COMMUNICATIVE CONTEXT, EXPECTATIONS, AND ADAPTATION IN PROSODIC PRODUCTION AND COMPREHENSION

BY

ANDRÉS BUXÓ-LUGO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Duane G. Watson
Professor Gary Dell
Professor Cindy Fisher
Professor Kara Federmeier
Assistant Professor Chigusa Kurumada, University of Rochester

ABSTRACT

It is generally assumed that prosodic cues that provide linguistic information are driven primarily by the linguistic content of an utterance. However, research has shown that information from different levels of language often interact and affect the production and comprehension of an utterance (e.g., Brown-Schmidt, 2009; Remez, 1981; Ganong, 1980). If prosody operates in a similar manner to other levels of language, speakers and listeners should be sensitive to things such as communicative context, sentence structure, and listener expectations. This paper explores this possibility through a variety of studies.

Part 1 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.

Part 2 investigates whether listener expectations influence the processing of intonational boundaries. In a boundary detection task, we manipulated a) the strength of cues to the presence of a boundary and b) whether or not a location in the sentence was a plausible location for an

intonational boundary to occur given the syntactic structure. In Experiment 1, listeners are instructed to report where they heard disfluencies, even though there were no disfluencies present in the recordings. Listeners report hearing disfluencies at the locations that had intonational boundaries, and they are equally likely to report them in licensed and unlicensed locations for boundaries. This suggests that listeners can interpret prosodic cues to boundaries as disfluencies, and that their expectations as to where disfluencies might occur are more flexible than they are for where boundaries might occur.

Experiment 2 makes use of a 2 trial version of the boundary detection paradigm to investigate whether listeners who reported hearing boundaries at unlicensed locations in previous studies (Buxó-Lugo & Watson, 2016) had done so because of adaptation to new types of input. The results replicate previous studies, and find no evidence for adaptation having occurred throughout those experiments. Lastly, we propose 2 studies to investigate whether listeners can adapt to new mappings between prosodic and syntactic structure.

ACKNOWLEDGEMENTS

I still can't believe I've been able to reach this point in my academic career. I'm sure it wouldn't have been possible without the enormous support I've gotten from my family, friends, and colleagues. I'll do my best to summarize their contributions here.

First, I'd like to thank my family. My parents Luis and Aida for always believing in me and for working so hard to provide me with the amazing opportunities I've had. My brother Alejandro for somehow keeping me sane and driving me insane at the same time, and for being one of the best friends I could ever have. Titi Tere, for sending me oversized sweaters and coffee from back home, without which I wouldn't have been able to survive grad school in Illinois.

I am thankful for all the professors who have in one way or another served as my mentors at Illinois. You've been incredibly kind towards me, and I feel very lucky to have met you. No doubt my work has greatly benefited from the questions, feedback, and comments I've gotten from all of you throughout the years. As I move forward, I know I will continue to ask myself what you would have thought about a certain issue, and I know my work will be better for it ("What would Kara think? What would Cindy say? What would Sarah ask? Aaron? John?" Etc.). I'd like to thank Gary for providing so much inspiration and support, and for being a great co-advisor, even though I wasn't technically his student. I especially want to thank Duane for being such a great advisor. He has been infinitely patient with me, and his advising has not only made me a better scientist, but also a better person.

Lastly, I'd like to thank my friends, who have always been there for me. I was fortunate to have started grad school alongside an amazing cohort of people. Each one of you has made my grad school experience so much better. Thanks to Nate for showing me how to see the world

through a different set of eyes. To Ariel, who got to know me and inexplicably still chose to be my close friend. To Cassie, with whom casual lunches would turn into inspiring conversations about research. And to Brian and Aldi, for the adventures I shan't mention here. I'd also like to thank all the students and postdocs that I got to befriend throughout my years in grad school (Rachel, Alison, Scott, Laurel, Maureen, Joe, Alex, Pam, John, etc.). You've been great friends, and have taught me so much. Grad school would have been worth it just for the chance of meeting all of you. Lastly, thank you to my old friends Eric, Jimmy, Guy, and Renee, who I've known for years and have been amazing cheerleaders throughout. Here's to many more years.

