Running Head: Acoustic prominence in distracted speech

Listeners perceive acoustic prominence differently for distracted and fluent speakers.

Jennifer E. Arnold, Giulia C. Pancani, and Elise Rosa

University of North Carolina, Chapel Hill, NC

Address correspondence to:

Jennifer Arnold

CB #3270 Davie Hall

Rm. 337B

Chapel Hill, NC 27599-3270

919-843-5737

fax: 919-962-2537

jarnold@email.unc.edu

ABSTRACT

Acoustically prominent and accented words can indicate information status, such that accented words tend to be used when the referent is discourse new, while reduced words are reserved for given information. But prosody can also reflect production difficulty, e.g. hesitant speech may reflect distraction. We examine how speech fluency modulates the information-status function of prosody. In two visual world eyetracking experiments, listeners responded to instructions like *Put the bagel on the circle. Now put the {bacon/BACON/ bagel/BAGEL} on the square*. The target was the object in the second sentence, which was either given or new, and either acoustically prominent or reduced. All subjects heard two blocks: a) fluent , and b) distracted, in which the speaker was supposedly distracted by a secondary task and spoke with longer word durations and hesitant pauses. Prominence was achieved with higher pitch in both blocks. In Exp. 1, the instructions were recorded naturalistically, and duration correlated with acoustic prominence only in the fluent block. In Exp. 2, the fluent targets were manipulated in *Praat* to produce longer distracted-sounding targets that were otherwise acoustically matched. The results from both experiments showed that distracted speech led to a strong discourse-new bias, and only a small relative effect of acoustic prominence. These findings demonstrate the importance of duration to the perception of acoustic prominence, and show that distraction can disrupt the usual information-status function of prosody.

Key Words: Distraction; Disfluency; Acoustic Prominence; Reference Comprehension

In spoken language, words vary in prosodic prominence. Some words are pronounced with emphasis, especially when they carry a pitch accent, while others are more reduced, especially when they are unaccented. This prosodic marking is one important signal to listeners about the speaker's intended message. For example, one function it plays is to help listeners identify the informational status of the speaker's words (Halliday, 1967; Hirschberg, 1995; Ladd, 1996). Pausing and accenting can also mark syntactic structure (inter alia, Ferreira, 2007; Kraljic & Brennan, 2005; Shafer, et al. 2000; Snedeker et al., 2003). Yet prosodic variation can also reflect things that are unrelated to the speaker's intended meaning. For example hesitant speech might indicate that the speaker is distracted. In this paper we examine how these combined sources of variation affect language comprehension. We specifically ask how distraction affects the comprehension of prosody.

We focus on the function of prosody to signal information status. Speakers tend to use accented and acoustically prominent pronunciations for information that is focused or new to the context, while they use reduced pronunciations for information that is given, topical, accessible, or predictable (e.g., Brown, 1983; Fowler & Housum, 1987; Halliday, 1967; see Baumann & Grice, 2006, for a discussion of finer-grained distinctions). The clearest example of this comes from words that are repeated in adjacent utterances, such as "*I'm eating pasta for lunch. Do you want pasta too?*" The second "pasta" is likely to be shorter, lower pitched, and less intelligible than the first one. By contrast, the first "pasta" is likely to carry a pitch accent, and be acoustically prominent and longer. This prosodic contrast has been explained in terms of a variety of informational properties. The second *pasta* is informationally "given", or "old", while the first one is "new" (Chafe, 1976; Prince, 1981, 1992; Schwartzchild, 1992), meaning that it has been evoked in the context already. It is also more predictable (Arnold, 1998; Jurafsky et al., 2001; Bell et al., 2009; Fowler & Housum, 1987), and not the focus of the second utterance (Rooth, 1992). For the purposes of this paper, we will use the cover term "given" for contexts like this in which a referential expression is repeated in adjacent utterances in a discourse.

There is strong evidence that listeners are sensitive to the information-status function of prosody, and they use it rapidly to constrain their interpretation of spoken words. Dahan, Magnuson, & Tanenhaus (2001) used a visual-world eyetracking paradigm to demonstrate that listeners expect a word to refer to given information if it is acoustically reduced, while they expect a new referent for words that are acoustically prominent. Using the same paradigm, Arnold (2008) replicated this effect with both adults and 4-5 year old children. Participants in these studies viewed a display like that in Figure 1, and followed instructions to move objects around, for example *Put the bacon on the circle. Now put the bacon on the square*. The critical word was the object in the second sentence, which was pronounced with either accented (BACON) or unaccented (bacon) prosody. The first sentence was also manipulated: in some trials the target object was also mentioned in the first sentence, establishing it as given. In the other trials, the target was not mentioned, making it new. These new trials instead mentioned an object with a similar-sounding name (i.e., a cohort competitor), e.g. the bagel. In both Dahan et al. (2001) and Arnold (2008), the unaccented trials led participants to look more quickly to the given cohort object, while this preference disappeared in the accented trials. Dahan et al. even found that accented target words induced a preference for the discourse-new object. The advantage of this paradigm is that it capitalizes on the incremental nature of language. As listeners hear the word *ba….,* the first syllable is consistent with both the target and the competitor. This allows researchers to identify listener's biases based on the discourse context. Similar comprehension patterns come from other experimental paradigms (e.g., Terken &

Nooteboom, 1994).

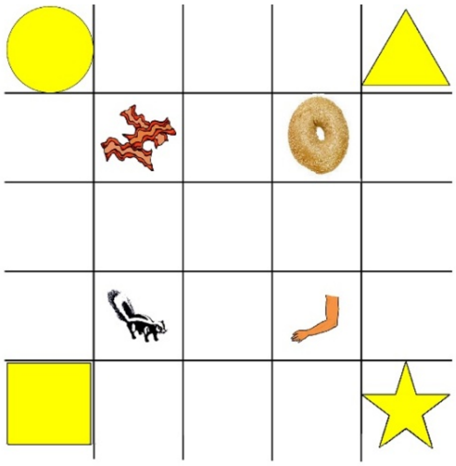

Figure 1. Sample visual display for both Experiments 1 and 2 (replicated from Arnold, 2008). The critical cohorts in this trial are the bacon and bagel.

Thus, experimental evidence suggests that the information-status function of prosody has robust effects on comprehension. Yet this evidence mostly comes from experimental tasks that use fluent speech in relatively clean situations, with constant speech rate. This contrasts with natural speech, in which fluency varies greatly. Even in a relatively fluent utterance, the duration of individual words is influenced by production processing considerations, such as the frequency of the word being produced (Bell et al., 2009). This raises questions about how listeners interpret prosodic variation. In the next section, we examine how prosody can also reflect other sources of information, and consider the implications for comprehension.

<u>Prosodic variation is ambiguous</u>

In English, the information-status function of prosody is primarily described in terms of the location of a pitch accent (Baumann & Hadelich, 2006; Ladd, 1996, Pierrehumbert &

Hirschberg, 1990). While pitch accents can vary in their contour (e.g., high, low, rising, falling, etc.), a more general distinction is between words that are accented and those that are not. In English, accents that convey discourse-newness are typically transcribed as H* or L+H* in the ToBI coding system (Beckman & Elam, 1997). These accents tend to be expressed in terms of longer duration, higher pitch, or in the case of L+H*, greater pitch movement (Ladd, 1996).

Yet the prosodic modulations that are associated with information status are not limited to accenting distinctions. Even among accented words, there is significant variation in word duration that correlates with things like given status (Bard & Aylett, 1999) and predictability (Watson, Arnold, & Tanenhaus, 2010). In a corpus analysis, Bell et al. (2009) found that word duration was longer on first mention than on subsequent mentions, even when controlling for intonational accent, as well as other probabilistic predictors like frequency. Breen et al. (2010) provide a detailed analysis of the acoustic parameters associated with focused and given information statuses, and demonstrate that speakers reliably use distinctions in duration, pitch, and intensity to mark both the location and breadth of the focused element in an utterance. Although these distinctions could be mediated by accent choice, they are also consistent with the view that duration, pitch, and intensity are independently modulated by information status. Consistent with this, Lam & Watson (2010) provide evidence that intensity and duration can reflect different aspects of the discourse context. In their production study, repeated words were shorter and carried less intensity, but predictability only affected intensity, and not duration. This suggests that listeners may possibly use duration, pitch, and intensity as separate signals.

A complicating factor is that duration, pitch, and intensity can also vary for other reasons. One of these reasons is speaker fluency. When speakers have production difficulty, it often results in disfluency, such as pauses, fillers such as *um,* and *uh*, repeated words, and other hesitations (Goldman-Eisler, 1968). These disfluent elements are associated with longer word durations on surrounding words (Bell et al., 2003). Even when speech isn't difficult enough to result in disfluent words per se, the ease of retrieving words affects word duration (Bell et al., 2009; Gahl, Yao, & Johnson, 2012). Conceptual difficulty also leads to longer word durations (Christodoulou, 2009), as does planning difficulty (Gillespie, 2011).

