, we would expect the following patterns across the literature: (1) In most studies, competitor and target fixations will show the same patterns, with statistical analyses showing bigger effect sizes for competitor analyses compared to target analyses; (2) on some proportion of studies, the effects will be reliable in competitor analyses but not reliable (or even trending in the opposite direction) in target analyses. The reverse is possible, but much less likely; and (3) the likelihood of seeing reliable effects for both targets and competitors will be greater with manipulations with bigger effect sizes. This explanation is of course speculative and will require explicit tests with new classes of analyses and richer data-sets. However, it provides a possible explanation for otherwise puzzling patterns in our data and in the larger set of visual-world studies of spoken-word recognition.

5. Although the three-way interaction was not significant for competitor fixations, visual inspection of competitor effects suggested a trend in the same direction: The interaction between context-type and target-word realisation was numerically larger early in the experiment (Figure 6). The interaction between context-type and target-word realisation was significant in a multilevel regression model with trial number centred as one standard deviation below the mean (i.e. early trials;  $\beta = -0.156$ ,  $SE = 0.065$ ,  $t = -2.397$ ,  $p < .05$ ) but was not significant in a model with trial number centred as one standard deviation above the mean (i.e. late trials;  $\beta = -0.105$ ,  $SE = 0.067$ ,  $t = -1.556$ ,  $p > .1$ ). Thus, the attenuation of target effects was qualitatively mirrored in competitor effects.
6. Because the model failed to converge with random effects by items in addition to by participants, we ran an additional model containing only random effects by items (intercepts and slopes for context-type, target-word realisation, trial number, and the interaction between context-type and target-word realisation). Effects were qualitatively similar between the two models, though the interaction between context-type and target-word realisation failed to reach significance, marginal or otherwise ( $\beta = -0.114$ ,  $t = -1.163$ ).

## References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264. doi:10.1016/S0010-0277(99)00059-1
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. doi: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. doi:10.1016/j.jml.2007.09.002
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using Eigen and R interfaces (R package, version 0.999375-42) [Computer software]. <http://CRAN.R-project.org/package=lme4>
- Bibyk, S., Heeren, W., Gunlogson, C., & Tanenhaus, M. (2011). *Gotta conundrum, gotta solution*. Poster presented at Experimental and Theoretical Advances in Prosody II, Montréal, QC.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.25) [Computer program]. Retrieved January 20, 2010, from <http://www.praat.org>
- Braun, B., & Chen, A. (2010). Intonation of ‘now’ in resolving scope ambiguity in English and Dutch. *Journal of Phonetics*, 38, 431–444. doi:10.1016/j.wocn.2010.04.002
- Brown, M., Dilley, L. C., & Tanenhaus, M. K. (2012). Real-time expectations based on context speech rate can cause words to appear or disappear. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 1374–1379). Austin, TX: Cognitive Science Society.
- 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. doi:10.3758/s13423-011-0167-9
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2012). Metrical expectations from preceding prosody influence spoken-word recognition. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 1380–1385). Austin, TX: Cognitive Science Society.
- Carbary, K., Brown, M., Gunlogson, C., McDonough, J. M., Fazlipour, A., & Tanenhaus, M. K. (under review). Anticipatory deaccenting in online language comprehension: A phonemic restoration study.
- Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language*, 39, 593–608. doi:10.1006/jmla.1998.2575
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204. doi:10.1017/S0140525X12000477
- 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. doi:10.1016/j.cognition.2008.04.004
- 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. doi:10.1016/j.wocn.2009.08.004



