

Effects of participant engagement on prosodic prominence

Andrés Buxó-Lugo
Dept. of Psychology and Beckman Institute
University of Illinois at Urbana-Champaign

Joseph C. Toscano
Dept. of Psychology
Villanova University

Duane G. Watson
Dept. of Psychology and Beckman Institute
University of Illinois at Urbana-Champaign

Corresponding author:

Andrés Buxó-Lugo
Dept. of Psychology
University of Illinois
603 E Daniel St.
Champaign, IL 61820
Phone: 919-259-5688
Fax: (217) 244-5876
Email: buxo2@illinois.edu

Effects of participant engagement on prosodic prominence

It is generally assumed that prosodic cues that provide linguistic information, like discourse status, are driven primarily by the information structure of the conversation. This paper investigates whether speakers have the capacity to adjust subtle acoustic-phonetic properties of the prosodic signal when they find themselves in contexts in which accurate communication is important. Thus, we examine whether the communicative context, in addition to discourse structure, modulates prosodic choices when speakers produce acoustic prominence. We manipulated the discourse status of target words in the context of a highly communicative task (i.e., working with a partner to solve puzzles in the computer game Minecraft), and in the context of a less communicative task more typical of psycholinguistic experiments (i.e., picture description). Speakers in the more communicative task produced prosodic cues to discourse structure that were more discriminable than those in the less communicative task. In a second experiment, we found that the presence or absence of a conversational partner drove some, but not all, of these effects. Together, these results suggest that speakers can modulate the prosodic signal in response to the communicative and social context.

In English, speakers tend to mark information that is new or unpredictable with prosodic prominence (e.g. Halliday, 1967; Fowler and Housum, 1987; Eady et al., 1986; Bard et al., 2000; Breen et al., 2010). Previous work has suggested that this prominence correlates acoustically with one or more different cues: an increase in the intensity of the sound, lengthening of the prominent word, and/or a change in fundamental frequency (F0). Researchers have typically assumed that speakers' decisions about which words are prominent are driven by grammatical knowledge (e.g., grammatical rules derived from syntax, information status, or phonology) that map the information status of words and phrases onto their acoustic realizations. However it is possible that speakers signal information status differently in different communicative contexts. In this paper, we investigate how the communicative context interacts with information structure to elicit different prosodic productions from speakers.

Communicative context has been found to be important in many areas of language production. For example, Brown-Schmidt (2009) found that partner-specific interpretation and perspective taking is more likely to occur in interactive dialogue settings. Furthermore, a speaker is more likely to take the addressee's perspective when it is more relevant to utterance goals (Yoon, Koh, & Brown-Schmidt, 2012). In fact, Clark (1997) argued that language, being primarily a joint activity, ought not to be studied "in a vacuum." Thus, a great deal of work suggests that interactive language use may differ in fundamental ways from less-interactive language use (Schober & Clark, 1989, Brown-Schmidt, 2005).

Because context effects appear to be ubiquitous in language production, we are interested in whether this is true specifically for the production of prosodic cues. Much of the work on this issue has focused on intonational boundaries, which are rhythmic junctures in speech that often correlate with syntactic boundaries. For example, some studies have found that the presence of syntactic ambiguity influences boundary placement, such that speakers place boundaries in

locations that will disambiguate the meaning of the sentence when they are aware of the ambiguity (e.g. Allbritton, Mckoon, & Ratcliff, 1996; Snedeker & Trueswell, 2003).

While some studies have focused on contextual effects on prosodic boundaries, acoustic prominence may be particularly sensitive to differences between communicative contexts. In English, prosodic cues often convey pragmatic and discourse information, the importance of which might vary across contexts. Indeed, studies have found that speakers signal information status differently when addressing infants and foreigners (Biersack, Kempe, & Knapp, 2005; Fernald & Mazzei, 1991). It is possible that speakers also change how they produce prosodic prominence based on communicative context even if the listener is an adult who shares their language. Evidence from computational linguistics provides motivated reasons to think that this is the case. Words that carry more information, and are consequently lower in predictability, are more likely to be prominent than less informative, more predictable words (Aylett & Turk, 2004).

This finding fits within a larger body of work that has found that speakers lengthen utterances by increasing both the duration and number of words at points of high information load in order to produce a uniform density of information over time for listeners, thereby facilitating communication (Levy & Jaeger, 2007; Jaeger, 2010). Moreover, these information theoretic effects have been found in the context of single conversations. If speakers can make subtle adjustments in the way prosodic prominence is implemented as a function of information load within a discourse, they may be able to do so across communicative contexts as well. In contexts in which a premium is placed on successful communication, speakers may attempt to make prosodic categories more distinct in order to facilitate comprehension. As such, a goal of this study is to investigate the acoustic dimensions along which prosodically-relevant cues to information status vary as a function of context.

Furthermore, understanding whether (and how) communicative contexts affect the way in which speakers produce acoustic prominence may allow us to better understand *what* the cues to acoustic prominence are. Researchers generally agree that some combination of F0 differences, duration, and intensity contribute to the perception of prominence (Fry, 1955; Lieberman, 1960; Beckman, 1986; Gussenhoven et al., 1997, Kochanski et al., 2005; Cole, Mo, & Hasegawa-Johnson, 2010; Lam & Watson, 2010; Breen et al., 2010 to name just a few studies). However, studies vary in which of these factors is found to be most important, and more importantly, there are few explanations for the discrepant findings across laboratories.

If one assumes that the presence and realization of acoustic prominence is wholly determined by discourse structure, then the communicative context in which new or focused words are elicited matters very little, and we might conclude that speakers simply fail to communicate prosodic prominence reliably. However, if prosodic prominence is sensitive to the goals and communicative context of the conversation, decisions about the types of tasks participants engage in become more important: different tasks may yield differences in the likelihood of detecting cues that correlate with prominence. This is important, as listeners need to integrate an array of prosodic cues in order to build informative prosodic representations. It is possible that some communicative contexts drive speakers to convey prosodic information by producing more discriminable cues, which would be more likely to help a listener identify the intended category. These differences in cue reliability across contexts may be most apparent when we consider the set of cues in aggregate, rather than simply looking at individual cues. Thus, an additional goal of this study is to determine the overall reliability of acoustic cues to prominence in more communicative contexts, relative to less communicative ones.

Information status in discourse

The strategy used in this study was to create contexts in which a speaker must convey referential information to a listener. We used two tasks: one in which speakers were more likely to be communicative and one in which they were less likely to be communicative.

Critically, across the tasks, target words and visual stimuli were held constant and differed only in the communicative context of the task. Participants in both tasks read aloud pairs of color sequences. The target word – the second word in the second sequence – was either new to the discourse, given, or contrasted with a color in the previous sequence:

(1a): *New* sequences: red blue green | gray **pink** black

(1b): *Given* sequences: gray pink black | gray **pink** black

(1c): *Contrastive* sequences: gray blue black | gray **pink** black

This allowed us to examine differences in the acoustic characteristics of words when they are focused (*contrastive* and *new*) and when they are not (*given*). We measured the acoustic prominence of the target word as determined by its duration, F0, and intensity, all of which are argued to correlate with prosodic emphasis (see Wagner & Watson, 2010 for a review).

The critical manipulation was whether the color sequences occurred in the context of a task in which speakers were more vs. less motivated to communicate effectively. In Experiment 1, the less communicative task was a simple color description paradigm, typical of a standard laboratory task, that the participant completed in isolation. Two sequences of colors appeared on a display, and the participants' task was to read the color sequences aloud. In the more communicative task, two participants worked together to navigate avatars through a series of puzzles in the computer game Minecraft. The puzzles were designed such that one participant had information that they needed to convey to their partner to solve the puzzle. This included the color description task: one participant was given a sequence of colors (the same ones used in the less communicative task) and the other had to enter that sequence as a "code" to unlock a door,

allowing them to proceed to the next room in the game. Thus, the game creates an immersive, highly engaging environment that allows us to study language use in a rich communicative context. Simultaneously, it provides precise control over the stimuli, allowing us to elicit production of specific words in different discourse contexts.