To everyone I've mentioned in this section: thank you. I hope I get to see y'all very soon.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION……1

CHAPTER 2: PART 1 – INTRODUCTION……...4

CHAPTER 3: PART 1 – EXPERIMENT 1……...10

CHAPTER 4: PART 1 – EXPERIMENT 2……..22

CHAPTER 5: PART 1 – GENERAL DISCUSSION AND CONCLUSIONS……...30

CHAPTER 6: PART 2 – INTRODUCTION…….34

CHAPTER 7: PART 2 – EXPERIMENT 1……...38

CHAPTER 8: PART 2 – EXPERIMENT 2……...48

CHAPTER 9: PART 2 – EXPERIMENT 3……...52

CHAPTER 10: PART 2 – GENERAL DISCUSSION AND CONCLUSIONS……..…66

CHAPTER 11: CONCLUSIONS……...…….……..68

REFERENCES…….70

APPENDIX A: PART 1 TABLES……..79

APPENDIX B: PART 2 TABLES……..82

CHAPTER 1: GENERAL INTRODUCTION

Prosody, the rhythm, intensity, and intonation of speech, is an integral part of spoken communication. Speakers use prosody to convey a wealth of information, and listeners often take advantage of this. Prosody might reveal, for example, the emotions or beliefs a speaker currently holds. If someone says "I won" with a falling intonation, listeners will interpret the message as a declaration. On the other hand, if someone says "I won?" with a rising intonation at the end, listeners will interpret the message as a question, perhaps also conveying the speaker's belief about their own chance of winning. Prosody can also inform listeners about the metalinguistic content in a message. For example, speakers might emphasize words that are important, unexpected, or new to the conversation at hand, and this might help listeners better understand or remember crucial parts of the message. Additionally, prosody interacts with many other levels of language, such as syntax, semantics, and pragmatics. As such, adequate prosody helps listeners decipher important parts of the linguistic structure of a message.

While the importance of prosody to spoken communication is acknowledged, our understanding of what mechanisms govern prosody production and comprehension is still limited. Part of this might be because prosody involves different kinds of cues (e.g. stress, word duration, intonation, intensity) conveying information about different levels of language, potentially at the same time. Despite this, prosody is often studied in highly specific scenarios with the intention to isolate a single relationship between prosodic cues and other levels of language. Perhaps because of this, researchers have proposed accounts of prosody that are mostly bottom-up, unidirectional, and isolated from other levels of language processing in both production (e.g., Schafer et al., 2000) and comprehension (e.g., Schafer, 1997).

While this approach of separating prosodic phenomena into separate, independent representations has yielded some valuable knowledge, researchers must also study how wider

contexts can affect prosodic production and comprehension. Because prosody is used to inform listeners about various levels of language, including speakers' belief states and communicative context, these aspects of language could interact and affect how humans produce and understand prosodic information. From a production standpoint, it would be beneficial to investigate what types of factors (including extra-linguistic factors) affect speakers' manifestation of prosody, and how these factors interact with each other. Similarly, for us to better understand how listeners process and comprehend prosodic information, it is necessary to investigate what types of information listeners consider when making inferences based on prosody, and how listeners weigh and consolidate these cues. Research from other areas of language processing has found that top-down effects, parallel processing, and contextual effects are common in language (e.g., Schober & Clark, 1989; Brown-Schmidt, 2005; Remez, 1981; Kim & Osterhout, 2005; Gibson et al., 2013). If the mechanisms governing over prosody are similar to those at other areas of language, other levels of language processing, as well as communicative context and listener expectations should also affect prosody. Understanding how these factors affect prosodic production and comprehension is crucial to developing models of prosody and understanding how prosodic processes fit with other levels of language production and comprehension. Additionally, if the same mechanisms that underlie other levels of language production and comprehension are responsible for prosody, it would allow researchers to investigate these processes in a special scenario, where the system has to cope with high degrees of interactivity, multiple sources for variability, and noisy, more confusable cues than in other levels of language.

This paper attempts to explore the effects of context and listener expectations on the production and comprehension of prosody. Part 1 focuses on the effect of communicative context on prosody production. One important debate in the field of prosody research is over what cues are used to mark prominence. We posit that a possible reason for why different

studies find different cues to be relevant is because they are not accounting for the communicative context. We support this idea with data from 2 experimental studies that manipulated the communicative context in which participants spoke to investigate whether this affected their prosodic manifestations of prominence.