The complications arising from fluency are compounded by the fact that production difficulty and information status are correlated (Arnold & Watson, under review). Speaking about new information is simply harder than speaking about information that has previously been retrieved at conceptual and/or linguistic levels of representation. Indeed, speakers are more likely to be disfluent when talking about new than given information (Arnold & Tanenhaus, 2011). Recent production studies have argued that information status effects on duration may be at least partly the result of production facilitation or difficulty (Kahn & Arnold, 2012; under review; Lam & Watson, 2010).

The relation between speech difficulty and prosody raises questions about how listeners use prosody to direct language comprehension. If a listener hears a longish, prominent-sounding word, do they infer that the word was difficult for the speaker to produce? Or do they interpret the prosodic cues as a signal for new information? And if listeners infer speaker difficulty, does this eliminate the inference of new information status?

The idea that listeners recognize the speaker's difficulty is supported by evidence that listeners use disfluency to guide speech comprehension. Disfluent noun phrases (e.g., *thee, uh, red…*) lead listeners to anticipate reference to an unfamiliar object (Arnold, Husdon Kam, & Tanenhaus, 2007; Barr, 2001), an unpredictable word (Corley et al. 2007), or something that has not been mentioned recently, i.e. something that is discourse-new (Arnold, Tanenhaus, Altmann, & Fagnano, 2004).

Even in the absence of other indicators of disfluency, longer pronunciations and hesitations may still signal that the speaker is having difficulty. In some cases, the source of difficulty may have nothing to do with the content of the speaker's message. A clear case of this is when the speaker is distracted by a secondary task.

Distraction affects communication

Our study examines the effects of distracted speech on the interpretation of prosody. In doing so, it also sheds light on questions about how communication is affected by contexts in which the speaker is distracted. Popular opinion suggests that people are increasingly involved in "multi-tasking", which inevitably leads to situations where their attention is divided among tasks. Distraction from speaking can have significant effects on driving (Briem & Hedman, 1995, Becic et al., 2010; Kubose et al., 2006), and distraction from multiple sources can affect pedestrian safety (Schwebel et al., 2012).

There is good reason to expect that distraction might affect speech. Intuitively, it might lead to more hesitation and disfluency, especially if the speaker tries to perform two tasks simultaneously. Becic et al. (2010) found that people who were distracted by driving were less accurate at both telling stories and remembering stories they heard. Rosa & Arnold (2011) found that in a story-telling task, speakers used more explicit referential expressions when they were distracted than when not. When speaking to distracted addressees, speakers in another study also produced longer and more explicit referential expressions (Rosa, Finch, Bergeson, & Arnold, 2013).

Yet no studies that we know of have examined how distracted-sounding speech affects the interpretation of prosody. The main question that we examine here is whether listeners mentally “excuse” a speaker’s disfluency when the context makes it clear that the disfluency stems from distraction. That is, the listener might interpret the speech in accordance with their knowledge that the speaker was distracted. If so, long pronunciations might not sound prominent within the context of distracted speech, and listeners may not exhibit a bias to consider discourse-new referents following prominent expressions. On the other hand, if listeners are swayed by the absolute prominence of a long-sounding word, they might erroneously take a distracted-sounding token to be evidence of discourse-newness.

To examine this question, we used Dahan et al.’s (2001) visual-world eyetracking experimental approach, as described above. In one block of trials, listeners were told that they were hearing speech that was produced by someone who was distracted with a secondary task; in another block they were told that they were hearing an undistracted speaker. The instructions in the distracted block sounded hesitant and disfluent, with longer pronunciations overall. The question was whether the interpretation of prosody would be contingent on the fluency of the context.

A handful of recent studies have identified ways in which the perception of prosody is contingent on contextual factors that modulate duration or pausing. For example, Clifton, Carlson, and Frazier (2006) presented listeners with sentences that included prosodic breaks in different places. When the pause occurred before or after a short constituent, listeners appeared to take the pause as a marker of the syntactic structure more frequently than when the pause flanked a long constituent. They attributed this to the fact that speakers often pause before or after long constituents, due to the processing demands posed by linguistically complex constituents (Watson & Gibson, 2004). What this means is that listeners essentially excuse the speaker for having produced a longer pronunciation on the basis of the processing involved in producing longer constituents, and do not take it as a signal that the pause marked a syntactic boundary.

A similar sensitivity to prosodic context comes from a study by Brown, Salverda, Gunlogson, and Tanenhaus (under review). Listeners heard words like *pan*, and *panda*, which contained the same phonological content during the first syllable. Importantly, the pronunciation of *pan* is longer when it stands alone than when it is embedded in *panda*. In Brown et al.'s first experiment, they showed that listeners used the length of the first syllable in *panda* as a cue to the word (*pan* vs. *panda*). But within a discourse context, listeners took a long *paaanda* as evidence that the referent was discourse-new, the preference to look at the competitor (e.g., the pan) disappeared. What this means is that when listeners heard a longish token within a discourse context, they attributed the token's prominence to the referent's discourse status, and not to the presence of a prosodic break after the syllable, as in the word *pan*.

Here we ask whether listeners display sensitivity to the contingent nature of prosody in a distracted-speech situation. If all the words in an utterance are hesitant and lengthened, a relatively long word might still sound reduced, and if so, may lead listeners to preferentially consider given referents. On the other hand, distraction might function differently from the studies mentioned above. Both constituent length (Clifton et al., 2006) and discourse status (Brown et al., under review) are salient aspects of the linguistic context. As such, they may be highly available for listeners to use during contingent calculations of prosodic signals. By contrast, distraction is a more complex contextual constraint. Identifying distraction requires the listener to calculate the speaker's mental state, a type of theory-of-mind processing, which may be difficult. This raises the possibility that listeners may not adjust their assessment of prosodic prominence to the distracted situation. If the association between prosody and discourse-status is primary, they may hear distracted, prominent-sounding expressions and instead expect a discourse-new referent.

This latter possibility is supported by the results from a study by Isaacs & Watson (2010). They used the Dahan et al. paradigm to examine the independent effects of pitch and duration on comprehension, using resynthesized tokens that had short or long duration, and high or low f0. They found a three-way interaction between duration, f0, and information status. When duration was short, listeners displayed the expected information-status biases associated with pitch: a discourse-new bias for high-f0, and a discourse-given bias for low-f0 tokens. By contrast, when duration was long, there was little effect of pitch overall. This suggests that lengthening a token may disrupt its ability to signal information status. On the other hand, their long and short target words occurred in identical, fluent contexts, raising questions about what happens in a distracted context.

Experiment 1

Subjects

36 native speakers of English participated in exchange for course credit. Four were excluded (1 for excessive track loss; 1 because one block of their data was lost; two were excluded to create even numbers of participants on each lists). This left 32 participants in the analysis, 4 on each list.

Design and Materials

Task overview. The task was identical to that for experiment 1 in Arnold (2008). Participants were asked to follow prerecorded instructions on a computer. On all trials, participants saw a grid with four yellow shapes in the corners. On each trial, a different set of four objects appeared in the positions shown in Figure 1.

Participants followed two movement instructions for each trial, for example *Put the bagel on the triangle. Now put the bacon on the circle*. As soon as the participant had completed the first movement, the second instruction played automatically. Each trial was initiated by clicking on a dot in the middle of the screen (which also served to automatically correct any drift in the calibration of the eyetracker). Each participant received two blocks of 32 trials (16 stimulus, 16 filler), one with fluent speech and one with distracted speech. Half of the participants began the experiment with the fluent speech block and the other began with the distracted speech block.

Equipment. We monitored participants' eye movements with a head-mounted Eyelink II eyetracker. Eye position was sampled at a rate of one datapoint every 4 msec (250 Hz), but the data were converted to a granularity of one data point for every 20 msec prior to analysis to speed processing (McMurray, 2002). Corneal reflection monitoring was used when possible; the pupil-only monitoring mode was used if we could not achieve adequate calibration with the corneal reflection mode. We analyzed only one eye, using the Eyelink automatic procedure for choosing the eye with better calibration.

The visual and auditory stimuli were presented on a PC computer running the ExBuilder software (an in-house software created at the University of Rochester; Longhurst, 2006), running on a PC computer with a 19” monitor (resolution: 1280 x 1024 pixels, refresh rate: 75 Hz).

Stimulus design. The visual display for each trial consisted of an array of four objects and four shapes on a screen (Fig. 1). For stimulus trials, two of the objects were cohort competitors (e.g., bacon/bagel) and the remaining two were unrelated distractor objects (e.g., skunk/arm).

Our primary manipulation was between fluent and distracted-sounding speech. Participants were told that a student had recorded the sound files in two blocks: one in which she was attentive, and simply following graphical cues that told her where each object should move, and one in which she was distracted because she was pressing a button every time she heard a beep in her headphones while recording the instructions. In fact, the third author (ER) recorded all of the instructions and tried to convey attentiveness and distraction in her speech. When recording the instructions, ER spoke normally in the fluent condition, and in the distracted condition she hesitated and elongated words, trying to convey distraction.

Each trial consisted of two movements. The first instruction mentioned one of the cohort objects. We defined the mentioned cohort as the given cohort, and the other cohort as the new cohort. The second instruction was the target instruction. The target object was always the same for all conditions of a trial (e.g., *bacon*), and the first sentence mentioned either the target object or the cohort competitor, such that the target word referred to either the given or new cohort.