It is important to note that the communication manipulation is actually a manipulation of an array of different factors: the presence of an interlocutor, the presence of engaging filler tasks, whether communication plays a role in meeting goals within the task, and the level of entertainment of the participants. The advantage of this strategy is that it allows us to test, at a very broad level, whether speakers' prosodic choices are sensitive to communicative context. However, because a number of factors can contribute to the communicative context, a disadvantage is that, if differences between conditions occur, it will be unclear which aspect of the manipulation is driving them. We address this by first testing for overall effects of communicative context on prosodic cues using the two tasks described above (Experiment 1), and then examine one factor that likely contributes to context effects, specifically, the presence or absence of an interlocutor in the less communicative task (Experiment 2).

Across the two experiments, we can view the different contexts as existing along a continuum: (1) the rich communicative context of the computer game, (2) the less communicative context with an active listener present, and (3) the less communicative context with no listener present. Thus, the predictions are as follows. If prosodic prominence is context invariant, there will not be a difference in the cues to prosodic categories between tasks. However, if prosodic prominence is modulated by the communicative context, we expect that there will be acoustic differences between the tasks, such that speakers provide more informative cues in the more communicative tasks (e.g., larger duration differences between focused and given words for the more communicative task than for the less communicative task). Lastly, if

these effects are observed and they are driven by the presence of an interlocutor, we expect the presence or absence of a listener alone will drive the effect. These hypotheses were tested across a set of two experiments.

Experiment 1

The first experiment was designed to determine whether communicative context, broadly, has an effect on acoustic cues to discourse prominence by comparing speakers' productions in a high vs. low communicative context, holding stimulus and task procedures constant across the two contexts. In each context, we measured word duration, intensity, mean F0, and F0 range for words produced in a focused context (*contrastive* and *novel* conditions) and words produced in a non-focused context (*given* condition). We also examined how the overall statistical reliability of the cues (Toscano & McMurray, 2010) varied across the two communicative contexts.

Method

Design. Participants performed either a less communicative or more communicative task. Within each one, participants read two sequences of three colors on each trial. We manipulated the information status of the color sequences such that the second sequence was either identical to the first (*given*), a completely different set of colors (*novel*), or had the same first and third colors, but a different second color (*contrastive*). Both the *novel* and *contrastive* conditions constitute focus contexts for prosodic prominence. The target word on each trial (i.e., those on which we took acoustic measurements) was the second color of the second sequence.

Each information status condition was repeated six times per subject. There were six color sets and participants produced all three conditions for each set, allowing us to measure acoustic differences between tokens of the same word across conditions. This resulted in a total of 18 critical trials for each participant (3 information status conditions \times 6 color sets). For the more communicative task, 18 filler trials were also included.

Subjects were randomly assigned to one of six trial order lists. The same lists were used across both tasks. Three of these lists were generated by randomly ordering the trials with the following constraints: (1) each of the six sets of color sequences occurred before it was presented again, (2) the specific order of color sets across the list was not repeated, (3) specific color sets could not repeat within two trials of each other, and (4) specific information status conditions could not repeat within one trial of each other (e.g., a *given* trial cannot be followed by another *given* trial). The remaining three lists were generated by reversing the trial order of the first three lists. For the more communicative task, each critical trial was followed by a random filler trial.

Participants in both tasks completed the experiment in a single one-hour session.

Participants. Seventy-two monolingual native-English speakers from the University of Illinois at Urbana-Champaign participated. Twenty-four pairs participated in the less communicative task, and twenty-four pairs participated in the more communicative task. For the more communicative task, we only analyzed the productions of the participant who was providing the color information to their partner. All participants provided informed consent, and received class credit as compensation. All reported normal hearing and normal or corrected-to-normal vision.

Materials. Stimuli consisted of colored squares corresponding to one of eight monosyllabic colors in English: *black, blue, brown, green, grey, pink, red, and white*.

Procedure. For the less communicative task, participants completed the experiment individually. Participants were presented with sequences of three colored squares on a computer screen (Fig. 1) and were instructed to name the colors they saw in each sequence from left to right. After naming the colors, participants pressed a key and were then presented with a blank screen for one second, followed by the next sequence. Participants were shown a fixation cross between trials in order to differentiate new trials from new sequences within that trial.

In the more communicative task, two naïve participants were seated in front of computers in different rooms and wore Sennheiser PC-360 headsets allowing them to communicate with each other and with the experimenter. The task consisted of a series of puzzles created in the multi-player computer game Minecraft (Bergensten & Persson, 2013) and MinecraftEdu (Koivisto, Levin, & Postari, 2013). The puzzles were organized into rooms within the game. Participants needed to communicate with each other and work together to solve each

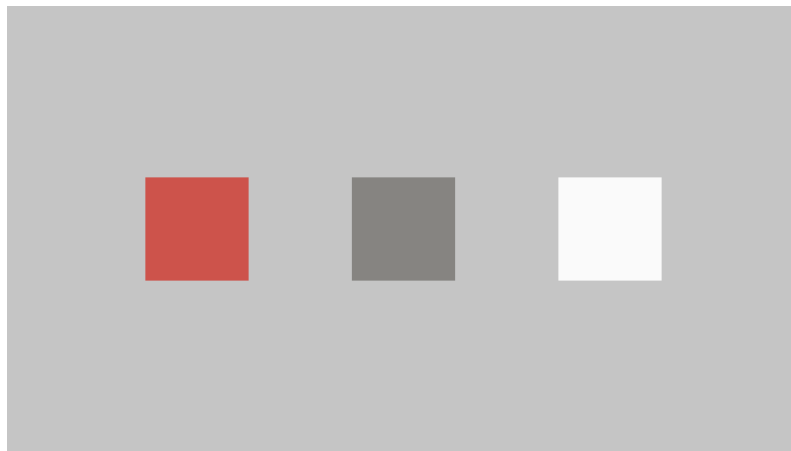


Fig. 1. Example of a critical trial for the less communicative task. The participant is presented with three colors, which she has to name out loud before proceeding to the next trial.

puzzle and proceed to the next room. At the beginning of the experiment, they were given a brief tutorial on how to operate the controls for the game. They were given enough time to practice until they felt ready to start the experiment. When both participants were ready, their characters in the game were moved to the room with the first puzzle. They were told that their goal was to work together to solve the puzzles in each room so that they could move on to the next. Filler trials included puzzles that were highly engaging and required interaction and reasoning to solve. When the participants solved a puzzle, a door in each subject's room opened and they could proceed to the next one. The characters in the game were separated by a wall during each critical trial, so that they could not see each other's rooms.