In Part 2, we investigate how listener expectations might affect prosodic processing. Given that prosody interacts with many other levels of language processing, listeners should develop expectations as to what prosodic manifestations accompany other linguistic phenomena. However, prosodic processing is often thought about as a strictly bottom-up process, with no influence from other levels of language contributing to the construction of prosodic structure. In Part 2, we focus on the connection between prosodic boundaries and syntactic structure. Through a series of experiments, we explore whether listener expectations based on syntactic structure affect their interpretation of prosodic structure, whether listeners are able to interpret the same cues as something other than prosodic phenomena (i.e., disfluencies), and whether listeners can use feedback to develop new mappings between prosodic and syntactic structures.

CHAPTER 2: PART 1 - INTRODUCTION

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). For example, if a speaker says "Pass me the KEYS… the keys on the table," The first instance of the word "keys" tends to sound more prominent or accented than the second instance of the word "keys." 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, & Knapton, 2005; Fernald & Mazzie, 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, Kochanksi 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.

2.1 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.

CHAPTER 3: PART 1 - 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.

3.1 METHOD

3.1.1 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 × 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.

3.1.2 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.

3.1.3 Materials

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

3.1.4 Procedure

For the less communicative task, participants completed the experiment individually. Participants were presented with sequences of three colored squares on a computer screen (Figure 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

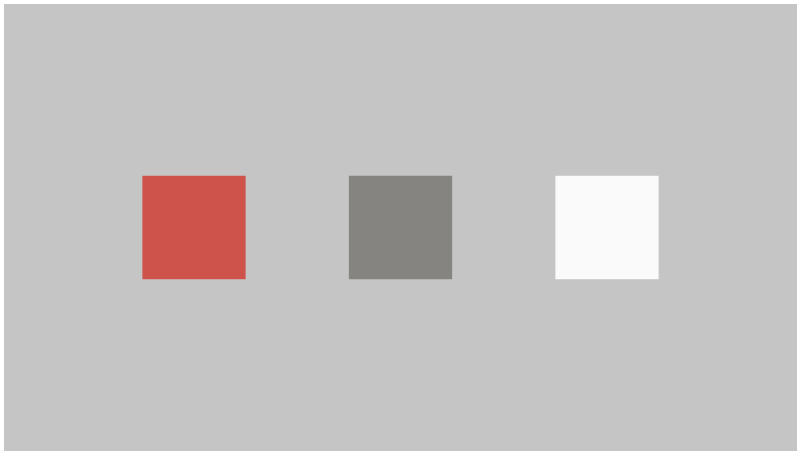

Figure 1. Example of a critical trial for the less communicative task. The participant is presented with three colors, which she has to name out loud before proceeding to the next trial.