We also manipulated the pronunciation of the target word in both fluent and distracted conditions, creating both acoustically prominent and acoustically reduced conditions. In the fluent condition, prominence was similar to the accented condition in Arnold (2008), while acoustic reduction was similar to the unaccented condition. The variation between prominent and reduced tokens was performed naturalistically, in which ER naturally varied the duration, pitch, and intonational phrasing of the utterances. An example instruction is shown in Table 1, and the average acoustic duration and pitch of the target words are shown in Table 2. Soundfiles for this example are available at http://arnoldlab.web.unc.edu/supporting-materials-for-arnold-pancani-rosa-2013/.

Table 1. Example instruction in each condition.

	Condition	First Sentence	Second Sentence (uppercase indicates prominence)
DISTRACTED	Prominent-New	Put.. thee.. bagel.. on the triangle,	now.. put .. the.. BACON.. on the star.
	Prominent-Given	Put.. thee.. bacon.. on the circle,	now.. put .. the.. BACON ..on the star.
	Reduced-New	Put.. thee.. bagel.. on the triangle,	now.. put .. the.. bacon .. on the star.
	Reduced-Given	Put.. thee.. bacon.. on the circle,	now.. put .. the.. bacon .. on the star.
FLUENT	Prominent-New	Put the bagel on the triangle,	now put the BACON on the star.
	Prominent-Given	Put the bacon on the circle,	now put the BACON on the star.
	Reduced-New	Put the bagel on the triangle,	now put the bacon on the star.
	Reduced-Given	Put the bacon on the circle,	now put the bacon on the star.

An examination of the means in Table 2 shows several things. First, the distracted stimuli were much longer than the fluent stimuli. This is consistent with the tendency for disfluent speech to be slower than fluent speech (Bell et al., 2003). In keeping with this, the distracted condition was slower for the entire utterance. The first three words “Now put the” comprise the “Introduction” to the utterance; the duration of the introduction was longer in the distracted than the fluent condition overall, and was even slightly longer in the reduced than the prominent conditions for both sets of stimuli. Thus, even though the distracted targets were longer than the fluent targets in an absolute sense, they actually represented a smaller proportion of the first four words of the utterance. Within both fluent and disfluent items, the prominent targets represented a greater proportion of the utterance than the reduced targets did.

Table 2. Prosodic properties of auditory stimuli for Exp. 1.

Condition	Avg. target duration	Avg. Syll. 1 Resid. Dur. *	Intro. Avg. dur.	Target proportion of first 4 words	Avg. pitch	Avg. Pitch mvt.	Max pitch	Min pitch
Fluent / prominent	480	-43	444	52%	250	115	307	192
Fluent / reduced	428	-81	480	47%	206	53	231	178
Distracted / prominent	661	57	1210	36%	239	104	291	187
Distracted / reduced	697	66	1549	31%	215	52	239	187

* The residual of the first syllable duration was calculated from a mixed effects linear regression model that included item identity as a random effect, and fixed predictors Number of syllables and Number of phonemes.

Second, for the fluent block, prominent tokens were somewhat longer than reduced tokens. However, in the distracted block, the reduction manipulation did not lead to shorter pronunciations for the reduced tokens, and in fact the reduced tokens were even a little longer than the prominent ones[1]. This pattern is apparent in both the average raw duration measure, and the average residual duration measure. Nevertheless, in both fluent and distracted blocks, prominence was marked by the proportion of the target duration, relative to other words in the utterance.

Third, the prominent and reduced conditions differed on pitch measures, for both fluent and distracted blocks. The prominent tokens had a higher average pitch, and higher maximum pitch than the reduced tokens, for both distracted and fluent stimuli. In summary, while

[1] The prominent target was longer than the reduced target in 24 out of the 32 fluent items, but only 11 out of the 32 distracted items.

prominence in the fluent items was marked by both pitch and duration, prominence in the distracted items was only marked by pitch.

Each of the 32 experimental items represented a set of four visual objects and a fixed target word. Each item appeared in the eight conditions resulting from the 2x2x2 cross between Distraction, Prominence, and Target object. There were 16 fluent fillers, and 16 distracted fillers. Each of the fillers also included two cohort objects in the visual display, but the instructions never referred to them. This was meant to reduce the expectation that the cohorts would be the objects that moved. All the fillers mentioned different objects in each instruction, and thus followed a "discourse-new" pattern.

Object and shape position were counterbalanced across both experimental and filler items. On critical items, the first instruction always moved the object to the shape that appeared in the same corner. This maintained the relative position of the target and competitor objects as closely as possible across given and new conditions.

List design. There were a total of 32 experimental trials and 32 filler trials. The experimental trials were divided into 8 lists, in a Latin square design. Thus, each list contained an equal number of trials in each condition, and each item occurred equally across lists in each condition. These were pseudorandomly intermixed with the fillers for each block.

Half of the participants heard the distracted block first and the other half heard the fluent block. Between blocks the experimenter would remind the participant about which kind of instructions they were about to hear.

Analytical method.

Dependent variable. Eye movement data throughout the paper are presented in terms of **looks**, where a look is defined as a fixation grouped together with the prior saccade. Saccades were identified using Eyelink's on-line parser, which uses a velocity and acceleration-based detection algorithm. Using McMurray's (2002) EyeLinkAnal program in Microsoft Access, looks were grouped by area of interest (target, competitor, unrelated 1, unrelated 2, other), using a square port around each target picture. In all eye movement analyses, trials were excluded if there was more than 33% track loss during the critical window or if the participant failed to fixate either target or competitor for the entire trial (n=9).

We are particularly interested in the proportion of time spent looking at the target object and its cohort competitor, starting at the point when the target word is encountered. The window we analyzed here was 300 to 1000 ms after the onset of the target word, following Arnold (2008) and Dahan et al. (2002). We calculated the ratio of target looks to competitor looks, using the empirical logit (Barr, 2009): natural log ((# samples target looks + .05)/(# samples given looks + .05)). The competitor was the cohort item whose name sounded like the target (e.g., bacon for the target bagel).

Statistical analysis. The effects of experimental conditions were evaluated in a mixed-effects linear regression model, using SAS proc mixed, including random effects for both participants and items, as well as random slopes for manipulated variables with respect to both participants and items (Barr, Levy, Scheepers, & Tily, 2013). The predictors in the model were the three manipulated variables (distraction, prominence, givenness) and the interactions between them. These predictors were centered. Models included random intercepts for subject and item, as well as random slopes for the critical predictors. Trials were weighted according to the method proposed by Barr (2009): 1/(target looks +.05) + 1/(comp looks + .05).

Our models also included two binary control variables, which indexed 1) whether the participant was already fixating the target object at the onset of the target word or not, and 2) whether the participant was already fixating the competitor object at the onset of the target word or not. These controls were important, given that our stimuli required a discourse context. This meant that fixations at the onset of the target word were systematically influenced by the preceding context. Nevertheless, the same statistical patterns emerged on the critical variables even if these controls were left out of the model. Other control predictors were not included, since our experimental design was balanced.

Results and Discussion

Figure 2 displays timecourse plots of looks to targets and competitors, beginning at the onset of the target word. In all conditions, there was a baseline effect in which participants tended to be looking at the new cohorts (black lines) slightly more than the given cohorts at the moment when the target word is heard (at 0ms on the figure). Nevertheless, within 300 ms, looks to either given or new cohort objects began to rise. These looks were taken as an indication that the comprehender was considering that object as the likely referent for the incoming stimulus.