The critical trials consisted of “combination lock” puzzles: One participant (*Player 1*) saw a sequence of three colored squares on the wall and had to read that sequence to the other participant (*Player 2*) who was able to enter it as a code using buttons corresponding to each of

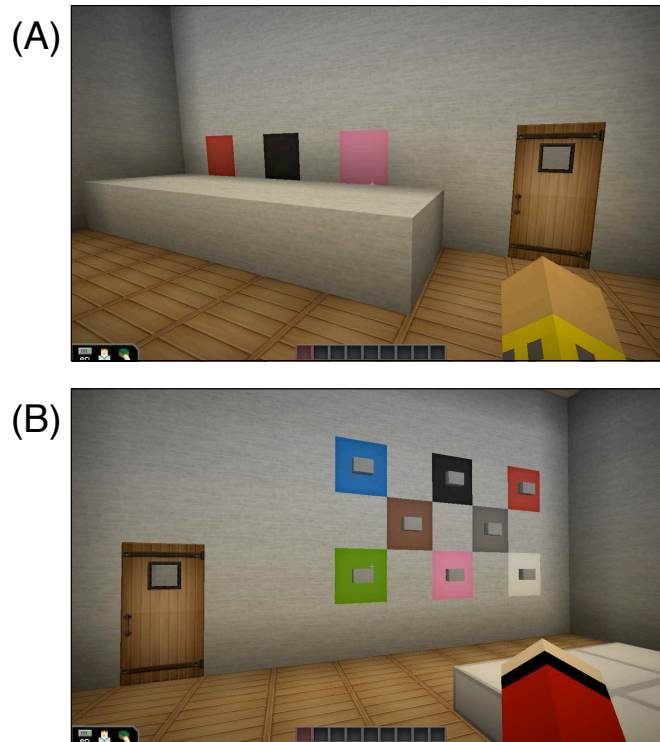


Fig. 2. Example of a critical trial in the computer-game task from the viewpoint of each participant. The participant in (A) is presented with the sequence, *red, black, pink*. They must give this information to their partner, the participant in (B), who must enter the sequence in as a code using the buttons on the wall.

the possible colors (Fig. 2). When the first sequence was entered correctly, the colors in *Player 1*'s room were replaced by a new sequence. Once the second sequence was entered, doors for both participants opened, and the participants continued to the next puzzle. Critically, the discourse structure of the target words in this task is identical to that of the targets in the less communicative task.

Data Analysis. For both tasks, participant's speech was digitally recorded at 44.1 kHz. Target words were manually transcribed using Praat (Boersma & Weenink, 2013) by marking

their onset and offset in a TextGrid. Word duration, mean intensity, mean F0, maximum F0, and minimum F0 values were then measured for each word. Word duration was log-transformed, and F0 values higher than 350 Hz were eliminated, as these were likely due to pitch doubling from speakers producing creaky voice.

There were a total of 864 critical trials across the two tasks. Seven participants were excluded from analysis because of either recording problems or because they could not appropriately solve the puzzles. Of these, two were from the less communicative task, and five were from the more communicative task. Thirty trials were discarded because they contained disfluent utterances or because the speaker did not say the words that the trial required. This left a total of 634 trials for analysis. Of these, 389 were from the less communicative task and 245 were from the more communicative task. There was a difference between these numbers because participants took much longer to finish the more communicative task and 11 participants were unable to complete all of the trials in the time allotted. On average, participants in the more communicative task completed 13.33 trials. Across both tasks, there were 204 given words, 216 new words, and 214 contrastive words.

Results

We analyzed four acoustic cues in the target words: (1) duration, (2) mean intensity, (3) mean F0, and (4) F0 range (i.e., maximum F0 – minimum F0). Figures 3-6 show the mean values for each of these cues across the three information status and two task conditions, indicating a number of differences between information status conditions and between the tasks. The data were analyzed in two ways. First, we used linear mixed effects models (LMEMs) to examine how cues differed as a function of task and information status, specifically comparing

the focus conditions (contrastive and new) with the non-focus condition (given).¹ Second, we asked how distinct the three information status conditions were for each task using the cue reliability metric from Toscano and McMurray (2010).

Mixed Effects Models. Table 1 summarizes the results of the LMEMs. In each analysis, trial number, information status (focus vs. non-focus), task (more- vs. less- communicative) and the information status \times task interaction were entered as fixed effects. Information status and task were effect coded (for information status, the two focus conditions, *contrastive* and *new*, were coded as +1, and *given* as -1; for task, low-engagement was coded as -1 and high-engagement as +1). Each fixed effect was then centered at zero. Subject was entered as a

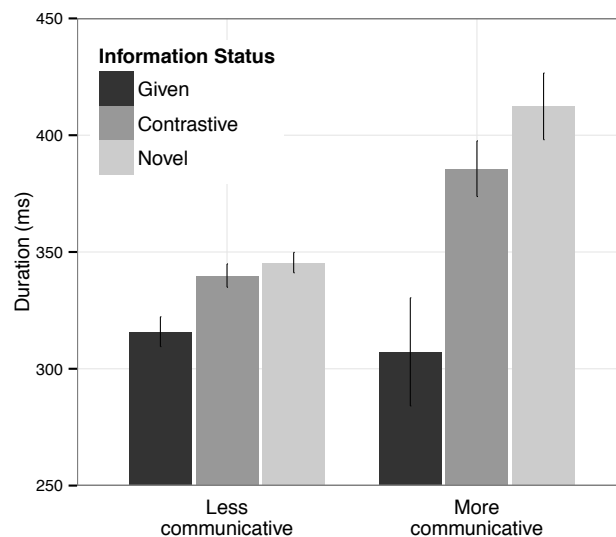


Fig. 3. Word duration as a function of information status and task. Overall, *contrastive* and *novel* words were longer than *given* words, and this difference was more pronounced for the more communicative task. Error bars indicate standard error.

random effect.² To determine the random effects structure, we used a backward-stepping model

¹ A visual inspection of the data revealed that the two conditions have similar cue values in both communicative contexts. For this reason, we collapse the two categories, and focus on the differences between the focus and non-focus conditions as a function of context.

² We also examined models with both by-subject and by-item (i.e., color word) random effects; these revealed the same pattern of results for the critical analyses (i.e., the interaction and main effect of information status within each task). Since there were only six different color words in the critical position in the lists, an item analysis likely does not have sufficient power to draw major conclusions. Thus, we present the by-subject models here.

comparison procedure to identify the most complex model justified by the data. Next, we used model comparison to test the significance of each fixed effect, comparing models in the following order: (1) a model with only random effects, (2) the previous model plus trial number, (3) the previous model plus information status, (4) the previous model plus task, and (5) the previous model plus the information status \times task interaction.

For duration, we found a main effect of information status, indicating that overall, target words were shorter in the *given* condition. More importantly, the interaction between task type and information status was significant, suggesting that the differences in durations between the information status conditions varied between the two tasks. There was a main effect of duration for both tasks: given targets had shorter durations than new and contrast targets, and these differences were larger for the more communicative task.

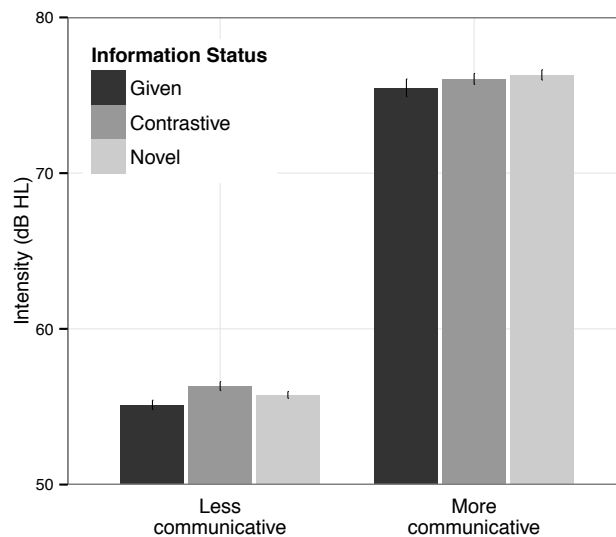


Fig. 4. Mean intensity as a function of information status and task. Overall, the high-communicative task had higher mean intensities (this could be due to microphone position for the recordings). There are also small differences between the information status conditions for the low-engagement task. Error bars indicate standard error.