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

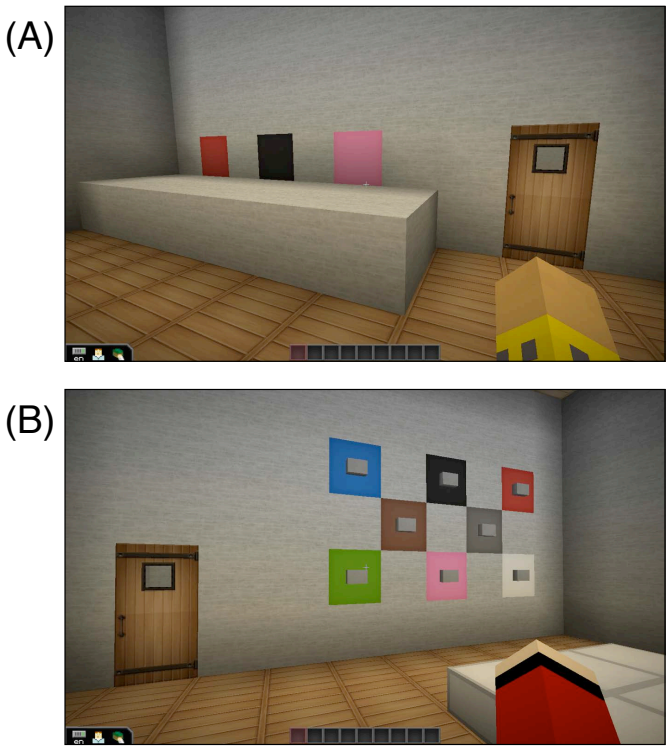

Figure 2. Example of a critical trial in the computer-game task from the viewpoint of each participant. The participant in **(A)** is presented with the sequence, *red, black, pink*. They must give this information to their partner, the participant in **(B)**, who must enter the sequence in as a code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

the possible colors (Figure 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.

3.1.5 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.

3.2 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).

3.2.1 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 × 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

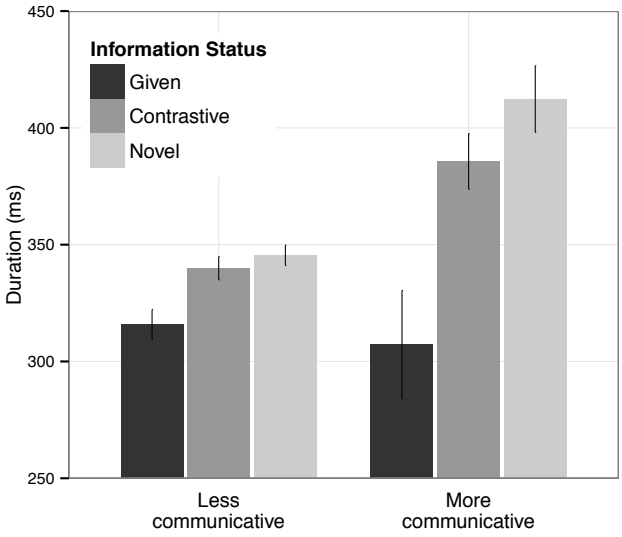

Figure 3. Word duration as a function of information status and task. Overall, *contrastive* and *novel* words were longer than *given* words, and this difference was more pronounced for the more communicative task. Error bars indicate standard error.

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 random effect.[2] To

[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.

determine the random effects structure, we used a backward-stepping model 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 × 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.

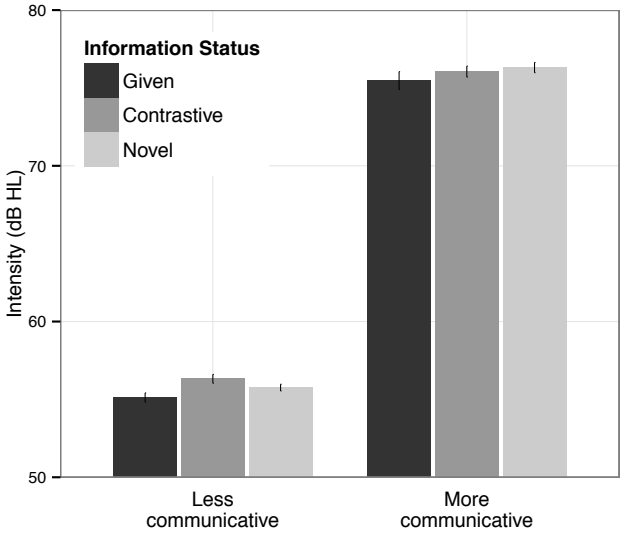

Figure 4. Mean intensity as a function of information status and task. Overall, the high-communicative task had higher mean intensities (this could be due to microphone position for the recordings). There are also small differences between the information status conditions for the low-engagement task.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