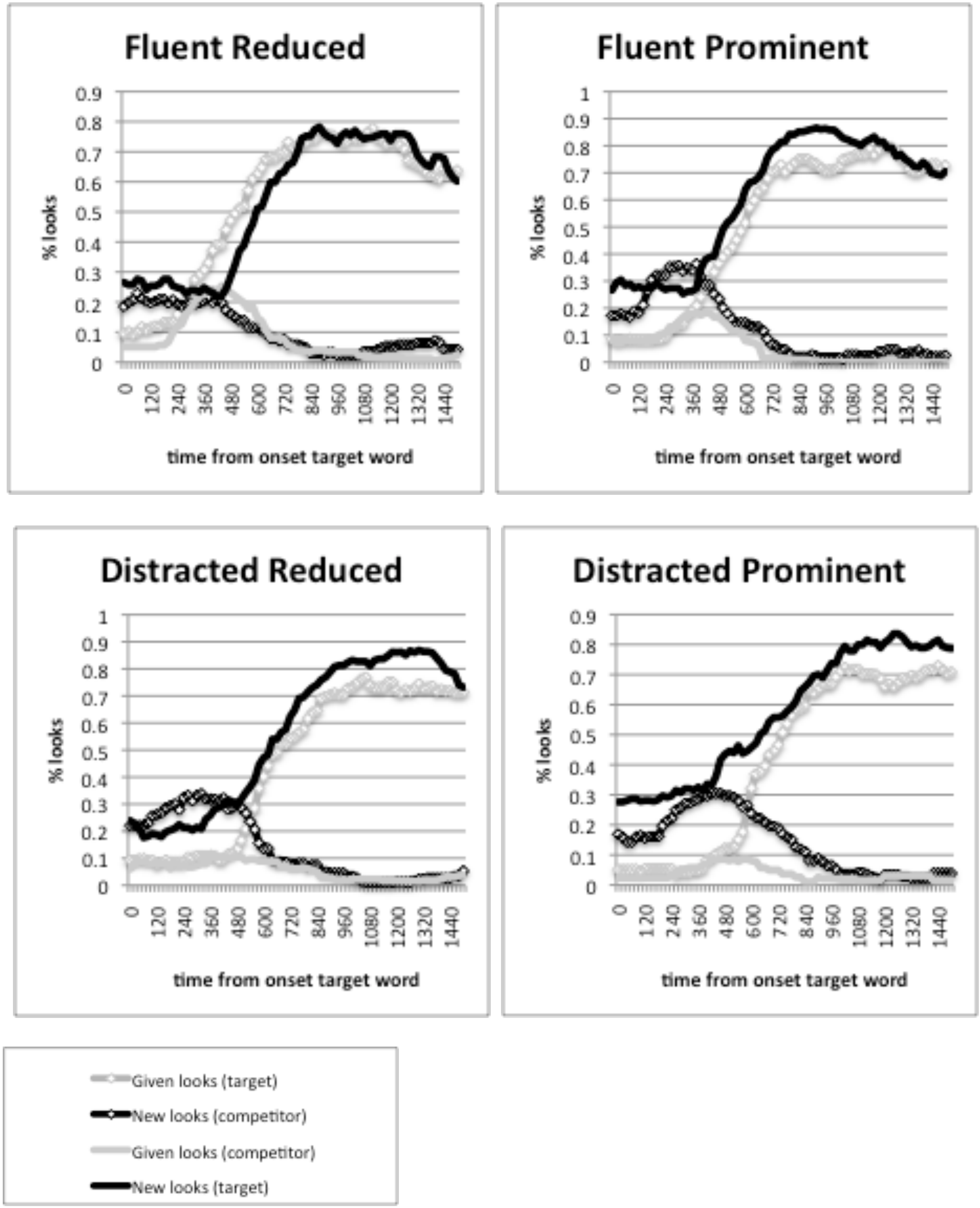

Figure 2. Results from Experiment 1. Proportion looks to given and new cohort objects at each time step, from the onset of the target word (0 ms) until 1500 ms after target word onset.

The comparisons of interest were between the given and new targets, and between given and new competitors. In the fluent condition, there was an early preference to look at the given target for reduced tokens, but the new target for prominent tokens. This replicates findings reported by both Dahan et al. (2001) and Arnold (2008) with the same paradigm. For the distracted condition, however, both prominent and reduced targets led to a bias toward the new cohort. These patterns can also be observed in Figure 3, which shows the average empirical logit of the target-to-competitor ratio, over the 300-1000 time window following target onset.

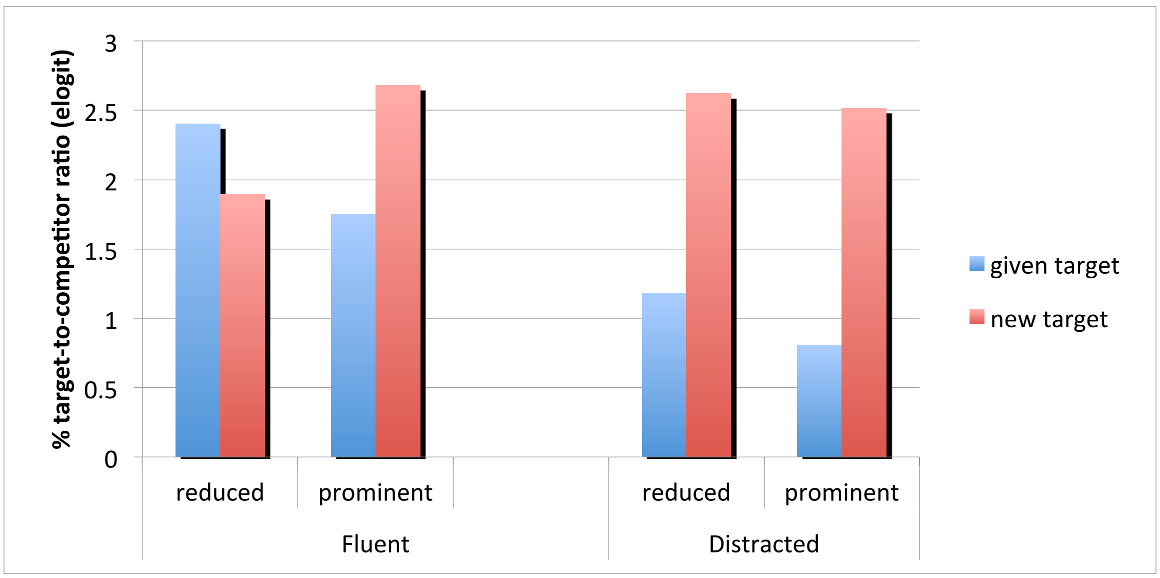

Figure 3. Exp. 1 results: Average empirical logit of the ratio of target-to-competitor looks during the period 300-1000 ms after target onset, in each condition.

s

The mixed-effects linear regression model confirmed these observations. Table 3 reports the parameter estimates, t-statistics, and p-values for each predictor in the model. The critical findings were the interactions with distraction. Distraction interacted with givenness: for distracted trials, the target ratio was higher when the target was new than when it was given; for fluent trials, there were about equal given- and new-target looks overall (although this pattern was modulated by a three-way interaction). There was also a three-way interaction between distraction, prominence and givenness. This reflects the fact that the prominence x target interaction held only for the fluent block.

In addition, we observed a main effect of distraction, which reflects the fact that there were more target looks overall in the fluent block. There was also a main effect of target, since there were generally more looks to the target when it was new than given. Prominence also interacted with givenness, in that looks to the given target were greatest for reduced tokens; this effect was carried by the fluent/reduced condition.

Table 3. Exp. 1: Results of the mixed effects analysis. Dependent variable = the empirical logit of the ratio of target looks to competitor-looks, from 300-1000 following the target onset.

	β (error)	DF	t	p	Sig.
FULL ANALYSIS					
Distraction (Distracted vs. Fluent block)	-0.62(0.18)	1005	-3.37	0.0008	**
Prominence (Prominent vs. Reduced)	-0.13(0.22)	1005	-0.58	0.5601	n.s.
Givenness (Given vs. New Target Referent)	-0.53(0.19)	1005	-2.83	0.005	**
Distraction * Prominence	-0.3(0.25)	1005	-1.17	0.241	n.s.
Distraction * Givenness	-1.1(0.25)	1005	-4.34	<.0001	***
Prominence * Givenness	-0.78(0.25)	1005	-3.08	0.002	**
Distraction * Prominence * Givenness	1.27(0.51)	1005	2.51	0.012	*
Looking at target at target word onset?	1.19(0.19)	1005	6.36	<.0001	***
Looking at comp. at target word onset?	-1.72(0.25)	1005	-6.98	<.0001	***
FLUENT ONLY					
Prominence (Prominent vs. Reduced)	0.04(0.25)	504	0.18	0.8601	n.s.
Givenness (Given vs. New Target Referent)	0.06(0.31)	504	0.19	0.8492	n.s.
Prominence * Givenness	-1.45(0.34)	504	-4.32	<.0001	***
Looking at target at target word onset?	1.03(0.25)	504	4.1	<.0001	***
Looking at comp. at target word onset?	-1.86(0.32)	504	-5.88	<.0001	***
DISTRACTED ONLY					
Prominence (Prominent vs. Reduced)	-0.3(0.28)	499	-1.07	0.2863	n.s.
Givenness (Given vs. New Target Referent)	-1.13(0.27)	499	-4.2	<.0001	***
Prominence * Givenness	-0.18(0.36)	499	-0.49	0.626	n.s.
Looking at target at target word onset?	1.38(0.27)	499	5.12	<.0001	***
Looking at comp. at target word onset?	-1.43(0.38)	499	-3.82	.0001	***

We also examined the fluent and distracted conditions in separate analyses, in order to directly assess the effect of acoustic prominence in each context. In the fluent condition, we found the critical interaction between target and prominence, replicating the findings of Arnold (2008) and Dahan et al. (2001). In the distracted condition, by contrast, there was no interaction, and only a main effect of givenness, due to the general preference for the new cohort in the distracted condition.

In all of these analyses, there were also robust effects of the control predictors: the target-competitor ratio was higher when the participant was already looking at the target when they first began to hear the target word, and it was lower when the participant was already looking at the competitor. This reflects the fact that our data consist of looks, which are categorical measures of where the participant was looking. People tended to look at the same object for at least a few hundred milliseconds. Thus, looks at the beginning of our measurement window are highly likely to be similar to the participant's location of fixation 300 msec earlier. Nevertheless, even taking this baseline effect into account, the other effects were robust. Additional analyses explored potential interactions between target-onset fixations and predictors of interest, and did not find any.

Our first analysis suggests that the distracted and fluent stimuli led to different effects of acoustic prominence on listeners' online biases. Yet we also know that the range of variability was greater in the fluent than distracted items. Does this difference account for our results? Or is there indeed less effect of acoustic prominence in the distracted items?