- Cooper, R. M. (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(1), 84–107. doi:10.1016/0010-0285(74)90005-X
- Cox, D. R. (1970). *The analysis of binary data*. London: Chapman and Hall.
- Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57, 483–501. doi:10.1016/j.jml.2007.01.001
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 498–513. doi:10.1037/0278-7393.30.2.498
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*, 47, 292–314. doi:10.1016/S0749-596X(02)00001-3
- Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218–244. doi:10.1037/0096-1523.28.1.218
- 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. doi:10.1016/j.jml.2010.06.003
- Dilley, L., & McAuley, J. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59, 291–311. doi:10.1016/j.jml.2008.06.006
- Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences*, 36, 211–212. doi:10.1017/S0140525X12002312
- Farmer, T. A., Monaghan, P., Misyak, J. B., & Christiansen, M. H. (2011). Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1318–1325.
- Fine, A. B., & Jaeger, T. F. (2011). Language comprehension is sensitive to changes in the reliability of lexical cues. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 925–930). Austin, TX: Cognitive Science Society.
- Frank, A., Salverda, A. P., Pontillo, D., Jaeger, T. F., & Tanenhaus, M. K. (in preparation). An event-based framework for analyzing visual world data.
- Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10, 244–249. doi:10.1016/j.tics.2006.04.002
- Kleinschmidt, D., & Jaeger, T. F. (under review). Recognize the familiar, generalize to the similar, and adapt to the novel.
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15. doi:10.1016/j.jml.2006.07.010
- Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, 121, 459–465. doi:10.1016/j.cognition.2011.08.015
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19, 332–338. doi:10.1111/j.1467-9280.2008.02090.x
- Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, 119(1), 23–42. doi:10.1016/j.cognition.2010.12.002
- Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Prosody and pragmatic inference: It looks like adaptation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 647–652). Austin, TX: Cognitive Science Society.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98–104. doi:10.1121/1.1908694
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177. doi:10.1016/j.cognition.2007.05.006
- Magnuson, J. S., Tanenhaus, M. K., and Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken-word recognition. *Cognition*, 108, 866–873. doi:10.1016/j.cognition.2008.06.005
- McClelland, J. L., & J. L. Elman. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- 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.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42. doi:10.1016/S0010-0277(02)00157-9
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from ‘lexical’ garden paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1), 65–91. doi:10.1016/j.jml.2008.07.002
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9, 453–467.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 89–234.
- Norris, D. G., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238. doi:10.1016/S0010-0285(03)00006-9
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Revill, K. P., Tanenhaus, M. K., & Aslin, R. N. (2008). Context and spoken word recognition in a novel lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1207–1223. doi:10.1037/a0012796
- Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye-movements in visual-world studies. *Acta Psychologica*, 137, 172–180. doi:10.1016/j.actpsy.2010.09.010
- Salverda, A. P., Dahan, D., & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90(1), 51–89. doi:10.1016/S0010-0277(03)00139-2
- Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of

- prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, 105, 466–476. doi:10.1016/j.cognition.2006.10.008
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (in press). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*. doi:10.1016/j.jml.2013.11.002
- Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1108–1117.
- Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659. doi:10.1016/S0022-5371(79)90355-4
- Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18, 427–440. doi:10.1016/S0022-5371(79)90237-8
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. doi:10.1126/science.7777863
- Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing ‘given’ information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37, 125–145.

### Appendix 1. Analysis of distractor-to-target and distractor-to-competitor saccades as a function of target-word realisation.

Here, in response to concerns raised by Michael Wagner, we consider an alternative explanation for the effect of lengthening on competitor fixations that we observed in Experiment 1 that does not appeal to the effects of prosodic phrasing on lexical activation. Instead, this account considers how lengthening of the initial syllable of the target word affects the temporal distribution of segmental information, and how this may affect the lexical interpretation of the speech signal. We will refer to this account as the point-of-disambiguation account.