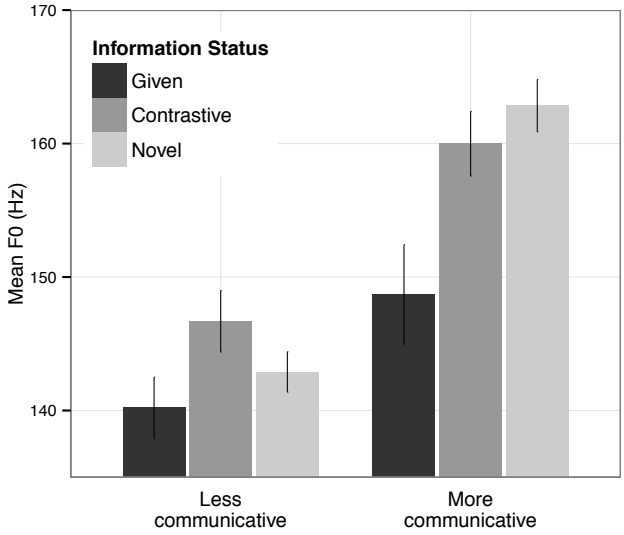

Figure 5. Mean F0 as a function of information status and task. Overall, the high-engagement task had higher mean F0 values. In addition, mean F0 varied as a function of information status for that task, with *novel* words having the highest mean F0 values, and *given* words having the lowest. Error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

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.

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.

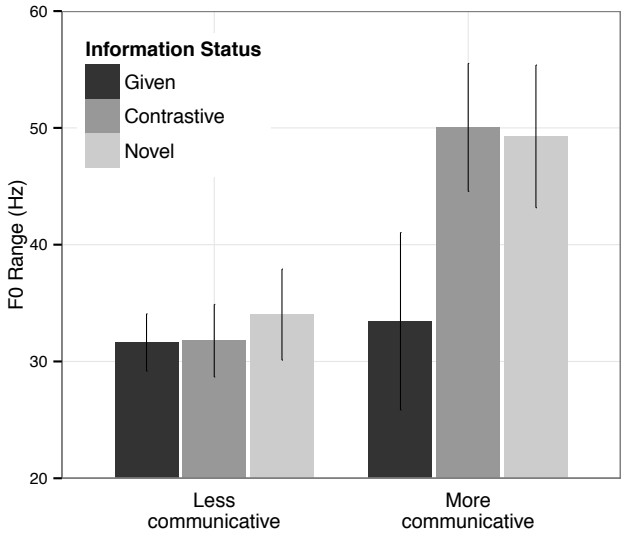

Figure 6. F0 range as a function of information status and task. Speakers used a larger F0 range in the high-engagement task, and this varied as a function of information status, with *novel* and *contrastive* words having larger F0 ranges than *given* words. Error bars indicate standard error.

3.2.2 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):

$$d_i = \frac{1}{\sigma_i}$$

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 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:

$$m_i = \sum_{a}^{K} \sum_{b}^{K} \sqrt{\frac{(\mu_{bi} - \mu_{ai})^2}{\sigma_{ai}^2 \, \sigma_{bi}^2}} /2$$

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.

[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.

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, 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.

3.3 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.

CHAPTER 4: PART 1 - 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.

4.1 METHOD

4.1.1 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.

4.1.2 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.

4.1.3 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*.

4.1.4 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.

4.1.5 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.

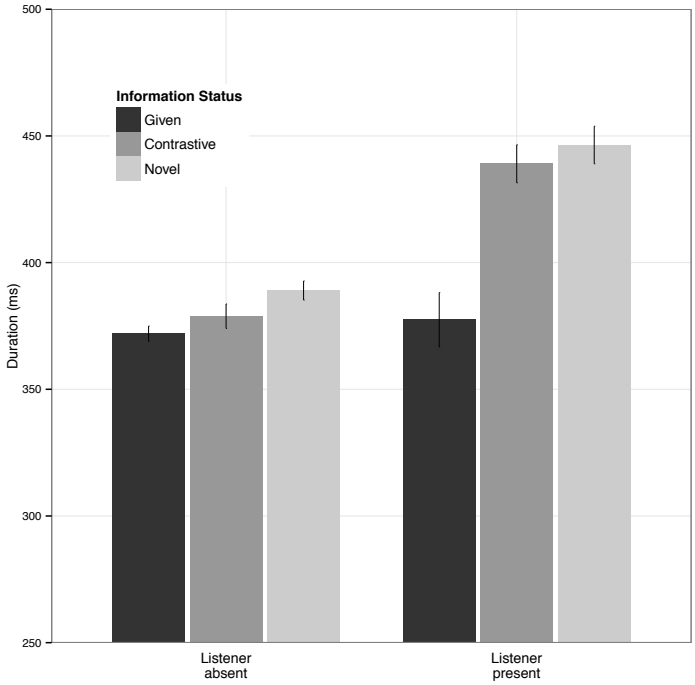

Figure 7. Word duration as a function of information status and task. Overall, *contrastive* and *novel* words were longer than *given* words, and this difference was more pronounced when there was a listener present. Error bars indicate standard error.