A corresponding analysis was run for mean intensity. There was a main effect of task, with the more communicative task having higher mean intensities. There was a marginal main

effect of information status, with contrastive and new conditions having a higher intensity than given conditions. The interaction between information status and task was not significant. Planned comparisons revealed a main effect of information status within the less communicative task, but not the high communicative task. However, the differences between the focus and non-focus conditions were extremely small (0.92 dB), and thus, may not actually be perceptible.

For mean F0, there was a main effect of information status, with focus conditions having higher mean F0 values than given conditions. Other effects were non-significant. Planned comparisons showed a significant effect of information status for the more communicative task, but only a marginal effect for the less communicative task. The less communicative task *contrastive* words also had numerically higher mean F0 values than both *novel* and *given* words.

Finally, for F0 range, there was a main effect of information status: *focus* conditions had a larger F0 range than *given* conditions. This effect was driven by differences in the more communicative task, and there was a significant information status by task interaction. Follow-up tests showed a main effect of information status for the more communicative task, but no effect for the less communicative task.

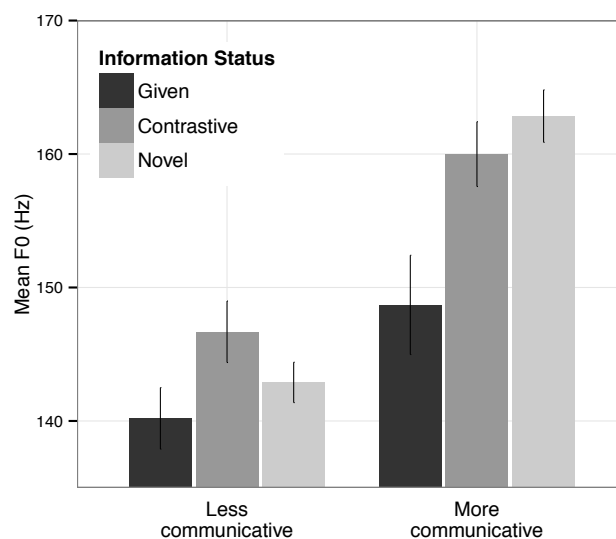


Fig. 5. Mean F0 as a function of information status and task. Overall, the high-engagement task had higher mean F0 values. In addition, mean F0 varied as a function of information status for that task, with *novel* words having the highest mean F0 values, and *given* words having the lowest. Error bars indicate standard error.

Thus, there were larger differences between focus and non-focus conditions for F0 range and duration in the more communicative context, and a significant difference for mean F0 in the more communicative context. This suggests that prominence is not context independent.

Cue Reliability Analysis. Next, we asked how distinct the three information status categories were for each task, given the set of four cues. To compute this, we used a simplified version of the cue reliability metric described in Toscano and McMurray (2010). This metric provides a measure of cue reliability that indicates how discriminable the categories along a given dimension are. Conceptually, this is similar to the d' statistic in signal detection theory and extends this idea to multimodal distributions (e.g., the distribution of *given*, *contrastive*, and *novel* categories along a cue dimension like *duration*). It is based on the Kalman filter approach for estimating cue reliability in a unimodal distribution (Jacobs, 1999; 2002):

$$d_i = \frac{1}{\sigma_i} \quad (1)$$

where d_i is the reliability of the cue and σ_i is the standard deviation of its estimate. Cues that provide highly variable estimates will have low reliabilities, while cues that have little variability in their estimates will have high reliabilities. The multi-modal cue reliability metric described by

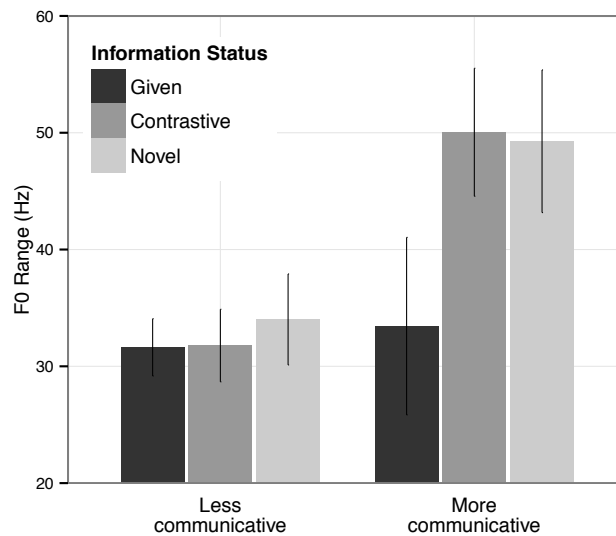


Fig. 6. F0 range as a function of information status and task. Speakers used a larger F0 range in the high-engagement task, and this varied as a function of information status, with *novel* and *contrastive* words having larger F0 ranges than *given* words. Error bars indicate standard error.

Toscano and McMurray (2010) extends this idea to allow for computing the reliability of acoustic cues in speech, where different modes correspond to different categories.

We used a simplified version of this measure to compute the reliability of each cue in the two tasks.³ For a given cue, the metric makes pairwise comparisons between each information status category according to:

$$m_i = \sum_a^K \sum_b^K \sqrt{\frac{(\mu_{bi} - \mu_{ai})^2}{\sigma_{ai}^2 \sigma_{bi}^2}} / 2 \quad (2)$$

where K is the total number of categories, m_i is the cue reliability, μ is the mean cue-value for a category, and σ is the standard deviation of cue-values for a category.

This yields a unitless measure of the overall reliability of the cue dimension (i.e., how easy it is to discriminate the categories along that dimension). If two categories are far apart along the cue dimension, the reliability of the cue will be higher than if they are close together. Similarly, if the variability within each category is high, the reliability will be lower. Thus, a cue dimension with two non-overlapping (i.e., distinct) categories will have a high reliability, and a cue with highly overlapping categories will have a low reliability.

The reliability for each cue in each task is shown in Table 2. Except for intensity (which provides very little information overall), cue reliability is higher (i.e. the categories are more distinct) for the more communicative task than for the less communicative task. The average reliability of the cues is approximately three times higher for the more communicative task.

To determine if the average cue reliabilities were different from chance, we ran Monte Carlo simulations. For each task, cue-values were randomly assigned to a condition (*given*,

³ The reliability metric given in Toscano and McMurray (2010) also includes terms for the likelihood of each category (to handle the fact that in their mixture model simulations, some categories had likelihoods near zero and, thus, should contribute little to the reliability estimates). Here, we simplify the equation by assuming that each category is equally likely and drop the likelihood terms.

contrastive, or *new*) and the average reliability of the set of cues was calculated. This process was repeated 10,000 times, producing a distribution of expected reliabilities. We then calculated p-values for the true cue reliabilities in each task from a normal distribution with the mean and standard deviation of the expected reliability distribution.

For the more communicative task, the mean reliability (0.42) was significantly greater than chance ($p < 0.001$; simulation mean: 0.19, simulation SD: 0.05), whereas the mean reliability for the less-communicative task (0.16) was not ($p = 0.415$; simulation mean: 0.15, simulation SD: 0.04). These results fit with the overall pattern of results seen with the LMEMs and suggest that speakers provide more reliable cues to information status in the more communicative task.

Discussion

One of the primary goals of Experiment 1 was to determine whether prosodic prominence was situationally dependent. We find evidence for this: speakers provided more informative cues to prosodic context, specifically via F0 and duration, in the more communicative task than in the less communicative task. Moreover, the reliability of the cues was higher for the more communicative task, suggesting that the categories are more distinct in this condition. Indeed, our simulations suggest that, overall, the set of cues is not informative at all in the *low-communicative* task.