Another way of looking at these data is to consider the acoustic parameters to be the predictors of interest, rather than the conditions (see Table 4). The reduced conditions corresponded to shorter duration and lower average pitch than the prominent conditions, and the fluent conditions were shorter than the distracted conditions. However, these correlations were not perfect. Moreover, duration did not pattern as cleanly with prominence in the distracted tokens as it did in the fluent. Therefore, if there is any hint of an acoustic prominence effect in the distracted block, it might show up in an analysis using duration and pitch as predictors, rather than the reduced/prominent conditions per se. In these analyses, we used the residual duration of the first syllable as a predictor. The residual duration represents the relative duration of the first syllable, given its expected length based on number of syllables and phonemes in the word, and the word identity as a random effect. The duration of the first syllable reflects the earliest information listeners receive about the word, which is hypothesized to guide their on-line comprehension biases.

Across all distracted tokens, there was slightly less variation in the average pitch for distracted (max: 282, min: 191, avg. 227, st. dev. 21) than fluent tokens (max: 271 Hz; min: 176; avg: 228; st. dev: 29), but they were roughly similar. There was greater variation in residual duration of the first syllable in the distracted tokens (max: 252, min: -52; avg: 62; st. dev. = 64) than the fluent tokens (max: 31, min: -208; avg: -62, st. dev.: 52), even though duration did not pattern with the reduced/prominent tokens for distracted utterances. Thus, the wider duration variation in the distracted condition might lead to a stronger effect.

The results of the analyses with duration and pitch as predictors are presented in Table 4. This analysis mirrored the previous one, in that again we found main effects of distraction and givenness, although the distraction effect barely missed significance ($p=.051$). Of interest is the finding that givenness interacted only with duration, and not with pitch. This suggests that the givenness x prominence interaction in the first analysis was carried by duration.

Table 4. Exp. 1: Statistical results from an analysis using duration and pitch as predictors instead of the prominence manipulation. Dependent variable = the empirical logit of the ratio of target looks to competitor-looks, from 300-1000 following the target onset.

	β (error)	DF	t	p	sig
FULL ANALYSIS					
Distraction	-0.505(0.259)	1003	-1.95	0.0513	(*)
First syllable residual duration	-0.001(0.002)	1003	-0.45	0.6529	n.s.
Target pitch	-0.001(0.006)	1003	-0.14	0.8896	n.s.
Givenness (Given vs. New Target Referent)	-0.76(0.275)	1003	-2.77	0.0058	***
Distraction x Resid. Duration	-0.005(0.004)	1003	-1.3	0.194	n.s.
Distraction x Pitch	-0.004(0.007)	1003	-0.58	0.5631	n.s.
Distraction x Givenness	-0.009(0.417)	1003	-0.02	0.98	n.s.
Resid. Duration x Givenness	-0.008(0.003)	1003	-3	0.0028	**
Pitch x Givenness	-0.006(0.006)	1003	-0.92	0.3556	n.s.
Resid. Duration x Givenness & Distraction	-0.004(0.006)	1003	-0.62	0.5334	n.s.
Pitch x Givenness & Distraction	0.024(0.012)	1003	1.98	0.0475	*
JUST DISTRACTED					
Givenness	-1.921(0.3)	499	-6.41	<.0001	***
First syllable residual duration	-0.004(0.002)	499	-2.19	0.0289	*
Target pitch	0.001(0.005)	499	0.17	0.8612	n.s.
Resid. Duration x Givenness	-0.007(0.004)	499	-2.03	0.0429	*
Pitch x Givenness	0.004(0.01)	499	0.35	0.727	n.s.
JUST FLUENT					
Givenness	0.264(0.389)	504	0.68	0.497	n.s.
First syllable residual duration	0.002(0.003)	504	0.94	0.3466	n.s.
Target pitch	0.001(0.004)	504	0.24	0.8136	n.s.
Resid. Duration x Givenness	-0.009(0.004)	504	-2.24	0.0253	*
Pitch x Givenness	-0.023(0.007)	504	-3.22	0.001	**

NOTE: These analysis did not include fixations at target onset, for simplicity; adding these to the model yields the same pattern of effects (except the main effect of givenness becomes a marginal effect in both the full analysis and the fluent-only analysis).

Also of interest is the fact that the distraction x duration x givenness effect was not significant. This suggests that duration leads to similar information-status biases in both the distracted and fluent blocks. Since the distracted tokens are longer overall than the fluent ones, they support a new-target bias. By contrast, we did find a three-way interaction with pitch (distraction x pitch x givenness). Pitch interacted with givenness for fluent, but not distracted tokens. This may have been due to the fact that there was greater pitch variability in the fluent block. On the other hand, it may also signal that the effect of pitch is disrupted in the distracted condition, where duration does not support the prominence signaled by pitch (see also Isaacs & Watson, 2011).

Another finding worth noting was that the Distraction x Givenness interaction was not significant, in contrast with the analysis in Table 3 (although in the individual analyses, Givenness was significant only for Distracted tokens, and not Fluent ones). This finding also speaks against one interpretation of the distraction effect. Since the fluent and distracted trials were presented in blocks, it is possible that participants could have inferred that the disfluent-sounding distracted trials would all be more likely to have new referents. If so, they may have generated a general expectation for new referents in this block, independently of the acoustic properties of the stimuli. Yet instead we found that duration x givenness was a stronger predictor than distraction x givenness, suggesting that the properties of the stimuli did matter in the distracted condition.

This interpretation is also supported by an analysis of distracted and fluent trials separately. In the distracted trials, we found a main effect of givenness (i.e., more looks to new targets overall), a main effect of duration (higher target ratio for shorter target words), and an interaction givenness x duration. There were no effects of pitch. By contrast, in the fluent trials, the only effects were the interactions givenness x pitch and givenness x duration.

Thus, duration signaled information status for both fluent and distracted tokens, but pitch did so only for fluent tokens. In addition, there was a general new bias (main effect of givenness) for distracted trials, but not for fluent trials.

In summary, there were two findings of interest. First, distracted-sounding speech led to an overall bias toward the discourse-new object. In both the reduced and prominent conditions, participants were more likely to look at the new than the given cohort object. One potential concern is that this bias is confounded with the baseline effect. In this paradigm, participants tend to look away from the object they have just moved, mostly likely because they have already processed its visual features, and are well aware of its location (see also Arnold, 1998; Brown et al., under review). Thus, the absolute number of looks to each cohort object is not a transparent indicator of which object is considered the better referent. Instead, looks should be interpreted as a signal of **relative** bias. In the fluent/reduced condition, the baseline new bias did not interfere with a rapid surge in looks to the given cohort. In contrast with this condition, the other three exhibited a relative new-cohort bias. These effects emerged even though the baseline effect was controlled in the model.

Second, acoustic prominence did not provide as strong a signal of information status in the distracted condition, compared to the fluent condition, such that there was no effect or interaction of prominence in the distracted analysis alone. In the fluent block, we replicated the interaction found in previous studies (Arnold, 2008; Dahan et al., 2001), in which reduced tokens led to more given-target looks, and prominent tokens led to more new-target looks. However, in the distracted condition, this effect disappeared.

There are two potential interpretations of the distraction effects that we found here. The most theoretically interesting interpretation is that distraction disrupts the function of prosody as a marker of information status, such that even relative prominence is not a good signal of information status. There are a number of possible explanations for why this might be, which we will take up in the general discussion.

Yet we must consider an alternative explanation, which is that perhaps there is a relative effect of prominence, even in distracted trials, but our materials were not set up to find it. The fluent and distracted targets were not perfectly matched in terms of the acoustic properties of the reduced and prominent conditions. This first experiment prioritized the naturalness of the spoken instructions, which were recorded as a whole, and recombined with entire context sentences. In essence, ER's portrayal of distraction was a performance. Thus, it is possible that the greater variability in the distracted tokens reflects a natural tendency for distraction to reduce the systematicity of prosodic cues. On the other hand, it is also possible that this variability in the stimuli is what accounted for the difference between contexts. If the distracted tokens sounded less prominent in the prominent condition, and less reduced in the reduced condition, it may have reduced the size of the effect.

The secondary analyses (using pitch and duration; Table 4) suggest that both of these interpretations may be partially right. We found an interaction between duration and givenness in all three analyses in Table 4, suggesting that duration provides an information-status signal for both fluent and distracted utterances. Nevertheless, we also found a main effect of givenness in in the distracted-trials analysis, showing that there was a general bias toward new targets overall. By contrast, this main effect did not obtain in the fluent-trials analysis. This suggests that while distraction may not fully disrupt the information-status role of prosody, it may still skew the interpretation towards new referents.

To further explore these possibilities, Experiment 2 sought to replicate the distraction effect using stimuli that were more closely matched on acoustic properties.

Experiment 2

The goals with this experiment were to replicate the findings from Experiment 1, but with a more tightly controlled set of auditory stimuli. While the sentences were produced naturally in experiment 1, here we created the distracted tokens by acoustically manipulating the fluent recordings from experiment 1. This meant that our stimuli were matched across conditions on all properties except duration and pausing.

Methods

Participants

41 native speakers of English participated in exchange for course credit. Five were excluded (2 for excessive track loss; 1 for technical problems, and 2 in order to even up the lists). This left 36 participants in the analysis, 4 or 5 on each list.