When the first syllable of the target word is lengthened, there is a longer period of time during which the speech signal is phonemically consistent with both target and competitor. Listeners therefore have more time to process the initial syllable of a lengthened target word, and more time to launch signal-driven saccades to both target and competitor pictures (resulting in increasing proportions of looks to both pictures) as the proportion of looks to the distractors decreases. This could result in a higher overall proportion of competitor (and target) fixations, with a later peak, when the initial syllable is lengthened. Prior to the availability of phonetic information from the second syllable, the proportion of fixations to the target and competitor picture should increase at approximately the same rate. Once listeners begin processing the second syllable of the target word, they start to favour the target over the competitor. This results in more fixations to the target. Importantly, the increase in target fixations (relative to competitor fixations) due to processing the second syllable of the target word would start earlier when the initial syllable of the target word is short compared to when that syllable is long. Therefore, in Experiment 1, when proportions of fixations to target and competitor are analysed across a relatively broad temporal window (i.e. a

window extending beyond the period of time during which listeners are presented with the initial syllable of the target word), the point-of-disambiguation account predicts (1) a higher overall proportion of competitor fixations and delayed peak in the long condition, and (2) an earlier rise in and/or higher overall proportion of fixations to the target in the short condition – from the point in time where the second syllable of the target word is processed. The statistical tests that we report in the main body of the text focus on a broad time window, and therefore do not necessarily distinguish between the predictions of the point-of-disambiguation account and the prosodic-boundary account.

However, the two accounts differ in the predictions that they make regarding listeners’ eye movements prior to the point of disambiguation. According to the point-of-disambiguation account, as listeners process the initial syllable of the target word, they should launch signal-driven saccades from distractor pictures to competitor and target pictures (which are both phonemically consistent with the input during this window), resulting in decreases in proportions of distractor fixations and increases in proportions of both competitor and target fixations. Therefore, the point-of-disambiguation account predicts that lengthening will increase fixations to competitor and target pictures to a similar extent. In contrast, the prosodic-boundary account predicts that lengthening will favour a short-word interpretation of the syllable (e.g. *pump*) over a long-word interpretation (e.g. *pumpkin*) because listeners anticipate an upcoming prosodic boundary on the basis of cues to duration contained within the initial syllable (Salverda et al., 2003), and such a boundary is consistent with a short-word interpretation but not with a long-word interpretation. Thus, only the prosodic-boundary account predicts that lengthening will result in more fixations to the competitor relative to the target during the processing of the initial syllable of the target word.

To distinguish between predictions of the point-of-disambiguation and prosodic-boundary accounts, we examined saccades from signal-irrelevant screen positions (i.e. any location on the screen other than the target or competitor picture) to the target or the competitor during the processing of the initial syllable of the target word (see Table A1). These data were assessed for each individual stimulus, taking into account a 200 ms delay for programming and executing an eye movement. Because the number of such events was relatively low, we combined the data from both experiments, collapsing across discourse context in Experiment 2. There were substantially more trials with a saccade from a distractor to a competitor when the initial syllable of the target word was long than when it was short. However, the opposite pattern, although numerically small, was found for saccades from a distractor to a target. A chi-square test showed that lengthening of the initial syllable of the target word had a different effect on the likelihood of competitor vs. target saccades ( $\chi^2 = 6.6$ ,  $p < .05$ ).

To address the possibility that effects of lengthening on target vs. competitor fixations were merely skewed in some way by the discourse manipulation in Experiment 2, we verified that differences in distractor-to-competitor vs. distractor-to-target saccades as a function of target-word realisation were observed when the analysis included only events from SR-contexts ( $\chi^2 = 6.0$ ,  $p < .05$ ) and only events from DR-contexts ( $\chi^2 = 3.2$ ,  $p < .10$ ). In addition, data from Experiment 1 alone show the same numerical pattern as the combined data-set. Statistical analysis of these data failed to reach significance, but this was likely due to insufficient power.

Taken together, these exploratory analyses suggest that the increase in competitor fixations observed in response to lengthened stimuli in the current study is not simply due to an effect of increased processing time on lexical activation, but instead results at least in part from the interpretation of segmental lengthening as a cue to prosodic phrasing.



Table A1.

	Saccade endpoint	
	Target	Competitor
Experiment 1 + Experiment 2		
Short	164	112
Long	161	170
Experiment 1 + Experiment 2 (SR-context)		
Short	107	67
Long	108	114
Experiment 1 + Experiment 2 (DR-context)		
Short	96	80
Long	88	109
Experiment 1 alone		
Short	41	38
Long	41	58