While we find evidence that prosodic prominence is context dependent, the tasks in Experiment 1 differed in a variety of ways, making it difficult to identify the source of variability in the execution of prosodic prominence. The tasks differed in level of participant engagement, whether there was a conversational partner present, the amount of thinking required in each task, the amount of fun the participants were having, as well as other features. This was done purposely in order to create tasks that were maximally different in how motivated speakers were to communicate effectively. That said, it is likely that these features differentially contribute to

communicative motivation. We address this in Experiment 2.

Experiment 2

In Experiment 2, used two tasks that were as similar to each other as possible, while still manipulating communicativeness. Specifically, participants performed the less-communicative referential communication task from Experiment 1 either alone or with a listener as a partner. If the presence of an interlocutor contributes to effects of communicative context, we expect speakers to differentiate discourse categories prosodically to a greater extent when an interlocutor is present compared to when they are not.

Method

Design. As in Experiment 1, participants read two sequences of three colors on each trial. We manipulated information status in the same manner as in the previous experiment, such that the target word was either given, contrastive, or new. The stimuli were the same as those used for the less communicative task in Experiment 1, and the target word on each trial (i.e., those on which we took acoustic measurements) was the second color of the second sequence. Experimental lists were also the same as in Experiment 1. Participants in both tasks completed the experiment in a single one-hour session.

Participants. Seventy-two monolingual native-English speakers from the University of Illinois at Urbana-Champaign participated. Twenty-four pairs participated in the listener-absent task, and twenty-four pairs participated in the listener-present task. For the listener-present task, we only analyzed the productions of the participant who was providing the color information to their partner (as we did for the more communicative condition in Experiment 1). All participants provided informed consent, and received class credit as compensation. All reported normal hearing and normal or corrected-to-normal vision.

Materials. Stimuli consisted of the same colored squares used in the less communicative

context of Experiment 1, with colors corresponding to one of eight monosyllabic color words in English: *black, blue, brown, green, grey, pink, red, and white*.

Procedure. The listener-absent task in Experiment 2 was identical to the less-communicative task in Experiment 1. The listener-present task was a modified version of the less-communicative task from Experiment 1. Participants were seated at computers in different rooms and wore Sennheiser PC-360 headsets allowing them to communicate with each other. Speakers saw two sequences of three colors per trial, which they had to communicate to their partners. Listeners then had to input these color sequences into their own computer by clicking on the correct colors in order to advance to the next trial. The array of color response options was the same as those used in the more communicative context of Experiment 1. The listener-absent and listener-present tasks were identical except for the presence of a listener, so any differences in how speakers signal information status are due to the presence of an interlocutor.

Data Analysis. For both tasks, participant's speech was digitally recorded at 44.1 kHz. Target words were manually transcribed using Praat (Boersma & Weenink, 2013) by marking their onset and offset in a TextGrid. As in Experiment 1, word duration, mean intensity, mean F0, and F0 range values were then measured for each word.

There were a total of 864 critical trials across the two tasks. Eight participants were excluded from analysis because they did not follow instructions, consistently used different color names, or because they later revealed that they were not monolingual speakers of English. Of these, five were from the listener-absent task, and three were from the listener-present task. Eighteen trials were discarded because they contained disfluent utterances or because the speaker did not say the words that the trial required. This resulted in a total of 702 trials for analysis. Of these, 341 were from the listener-absent task and 361 were from the listener-present task. Across both tasks, there were 229 given words, 235 new words, and 238 contrastive words.

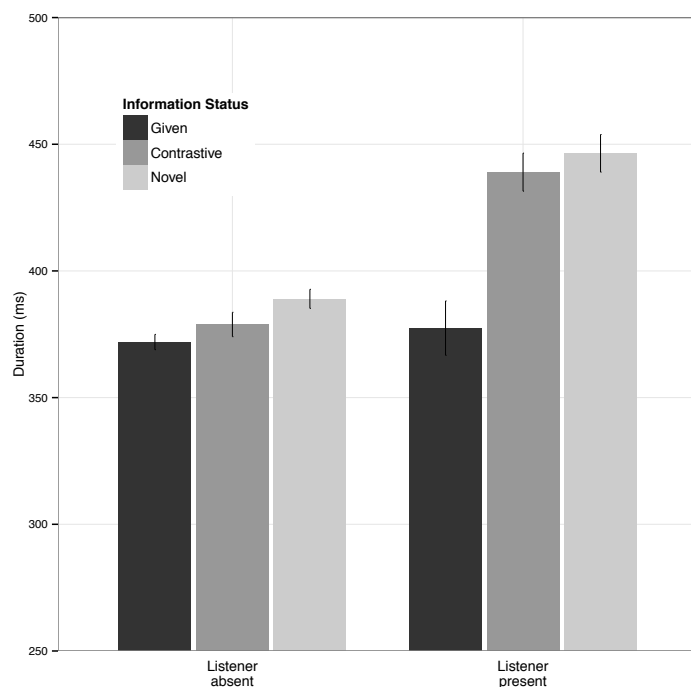


Fig. 7. Word duration as a function of information status and task. Overall, *contrastive* and *novel* words were longer than *given* words, and this difference was more pronounced when there was a listener present. Error bars indicate standard error.

Results

As in Experiment 1, we analyzed the target words' duration, mean intensity, mean F0, and F0 range. Figures 7-10 show the mean values for each of these cues across the three information status and two task conditions. The figures suggest that there are some differences between information status conditions and the tasks, though the differences do not appear to be as robust as those observed in Experiment 1. As before, we used LMEMs and the cue reliability metric to examine how cues differed as a function of task and information status, and to see how reliable the cue distributions were for each task.

Mixed Effects Models. Mixed effects models were built using the same steps described in Experiment 1. For duration, there was a main effect of information status, indicating that overall, target words were shorter in the *given* condition. More importantly, the interaction between task type and information status was significant, suggesting that the differences in durations between the information status conditions varied between the two tasks. Planned

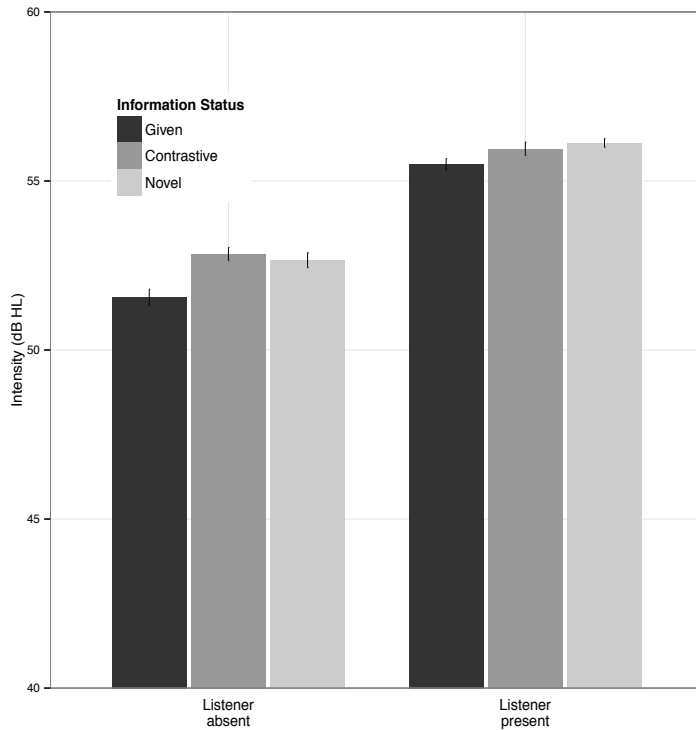


Fig. 8. Mean intensity as a function of information status and task. Overall, the listener-present task had higher mean intensities. There are again small differences between the information status conditions for the listener-absent task. Error bars indicate standard error.

comparisons reveal that there was a main effect of duration only for the listener-present task, suggesting that speakers lengthened focused words for the benefit of a listener. Although focused words were numerically longer in the listener-absent task, this difference was not significant. This is similar to the pattern observed in Experiment 1.