Design, Equipment, and Procedure

The exact same design, equipment, procedure, and visual stimuli were used as in Experiment 1. The only difference was in the auditory stimuli.

Stimulus design. The stimuli for the fluent conditions were the exact same recordings as those used in Experiment 1, except for a few. The new distracted stimuli were created by following the procedure in Figure 4. First we copied the target words from the fluent set into the distracted instructions, using the software Praat. We then lengthened all of the spliced target words until they sounded naturalistic in the distracted context. In some cases pauses were added or lengthened either before or after the spliced target word in order to make the utterance sound natural. In a few (n=9) instances the resulting utterances failed to sound naturalistic and were instead re-recorded by ER.

Step 1: Cut the target word out of the fluent utterance	Now put the [bacon] on the....
Step 2: Splice the target into the disfluent context sentence	Now . . put . . the . . [bacon] . . on . the....
Step 3: Lengthen the target to make it fit the distracted context	⇐ ⇒ Now . . put . . the ..[b a c o n].. on . the....

Figure 4. Procedure for creating the distracted stimuli in Experiment 2.

The final set of distracted stimuli matched the fluent stimuli in target pitch (see Table 5), such that for both sets of stimuli, the prominent condition used a higher average, max, and mean pitch than the reduced condition. In both distracted and fluent stimuli, the prominent condition was also longer than the reduced condition. Thus, in this experiment, prominence was marked by both pitch and duration for the distracted items as well as for the fluent items.

Table 5. Prosodic properties of auditory stimuli for Exp. 2.

Condition	Average duration	Average pitch	Max pitch	Min pitch
Fluent / prominent	481	251	308	194
Fluent / reduced	428	206	230	182
Distracted / prominent	686.	251	310	191
Distracted / reduced	645	203	223	183

Results and Discussion

Figure 5 displays timecourse plots of looks to given and new cohort items, beginning at the onset of the target word. The eye movement plots looked much like the ones in Experiment 1, although there were some differences. As before, the fluent/reduced condition displayed the most early looks to the given target, compared with the other three conditions. Also, as before, there were more new-target looks overall in the distracted conditions. As shown in Figures 5 and 6, both distracted conditions led to a greater target-competitor ratio for the new-target conditions, and a smaller one for the given-target conditions. This finding is reflected in a significant interaction between distraction and givenness (see table 6), replicating this result from experiment 1. The statistical analysis also revealed a main effect of distraction (more target looks in the fluent condition), a main effect of givenness (higher target ratio for new targets overall), and an interaction between prominence and givenness, reflecting the critical interaction found in Experiment 1 as well.

The most important difference from Experiment 1 was that there was no three-way interaction between distraction, prominence, and target. This is interesting because it shows that with more closely controlled stimuli, we found an interaction between prominence and givenness for both fluent and distracted stimuli. Moreover, when we analyzed each condition separately, we found this interaction for both fluent and distracted conditions. An examination of the fluent conditions in figures 5 and 6 suggests that the interaction looks visually smaller than in Experiment 1. This is somewhat surprising, since the fluent conditions used nearly identical auditory stimuli; we expect that this difference represents normal variation. Nevertheless, the contrast between prominent and reduced conditions had the expected effect.

Yet we still found a contrast between the fluent and distracted trials, in that the distracted condition displayed a consistent bias toward the new cohort objects. This was reflected in the interaction between distraction and givenness, which showed that the new-objects bias was stronger in the distracted than the fluent trials. In addition, the distracted-trials-only analysis found a main effect of givenness (i.e., more target looks when it was new), while the fluent-trials analysis also showed a marginal effect of givenness.

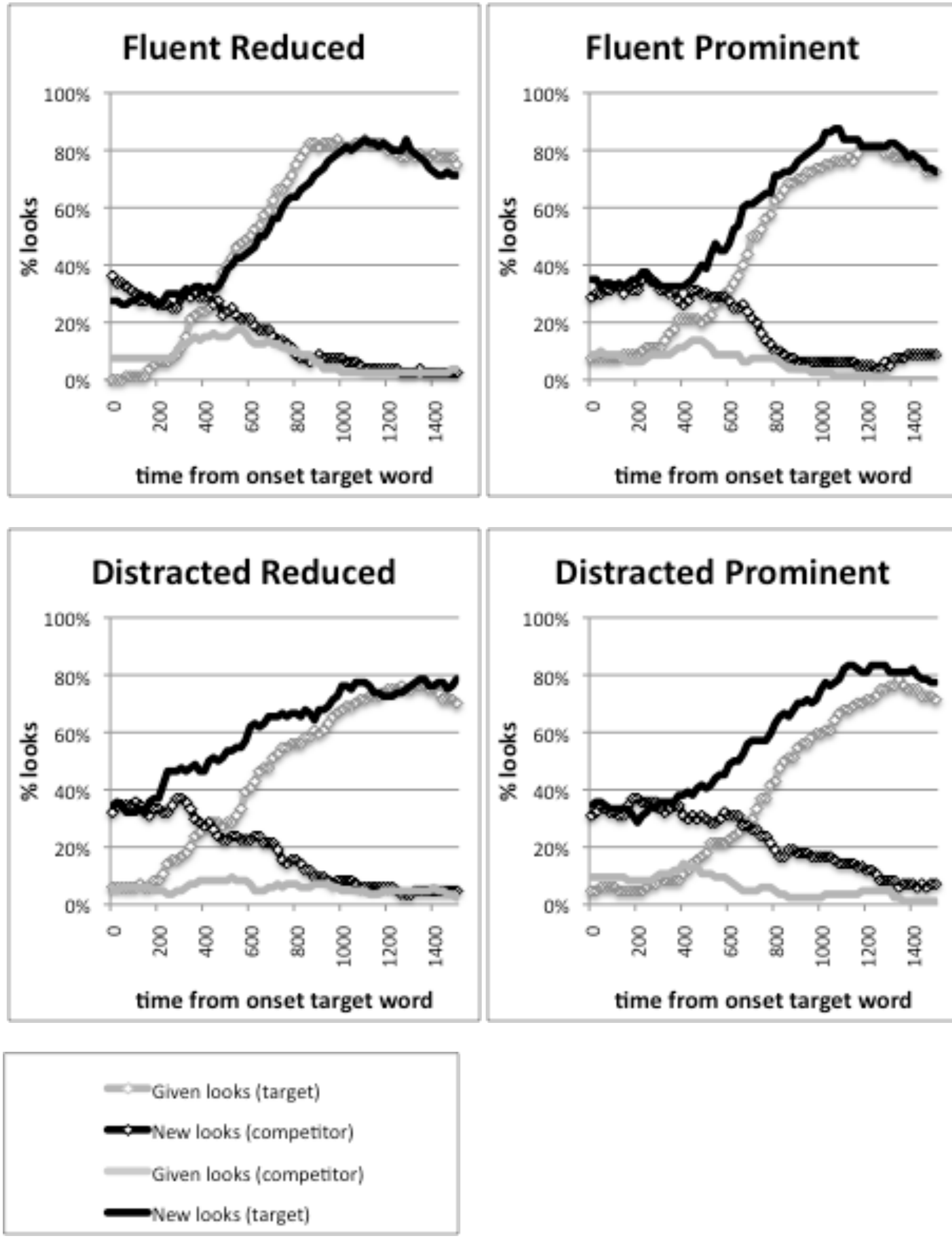

Figure 5. Results from Experiment 2. Proportion looks to given and new cohort objects, from the onset of the target word (0 ms) until 1500 ms after target word onset.

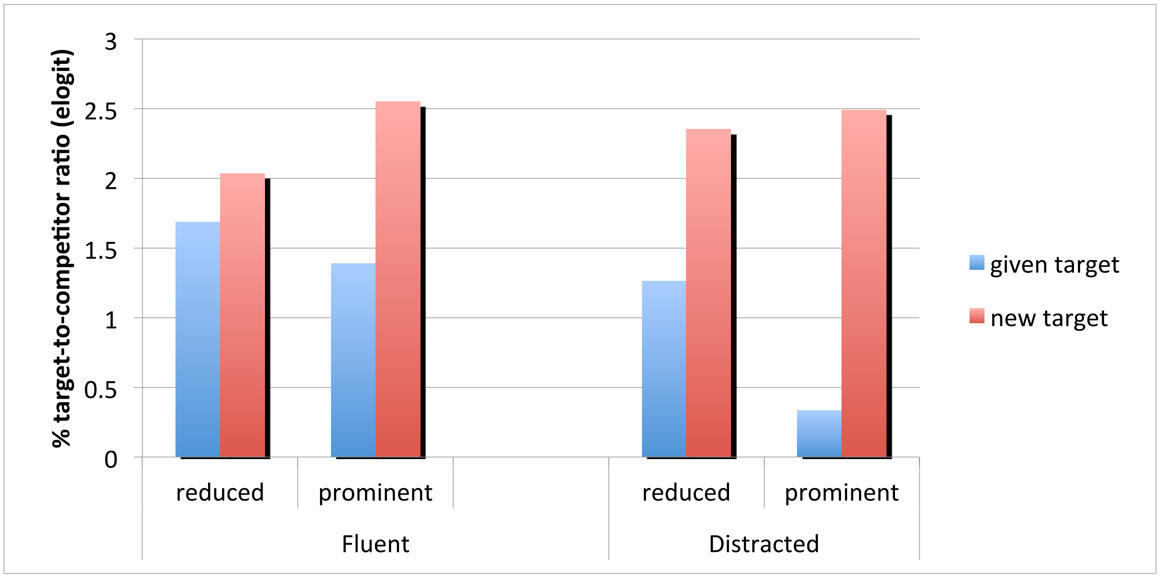

Figure 6. Exp. 2 results: Average empirical logit of the ratio of target-to-competitor looks during the period 300-1000 ms after target onset, in each condition

In summary, Experiment 2 demonstrated two things. 1) The longer pronunciation of both "reduced" and "prominent" words in a distracted context led to a stronger expectation of a discourse-new referent than in a fluent context, despite the fact that the distracted context supported this longer pronunciation. Thus, the information-status function of prosody was disrupted by a general new-bias in the distracted condition. 2) There is nevertheless a relative effect of prominence in the distracted context, such that reduced tokens lead to relatively more given-object looks than prominent tokens do.