A corresponding analysis was run for mean intensity. There was a main effect of task, with the listener-present task having higher mean intensities. There was also a main effect of information status, with contrastive and new conditions having a higher intensity than given conditions. The interaction between information status and task was not significant. Planned comparisons revealed a main effect of information status for both tasks, with the given condition having a lower intensity. However, as in Experiment 1, the absolute value of these differences was quite small (1.11 dB) suggesting that there was little perceptual information for listeners to

gain from the intensity cue.

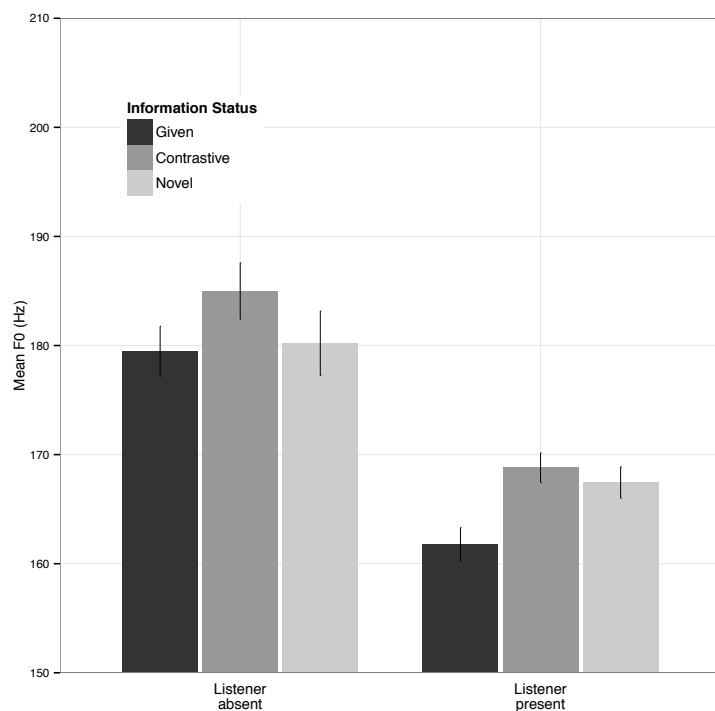


Fig. 9. Mean F0 as a function of information status and task. As opposed to Experiment 1's more-communicative task, the listener-present task had lower mean F0 values overall. In addition, mean F0 varied as a function of information status for that task, with *contrastive* words having the highest mean F0 values, and *given* words having the lowest. Error bars indicate standard error.

For mean F0, there was a main effect of information status, with focus conditions having higher mean F0 values than given conditions. Other effects were non-significant. Planned comparisons showed a significant effect of information status only for the listener-present task.

Finally, for F0 range, there was a main effect of task: the listener-absent condition had larger F0 ranges than listener-present condition. However, unlike in Experiment 1, there was no effect of information status or an interaction between information status and task. Follow-up tests showed no effect of information status for either task.

Thus, we see that there were larger differences between focus and non-focus conditions for duration and mean F0 in the listener-present condition (similar to Experiment 1), but that listener presence did not clearly modulate the use of intensity or F0 range. This suggests that the

presence of an interlocutor can drive some of the communicative context effects observed in

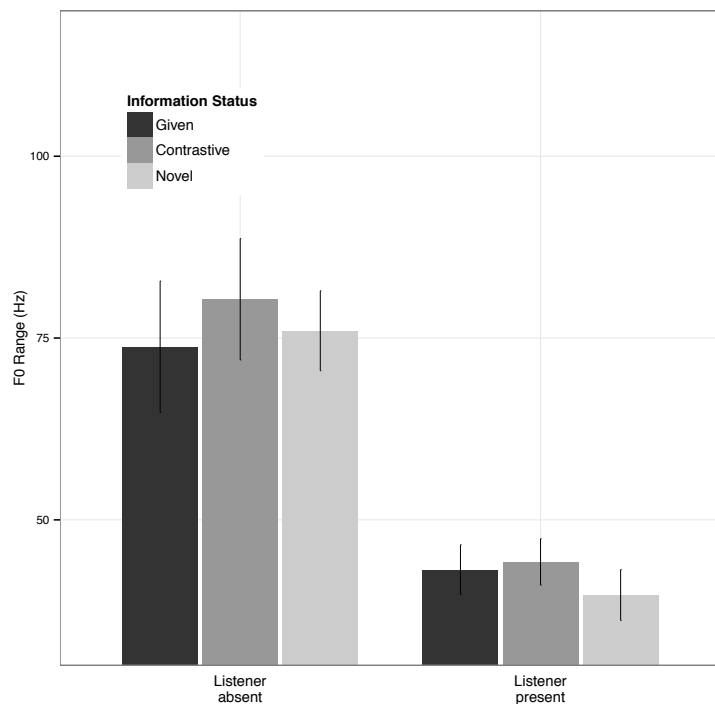


Fig. 10. F0 range as a function of information status and task. Speakers used less F0 range in the listener-present task. Unlike in Experiment 1, there is no longer an interaction between information status and task. Error bars indicate standard error.

Experiment 1, but it does not contribute to all of the effects observed; other factors, such as participant engagement, must also play a role.

Cue Reliability Analysis. Next, we asked how distinct the three information status categories were for each task, using the same metric we used for the cue reliability analysis in Experiment 1. The reliability for each cue, along with the average cue reliability, for each task is shown in Table 4. Overall, mean cue reliability for the listener-absent task was 0.14, the same as the low-engagement task from Experiment 1. This is expected, since these experimental conditions are identical. This value was not statistically different from chance ($p=0.723$; simulation mean: 0.16, simulation SD: 0.05), indicating that, overall, the cues did not clearly convey differences in information status in the listener-absent condition.

Mean cue reliability for the listener-present task was 0.26. This value was significantly different from chance ($p=0.011$; simulation mean: 0.16, simulation SD: 0.04) but is considerably lower than the cue reliability in the high-engagement task of Experiment 1 (0.42). This mirrors the LMEM results: listener presence contributes to speakers providing more reliable cues overall, although cue reliability is poorer than in the high communicative context of Experiment 1.

Discussion

To summarize, in Experiment 2, we explored how the presence of a listener affects how speakers signal information status. We find that word duration and mean F0 differences between focused and non-focused conditions are greater when there is a listener present compared to when there is not. Additionally, speakers produce more reliable cues overall when there is a listener present. However, unlike Experiment 1, we saw no effect of listener presence on how listeners use F0 range to signal information status, and the overall cue reliability was lower than in the more-communicative task used in Experiment 1. This suggests that some of the differences we saw in Experiment 1 were due to the fact that the more-communicative task included a listener, but listener-presence alone cannot account for all the differences between the more-communicative task and the less-communicative contexts. More generally, in both Experiments, the communicative contexts in which the utterances were produced modulated the reliability of the prosodic cues to discourse structure.

General Discussion and Conclusions

The main goal of this study was to investigate how different communicative contexts elicit different manifestations of prosodic prominence. In Experiment 1, we found evidence that speakers differentiate focused and given words to a greater extent when participating in more-communicative tasks. Specifically, speakers signaled focused words using longer word durations, higher mean F0 values, and larger F0 ranges. Additionally, cue reliability was higher

for the more-communicative task, demonstrating that the prosodic categories are more distinct in this condition. Indeed, our simulations suggest that, in aggregate, these cues are not informative at all in the less-communicative task.

In Experiment 2, we compared the prosody of two carefully matched communicative contexts that differed only in the presence of an interlocutor. We still find that word duration and mean F0 differences between focus and non-focus conditions are greater when the communicative stakes are higher (i.e. a listener present). Additionally, speakers produce more reliable cues overall when there is a listener present. However, unlike Experiment 1, we saw no effect of listener presence on how listeners use F0 range to signal information status, and the overall cue reliability was lower than the more-communicative context examined in Experiment 1. This demonstrates that listener presence is important for signaling acoustic prominence, but it is potentially only one of a variety of factors that contribute to the communicative context.

These results suggest that more communicative contexts elicit larger differences between prosodic categories, even when the task itself is not more difficult or engaging. It is also likely that the more robust differences seen in Experiment 1 were due to the task being even more communicative than the listener-present task used in Experiment 2.

Together, results from Experiments 1 and 2 suggest that speakers modulate how much prosodic information they convey based on the communicativeness of the present context. This can be seen across the continuum of contexts we examined. Talkers provided the most reliable information about prosodic prominence in the rich, immersive game-based task used in Experiment 1, less reliable information in the non-engaging task with a listener present, and the least reliable information in the non-engaging task with no listener present.

Although it might be tempting to attribute the effects seen in these experiments to non-communicative factors such as engagement or task difficulty, it is important to point out that

these accounts would not explain the effects seen in Experiment 2, where the two conditions did not differ meaningfully in engagement or in difficulty. Because of this, we conclude that communicative context is responsible for the effects described above.

Another potential concern is that the results presented here are simply the result of participants being more emotionally aroused in the more communicative tasks, and are not the result of speakers subtly manipulating the acoustic form of prosodic prominence to optimize communication. It is well known that high levels of arousal can lead to increased F0 excursions and more F0 variability (see Juslin & Laukka, 2003 and Scherer, 2003 for a review). In fact, Ladd (2008) draws a distinction between linguistic prosody, which correlates with linguistic structure, and paralinguistic prosody, which correlates with emotional arousal. Moreover, an explanation based on communication and one based on arousal are not mutually exclusive. It is possible that arousal may be a mechanism by which speakers make prosodic cues more distinct in contexts where communication is particularly important. A speaker who is angry, or very excited, may heighten prosodic distinctions because these are contexts in which linguistic communication is most important.

However, these effects would not produce the pattern of results observed in the current study. That is, there is no reason to think that increased emotional arousal would lead to greater differences as a function of information status. An arousal-based explanation cannot explain why the information status categories were more acoustically distinct in the more communicative tasks. Previous studies have found a wider F0 range and greater pitch excursions, *overall*, in emotional speech. This predicts a main effect of task such that the more arousing task should elicit greater F0, intensity, and duration across all information status categories. Indeed, we see this in the current data, in the main effect of task on intensity. Critically, however, there was also an interaction between task and information status for duration and F0 mean (for both

experiments) and F0 range (in Experiment 1) such that difference between information status categories are greater in the more communicative contexts than in the less communicative ones. This is not predicted by an arousal explanation. Additionally, while the listener-present task in Experiment 2 could have lead to higher levels of emotional arousal, the results were not an attenuated version of the results from Experiment 1, but rather a categorically different pattern, where duration and mean F0 differences resembled the more communicative task but F0 range differences resembled the less communicative task. Consequently, we think it is unlikely that the effects here can be explained by paralinguistic prosody or emotional arousal.

The use of communicative tasks, like the Minecraft task used in this study, may allow us to ask questions that lie at the heart of recent debates in the literature about whether speaker preferences or listener preferences drive the distribution of linguistic regularities in language (MacDonald, 1999; MacDonald, 2013; Tanenhaus, 2013). MacDonald (2013) argues that the process of language production is more computationally expensive than the process of language comprehension. Consequently, speakers' linguistic choices are constrained primarily by ease of production, rather than an optimization of the linguistic signal for listeners' comprehension. However, in the current experiment, we see that in contexts in which communication is more important, speakers make distinctions between prosodic categories more clear. Tanenhaus (2013) and Jaeger (2013) argue that optimizing information for a listener may not be as computationally complex as intuitions might suggest.

Indeed, in Experiment 2, we find evidence suggesting that speakers are modulating prosodic prominence for the benefit of the listener via word durations. Thus, it is likely that there are both, production-centric and listener-centric sources to prosodic variability, and it is possible that these factors are reflected in different cues, as we see in our studies. Tanenhaus (2013) points out that experimental production tasks typically do not include interactive

conversations, rich context, complex goal structures, and continual feedback from listeners, and that all of these factors may help in mitigating the complexity of optimizing utterances for listeners in real conversation. Game-based platforms potentially provide the psycholinguist with the tools to design complex, context-rich, interactive experiments that have the capacity to answer the types of questions raised above.

To conclude, speakers modulate prosodic prominence in fine-grained ways to improve the discriminability of prosodic categories. In particular, speakers improve discriminability more often in contexts in which a premium is placed on communication. The overall communicative context in which a conversation occurs can have consequences not only for whether prominence occurs, but also for how discriminable the cues to prominence are. The factors driving these effects include, but are not limited to, the presence of an interlocutor, suggesting that studying discourse processing must entail an understanding of the rich communicative contexts that characterize real language use. The studies presented take important steps towards deciphering how prosodic prominence varies among different communicative contexts.

Acknowledgements

We would like to thank Dominique Simmons, Luis Paneque, Samantha Jensen, and Kelsey Mills for assistance with data collection and coding, and Cheyenne Munson Toscano for help creating the Minecraft map. ABL was supported by NIH Grant T32-HD055272. JCT was supported by a Postdoctoral Fellowship from the Beckman Institute. DGW is supported by R01 DC008774 and the James S. McDonnell foundation.

References

- Allbritton, D.W., McKoon, G., & Ratcliff, R. (1996). The reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 714-735.
- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47, 31-56.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42 (1), 1-22.
- Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Dordrecht, Netherlands: Foris Publications.
- Bergensten, J. & Persson, M. (2013). Minecraft [Computer program]. Stockholm, Sweden: Mojang.
- Biersack, S., Kempe, V., & Knapton, L. (2005). Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. In: *Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisbon, 2401-2404.
- Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.49, retrieved 13 May 2013 from <http://www.praat.org/>
- Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25, 1044-1098.
- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, 61, 171-190.

- Brown-Schmidt, S. (2005). *Language Processing in Conversation*. (Doctoral dissertation), University of Rochester.
- Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, 23, 567-598.
- Cole, J., Mo, Y., Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1, 425-452.
- Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., & Lotts, D. W. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, 29, 233-251.
- Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, 27, 209-221.
- Fowler, C. A., & 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, 489-504.
- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765-768.
- Gussenhoven, C., Repp, B.H., Rietveld, A., Rump, W.H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, 102, 3009-3022.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. Part 2. *Journal of Linguistics*, 3, 199-244.
- Jacobs, R.A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, 6, 345-350.
- Jacobs, R.A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, 39, 3621-3629.

- Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, 4, 230. doi: 10.3389/fpsyg.2013.00230
- Jaeger, T. F. (2010). Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology*, 61(1), 23-62.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotion in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770-814.
- Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118, 1038-1054.
- Koivisto, S., Levin, J., & Postari, A. (2013). MinecraftEdu [computer program]. Joensuu, Finland: Teacher Gaming.
- Ladd, D. R. (2008). *Intonational phonology (2nd edn)*. Cambridge, UK and New York, NY: Cambridge University Press.
- Lam, T. Q. & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, 38, 1137-1146.
- Levy, R. & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Scholkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems (NIPS)*, 19, 849-856. Cambridge, MA: MIT Press.
- Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 32(4), 451-454.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 1-16.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney. (Ed.), *The Emergence of*

- Language* (pp. 177-196). Mahwah, NJ: Erlbaum.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227-256.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211-232.
- Snedeker, J. & Trueswell, J. (2003). Using Prosody to Avoid Ambiguity: Effects of Speaker Awareness and Referential Context. *Journal of Memory and Language*, 48, 103-130.
- Tanenhaus, M. K. (2013). All P's or mixed vegetables? *Frontiers in Psychology*, 4, 234.
doi:10.3389/fpsyg.2013.00234
- Toscano, J. C. & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34, 434-464.
- Wagner, M. & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25, 905-945.
- Yoon, S. O., Koh, S., & Brown-Schmidt. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, 19, 699-707.

Table 1. Summary of LMEM results (Experiment 1).^a

	Log-duration	Intensity	Mean F0	F0 range	
Trial Number	b=-0.003, SE=0.002 $\chi^2(1)=3.19$, p=0.074	b=0.001, SE=0.034, $\chi^2(1)=0.25$, p=0.619	b=0.316, SE=0.194, $\chi^2(1)=2.33$, p=0.127	b=-0.052, SE=0.28, $\chi^2(1)=0.073$, p=0.787	
Information status	b=0.085, SE=0.013, $\chi^2(1)=21.55$, p<0.001	b=0.34, SE=0.13, $\chi^2(1)=3.87$, p=0.049	b=3.47, SE=1.05, $\chi^2(1)=10.57$, p=0.001	b=3.52, SE=1.84, $\chi^2(1)=3.78$, p=0.052	
Task	b=0.035, SE=0.028, $\chi^2(1)=3.50$, p=0.061	b=10.1, SE=0.95, $\chi^2(1)=53.28$, p<0.001	b=7.50, SE=6.20, $\chi^2(1)=1.45$, p=0.229	b=5.75, SE=3.37, $\chi^2(1)=0.57$, p=0.449	
Information status × Task	b=0.052, SE=0.013, $\chi^2(1)=11.89$, p<0.001	b=-0.18, SE=0.13, $\chi^2(1)=1.67$, p=0.197	b=1.41, SE=1.09, $\chi^2(1)=1.67$, p=0.196	b=3.89, SE=1.89, $\chi^2(1)=4.14$, p=0.042	
Task = <i>more communicative</i>	Trial number	b=0.003, SE=0.004, $\chi^2(1)=1.15$, p=0.284	b=-0.043, SE=0.043, $\chi^2(1)=0.857$, p=0.355	b=0.085, SE=0.337, $\chi^2(1)=0.031$, p=0.861	b=0.36, SE=0.56, $\chi^2(1)=0.309$, p=0.579
	Information status	b=0.147, SE=0.030, $\chi^2(1)=11.59$, p<0.001	b=0.12, SE=0.22, $\chi^2(1)=0.29$, p=0.588	b=5.21, SE=1.70, $\chi^2(1)=9.18$, p=0.002	b=7.82, SE=2.85, $\chi^2(1)=7.39$, p=0.007
Task = <i>less communicative</i>	Trial number	b=-0.006, SE=0.002, $\chi^2(1)=6.92$, p=0.009	b=0.030, SE=0.045, $\chi^2(1)=0.0004$, p=0.983	b=0.42, SE=0.24, $\chi^2(1)=3.02$, p=0.082	b=-0.28, SE=0.33, $\chi^2(1)=0.714$, p=0.398
	Information status	b=0.043, SE=0.012, $\chi^2(1)=10.22$, p=0.001	b=0.488, SE=0.157, $\chi^2(1)=6.04$, p=0.014	b=2.41, SE=1.33, $\chi^2(1)=3.29$, p=0.070	b=0.61, SE=1.84, $\chi^2(1)=0.11$, p=0.739

^a Model coefficients and standard errors are from models that include all the terms for that analysis (i.e., the models including the interaction for the omnibus analysis and the models containing the information status term for the follow-up analyses).

Table 2. Cue reliability results (Experiment 1).

Task	Log-duration	Intensity	Mean F0	F0 Range	Average	Different from chance?
<i>High-engagement</i>	0.82	0.12	0.33	0.43	0.42	Yes ($p < 0.001$)
<i>Low-engagement</i>	0.34	0.13	0.11	0.06	0.16	No ($p = 0.415$)

Table 3. Summary of LMEM results (Experiment 2).

	Log-duration	Intensity	Mean F0	F0 range	
Trial Number	b=-0.005, SE=0.002 $\chi^2(1)=7.763$, p=0.005	b=0.033, SE=0.019, $\chi^2(1)=2.875$, p=0.090	b=0.066, SE=0.157, $\chi^2(1)=0.183$, p=0.669	b=-0.354, SE=0.466, $\chi^2(1)=0.596$, p=0.440	
Information status	b=0.055, SE=0.009, $\chi^2(1)=35.988$, p<0.001	b=0.451, SE=0.106, $\chi^2(1)=18.173$, p<0.001	b=2.378, SE=0.870, $\chi^2(1)=7.585$, p=0.006	b=0.223, SE=2.579, $\chi^2(1)=0.008$, p=0.930	
Task	b=0.052, SE=0.038, $\chi^2(1)=1.997$, p=0.158	b=1.735, SE=0.726, $\chi^2(1)=5.603$, p=0.017	b=-7.862, SE=6.918, $\chi^2(1)=1.332$, p=0.248	b=-16.970, SE=4.352, $\chi^2(1)=13.546$, p<0.001	
Information status \times Task	b=0.035, SE=0.009, $\chi^2(1)=15.998$, p<0.001	b=-0.143, SE=0.106, $\chi^2(1)=1.840$, p=0.175	b=0.875, SE=0.871, $\chi^2(1)=1.013$, p=0.314	b=-1.392, SE=2.582, $\chi^2(1)=0.293$, p=0.588	
Task = <i>listener-present</i>	Trial number	b=-0.002, SE=0.002, $\chi^2(1)=0.734$, p=0.392	b=0.025, SE=0.027, $\chi^2(1)=0.908$, p=0.341	b=0.107, SE=0.186, $\chi^2(1)=0.390$, p=0.532	b=-0.715, SE=0.379, $\chi^2(1)=3.575$, p=0.059
	Information status	b=0.088, SE=0.018, $\chi^2(1)=15.995$, p<0.001	b=0.301, SE=0.153, $\chi^2(1)=3.759$, p=0.053	b=3.113, SE=1.069, $\chi^2(1)=7.618$, p=0.006	b=-0.719, SE=2.316, $\chi^2(1)=0.099$, p=0.753
Task = <i>listener-absent</i>	Trial number	b=-0.007, SE=0.002, $\chi^2(1)=12.946$, p<0.001	b=0.042, SE=0.028, $\chi^2(1)=2.222$, p=0.136	b=0.032, SE=0.256, $\chi^2(1)=0.011$, p=0.916	b=0.046, SE=0.871, $\chi^2(1)=0.002$, p=0.964
	Information status	b=0.018, SE=0.011, $\chi^2(1)=2.737$, p=0.098	b=0.599, SE=0.152, $\chi^2(1)=12.157$, p<0.001	b=1.563, SE=1.533, $\chi^2(1)=1.048$, p=0.306	b=1.767, SE=5.690, $\chi^2(1)=0.099$, p=0.753

Table 4. Cue reliability results (Experiment 2).

Task	Log-duration	Intensity	Mean F0	F0 Range	Average	Different from chance?
<i>Listener-present</i>	0.60	0.22	0.08	0.13	0.26	Yes (p=0.011)
<i>Listener-absent</i>	0.19	0.19	0.11	0.06	0.14	No (p=0.723)