Table 6. Experiment 2, Results of the mixed effects analysis. Dependent variable = the empirical logit of the ratio of target looks to competitor-looks, from 300-1000 following the target onset.

	β (error)	DF	t	p	Sig.
FULL ANALYSIS					
Distraction (Distracted vs. Fluent block)	-0.48(0.22)	1128	-2.22	0.0269	*
Prominence (Prominent vs. Reduced)	-0.14(0.15)	1128	-0.94	0.349	n.s.
Givenness (Given vs. New Target Referent)	-0.51(0.24)	1128	-2.1	0.036	*
Distraction * Prominence	-0.36(0.24)	1128	-1.49	0.136	n.s.
Distraction * Givenness	-0.78(0.24)	1128	-3.25	0.001	**
Prominence * Givenness	-0.99(0.24)	1128	-4.13	<.0001	***
Distraction * Prominence * Givenness	0.15(0.48)	1128	0.3	0.762	n.s.
Looking at target at target word onset?	1.18(0.19)	1128	6.21	<.0001	***
Looking at comp. at target word onset?	-1.4(0.2)	1128	-7.07	<.0001	***
FLUENT ONLY					
Prominence (Prominent vs. Reduced)	0.04(0.18)	567	0.23	0.8177	n.s.
Givenness (Given vs. New Target Referent)	-0.13(0.28)	567	-0.44	0.658	n.s.
Prominence * Givenness	-1.07(0.34)	567	-3.11	0.002	**
Looking at target at target word onset?	0.97(0.3)	567	3.25	0.001	**
Looking at comp. at target word onset?	-1.54(0.3)	567	-5.09	<.0001	***
DISTRACTED ONLY					
Prominence (Prominent vs. Reduced)	-0.33(0.25)	559	-1.3	0.1925	n.s.
Givenness (Given vs. New Target Referent)	-0.89(0.27)	559	-3.34	0.001	**
Prominence * Givenness	-0.98(0.33)	559	-3	0.003	**
Looking at target at target word onset?	1.26(0.25)	559	5.06	<.0001	***
Looking at comp. at target word onset?	-1.43(0.26)	559	-5.51	<.0001	***

General discussion

The results of two experiments revealed two primary findings. First, we observed a small, relative effect of acoustic prominence in the distracted condition. In Experiment 1, this only emerged in the second analysis, where duration interacted with givenness for distracted trials. In Experiment 2, both pitch and duration were manipulated to reflect prominence, and we found an interaction between the prominence condition and givenness. This finding is consistent the prediction that a word's acoustic prominence is perceived relative to the speed of the utterance that the word occurs in.

The second and more important finding sits on top of the relative effect of prominence: listeners responded differently to acoustic prominence from distracted and fluent speakers. In particular, distraction appeared to disrupt the information-status function of prosody. The fluent trials replicated the well-known tendency for reduced and unaccented words to lead to an expectation for a given referent. However, this effect was heavily weakened in the distracted condition. The most consistent indicator of this contrast was that distraction interacted with givenness in both experiments, revealing a general preference to look at the discourse-new cohort item in the distracted conditions. This general new bias dampened the effect of prominence in the distracted condition.

There are several possible interpretations of the distraction effect. Here we consider four explanation, which we refer to as: 1) absolute prominence, 2) f0 slope, 3) all bets are off, and 4) distraction as disfluency explanations.

One possibility (the "absolute prominence" explanation) stems from the observation that the distracted target words were objectively more prominent overall than the fluent ones. The distracted targets were longer than the fluent ones, even the ones in the reduced condition. If duration contributes to the perception of acoustic prominence, this should make the targets sound even more prominent. Duration could contribute to perceived prominence in one of two ways. Perceived prominence might be a function of the number of acoustic cues that sound prominent. Thus, long duration in the absence of a pitch excursion may sound partially prominent. Or it may be that listeners are most influenced by acoustic cues that are consistent with each other, so when duration and pitch together signal prominence, it serves as a stronger cue to information status than when the cues are inconsistent (as they were in the distracted block).

This interpretation is supported by the fact that in Experiment 1, the ratio of target-competitor looks was predicted by an interaction between duration and givenness. That is, longer targets led to a greater preference for discourse-new referents. This finding contrasts with a straightforward application of the assumption that the criteria for using length as a cue to prominence should be set in relation to the speech rate of the surrounding discourse (e.g., see Bell et al., 2009; Ladd, 1996). Even though we found a small relative effect of prominence, our findings suggest that the interpretation of acoustical cues may be driven by absolute prominence as well, where the strong discourse-given bias is reserved for cases where the target is highly reduced.

Another possibility (the "f0 slope explanation") is that the distraction effect is driven by differences in f0 slope. Isaacs & Watson found stronger discourse biases for shorter than longer target words, even though both had the same pitch excursion, and suggested the difference may have been due to the faster f0 rise in shorter words. Our fluent and distracted stimuli differed in duration, which mean that even when f0 was comparable (as in Exp. 2), the slope should still differ. In the distracted stimuli our targets were manipulated to be longer, which likely led to a slower rise in f0. This flatter slope may have decreased the perception of prominence and the strength of the resulting discourse-new bias.

An alternative explanation, which we deem unlikely, is that distraction signals that "all bets are off". Perhaps in a distracted context, listeners stop drawing inferences about the likely referent on the basis of prosodic prominence. An extreme version of this explanation would suggest that listeners cease processing incrementally, and simply wait to hear the entire expression. However, this does not seem to be the case here. The distracted tokens were about 200 msec longer on average than the fluent tokens. If listeners were waiting to hear the entire word before beginning to process it, we would expect a much later rise in fixations in the distracted than fluent contexts. Nevertheless, a visual inspection of looks in Figures 2 and 5 shows that looks to the new targets began to rise around the same time (by 400 msec after target onset) for fluent/prominent and both distracted contexts, for both experiments. Even a weaker version of this explanation would not explain why we still see a relative prominence effect in the distracted trials in Experiment 2.

A final possibility is that distraction in this experiment functions like disfluency does in general. As observed by Arnold et al. (2004), a disfluent utterance leads to an expectation of a discourse-new referent. They offered two possible explanations. One is that listeners made an attribution of the disfluency to the speech context. Reference to new information is plausibly related to speech difficulty, and listeners may have inferred that a disfluent phrase referred to something not previously mentioned. This explanation seems unlikely to account for the current findings, since the disfluency stemmed from distraction (a secondary task), and therefore should not have been attributed to difficulty with new information. On the other hand, another possibility is that Arnold et al. (2004)'s data resulted from an automatic association between disfluency and reference to new information, based on the fact that reference to new information leads to a higher disfluency rate than reference to given information does (Arnold & Tanenhaus, 2011). This kind of automatic effect may have contributed to the current findings. However, other research suggests that disfluency effects are not simple automatic associations (Arnold et al., 2007), making this explanation less likely.

In summary, the contrast between distracted and fluent conditions is most likely explained by the properties of the stimuli. Under the "absolute prominence" explanation, distracted speech leads to a greater new-information bias because of the absolute prominence of the speaker's pronunciations. This view suggests that prominence is shaped by multiple production constraints simultaneously, creating a single signal for comprehenders to interpret. Sometimes a referring expression is prominent and accented because the speaker intends to signal that the referent is new or contrastive. Sometimes the referring expression is lengthened and prominent-sounding because the speaker is disfluent, as occurs with distraction. In either case, the listener hears prominence and automatically orients towards discourse-new referents, in accordance with the most typical usage of prominent prosodic markings. An alternate interpretation stems from the f0 slope-differences explanation. On this view, the speaker intends to mark prominence prosodically, but the hesitation resulting from distraction interrupts the speaker's ability to do so effectively. Either way, it appears that reduced tokens facilitate comprehension of reference to given information, but only when the entire context supports a clear perception of the token's reduction, i.e. in a fluent context.

Regardless of why our distraction effect occurred, it clearly showed that distraction supports a bias toward discourse-new information. We might have found instead that distraction would cause listeners to shift their category of "reduced" in the distracted context, and lead to a given preference even for the prominent tokens. That is, listeners may have heard prominence, and mentally "excused" it as the effect distraction, and not discourse-new status per se. However, we found that listeners do not appear to excuse distraction. Instead, distraction enhanced the tendency to expect new-target references.

In this sense, our findings initially appear to differ from those of Clifton et al. (2006) and Brown et al. (2013), who both reported that the effect of prosody was contingent on the context. By contrast, we found that duration increased discourse-new looks, even in the context of a long and distracted utterance. However, our experiment differed from both of theirs, in that the distracted tokens were even longer than the fluent ones. In addition, the relative effect of prominence observed in Exp. 2 is consistent with the broader conclusion that the context of a word affects how its prosody is interpreted, similar to the conclusions of Clifton et al. and Brown et al. Moreover, our results are consistent with Brown et al.'s finding that the information-status effect of prosody is primary. Here too, our distracted and relatively long "reduced" tokens are taken to signal discourse newness, even in a context in which they might have been taken merely to signal distractedness.

Our findings are consistent with the view that multiple signals support the interpretation of a word as discourse-given or new. Our paradigm was set up to highlight the listeners' on-line biases, by including a cohort competitor. This competitor object provided an alternative interpretation for the target word that was temporarily plausible. When the prosodic cues supported this interpretation, we observed greater competition from the cohort competitor. Under these conditions, we saw variation in the target-competitor ratio, which indexed the strength of the listeners' bias toward the target referent as the target word unfolded.

This variation supported the assumption that both duration and pitch variation contribute to information status biases during comprehension, and they appear to contribute independently. For example, in the distracted trials of experiment 1, we only saw an interaction between duration and givenness, and not an interaction between pitch and givenness. This shows that even when pitch is not informative (possibly due to the moderate degree of pitch variation in these items), duration is still informative about information status. These findings are consistent with the idea that accenting and acoustic prominence are gradient markers of information status (Baumann, Grice, & Steindamm, 2006).

In sum, our findings show that duration contributes to prosodic prominence independently from pitch. This is important, because the duration of target words can vary for a number of different reasons that are not related to accenting or information status. One such source of duration variation is distraction, which we explored here. With long, hesitant instructions, there was no preference for discourse-given referents, even with relatively reduced tokens.

More broadly, our results suggest that listening to distracted speakers interrupts the normal ability to use prosodic cues to information status. This confirms the intuitive sense that it's hard to converse with an inattentive interlocutor, and shows one specific way in which distraction affects the signal. These results strongly support the conclusion that distraction matters for reference comprehension.

AUTHOR NOTE

This research was supported by NSF grant BCS-0745627. We are grateful to the following people for their assistance collecting and analyzing the data: Lap-Ching Keung, Jennifer Little, Sarah Lineberry, Richard Harenberg, Michael Parrish.

REFERENCES

Arnold, J. E. (1998). Reference form and discourse patterns. Unpublished doctoral dissertation, Stanford University.

Arnold, J. E. (2008). *THE BACON* not *the bacon*: How children and adults understand accented and deaccented noun phrases. *Cognition, 108*, 69-99.

Arnold, J. E. & Tanenhaus, M. K. (2011). Disfluency effects in comprehension: how new information can become accessible. In Gibson, E., and Perlmutter, N. (Eds), *The processing and acquisition of reference*, Cambridge: MIT Press, pp. 197-217.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 914-30. doi:10.1037/0278-7393.33.5.914

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: disfluency and reference resolution. *Psychological Science, 15*(9), 578-582. doi:10.1111/j.0956-7976.2004.00723.x

Arnold, J. E. & Watson, D. G. (under review). Synthesizing meaning and processing approaches to prosody: performance matters. Ms., University of North Carolina at Chapel Hill.

Bard, E. G., & Aylett, M. P. (1999). The dissociation of deaccenting, givenness and syntactic role in spontaneous speech. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), Proceedings of the XIVth International Congress of Phonetic Science (pp. 1753–1756). San Francisco.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cave´, I. Gua¨itella, & S. Santi (Eds.), *Oralite´ et gestualite´: Interactions et comportements multimodaux dans la communication* [Orality and gestuality: Multimodal interaction and behavior in communication] (pp. 597–600). Paris: L'Harmattan.

Barr, D. J. (2009). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language,* 457-454.

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(3):255-278.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*(10), 1636-1657. doi:10.1016/j.pragma.2005.03.017

Baumann, S., Grice, M. & Steindamm, S. (2006). Prosodic Marking of Focus Domains - Categorical or Gradient? Proceedings SpeechProsody 2006, Dresden, Germany. 301-304.

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin & Review*, *17*(1), 15-21. doi:10.3758/PBR.17.1.15

Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labelling, Version 3. The Ohio State University Research Foundation. Available from: http://www.ling.ohio-state.edu/~tobi/ame_tobi/ labelling_guide_v3.pdf.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60*(1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C. Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113(2),* 1001-1024.

Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General, 134*, 207-221.

Birch, S., & Clifton, C., Jr. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. Journal of Memory and Language, 47, 571-588.

Breen, M., Fedorenko, E., Wagner, M., Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*. *25* (7), 1044-1098.

Briem, V., & Hedman, L. R. (1995). Behavioural effects of mobile telephone use during simulated driving. *Ergonomics, 38*(12), 2536-2562.

Brown, G. (1983). Prosodic structures and the Given/New distinction. In D. R. Ladd & A. Cutler (eds.*), Prosody: Models and measurements, 67–77*. Berlin: Springer.

Brown, M., Salverda, A. P., Gunlogson, C., and Tanenhaus, M. K. (under review). Rapid integration of prosodic and discourse cues during spoken-word recognition. *Language and Cognitive Processes*.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), Subject and topic (pp. 25–56). New York: Academic Press.

Christodoulou, A. 2009. Thinking prosody: how speakers indicate production difficulty through prosody. Chapel Hill, NC: University of North Carolina at Chapel Hill, Unpublished Masters Thesis.

Clifton, C., Jr., Frazier, L., & Carlson, K. (2006). Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin & Review*, 13, 854-861.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect the language comprehension. Cognition, 105, 658-668.

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.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, *22*, 1151-1177.

Fowler, C., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*(5), 489-504.

Gahl, S. (2008). "Thyme" and "Time" are not homophones. Word durations in spontaneous speech. Language 84(3), 474-496.

Gahl, S., Yao, Y., & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language 66,* 789-806.

Gillespie, M. (2011). *Agreement computation in sentence production conceptual and temporal factors*. Unpublished doctoral dissertation, Northeastern University, Boston, MA.

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech. New York: Academic Press.

Gussenhoven, C. (1983). *A semantic analysis of the nuclear tones of English*. Bloomington, Indiana: Indiana University Linguistics Club.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Hirschberg, J. (1995). Pitch accent in context: Predicting intonational prominence from text. Artificial Intelligence, 63, 305–340.

Isaacs, A. M. & Watson, D. G. (2010). Accent detection is a slippery slope: Direction and rate of F0 change drives comprehension. *Language and Cognitive Processes*, 25, 1178-1200.

Jurafsky, D, Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In Bybee, Joan and Paul Hopper (Eds.). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. 229-254.

Kahn, J. & Arnold, J.E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*.

Kahn, J. & Arnold, J.E. (under review). *Lexical activation effects on spoken word reduction*. Ms., University of North Carolina at Chapel Hill.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194-231.

Kubose, T. T., Bock, J. K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology, 20*, 43-63. DOI: 10.1002/acp.1164

Ladd, R. (1996). *Intonational Phonology***.** Cambridge: University Press.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: why repeated referents have reduced prominence. *Memory & Cognition*, *38*(8), 1137-46. doi:10.3758/MC.38.8.1137

Longhurst, E. (2006). ExBuilder. [Computer program]

McMurray, B. (2002). EyelinkAnal, version 1.5. Analysis scripts written in Microsoft Access.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.*), Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Prince, E. (1981). Toward a taxonomy of given-new information. In P.Cole, (Ed.). *Radical Pragmatics* (pp. 223-256). NY: Academic Press.

Prince, E. (1992). The ZPG letter: subjects, definiteness, and information-status. In Thompson, S. and Mann, W., (Eds.). *Discourse description: diverse analyses of a fund raising text* (pp. 295-325). Philadelphia/Amsterdam: John Benjamins B.V.

Rooth, M. (1992). A theory of focus interpretation. Natural Language Semantics, 1, 75 – 116.

Rosa, E. , Finch, K., Bergeson, M. & Arnold, J.E. (2013). The Effects of Addressee Attention on Prosodic Prominence. *Language and Cognitive Processes*.

Rosa, E. C. & Arnold, J. E. (2011). The role of attention in choice of referring expressions. *Proceedings of PRE-Cogsci: Bridging the gap between computational, empirical and theoretical approaches to reference*, Boston, 2011.

Schafer, A., Speer, S., Warren, P., & White, S. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*, 169-182.

Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics, 7,* 141–177.

Schwebel D, Stavrinos D, Byington K, Davis T, O'Neal E, de Jong D. (2012). Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis And Prevention*, *45*, 266-271. Doi: 10.1016/j.aap.2011.07.011

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103-130.

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.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language & Cognitive Processes,19*, 713-755