4.2 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 to 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.

4.2.1 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 comparisons reveal that

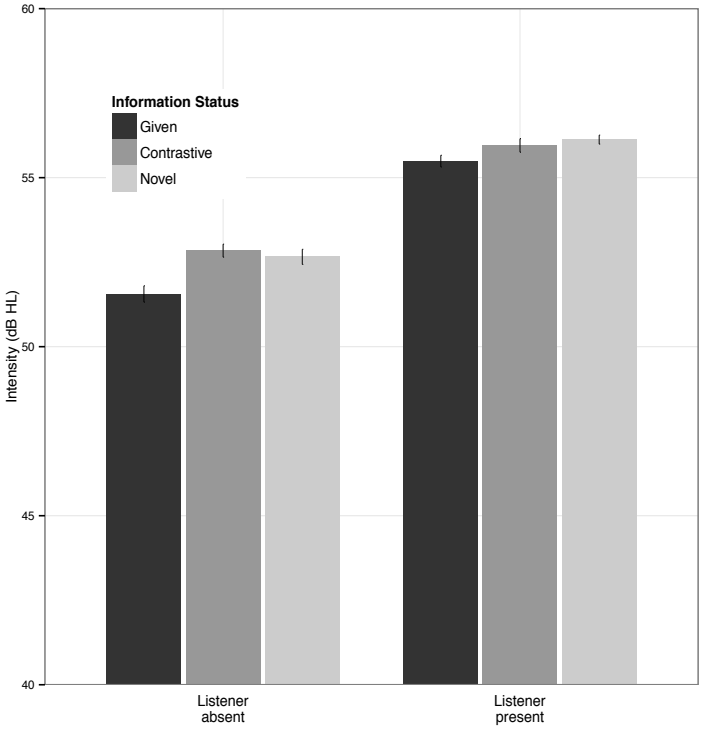

Figure 8. Mean intensity as a function of information status and task. Overall, the listener-present task had higher mean intensities. There are again small differences between the information status conditions for the listener-absent task. Error bars indicate standard error.

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,

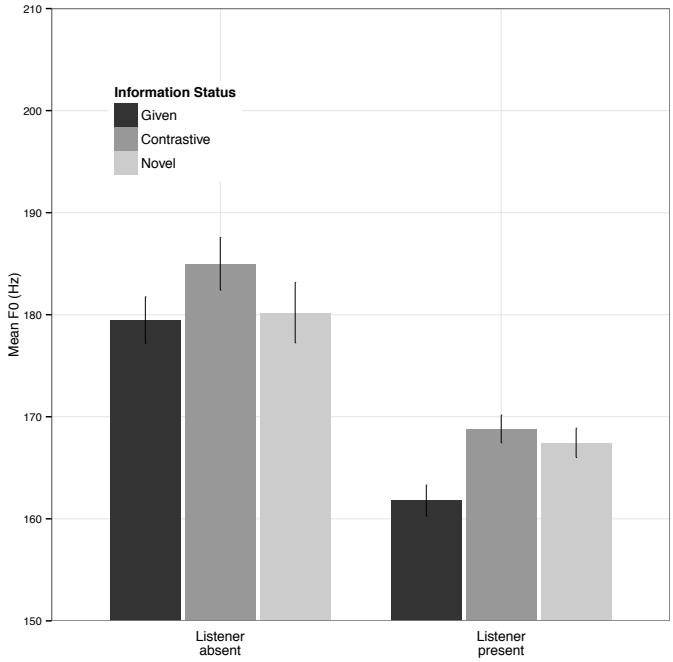

Figure 9. Mean F0 as a function of information status and task. As opposed to Experiment 1's more-communicative task, the listener-present task had lower mean F0 values overall. In addition, mean F0 varied as a function of information status for that task, with *contrastive* words having the highest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

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.

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

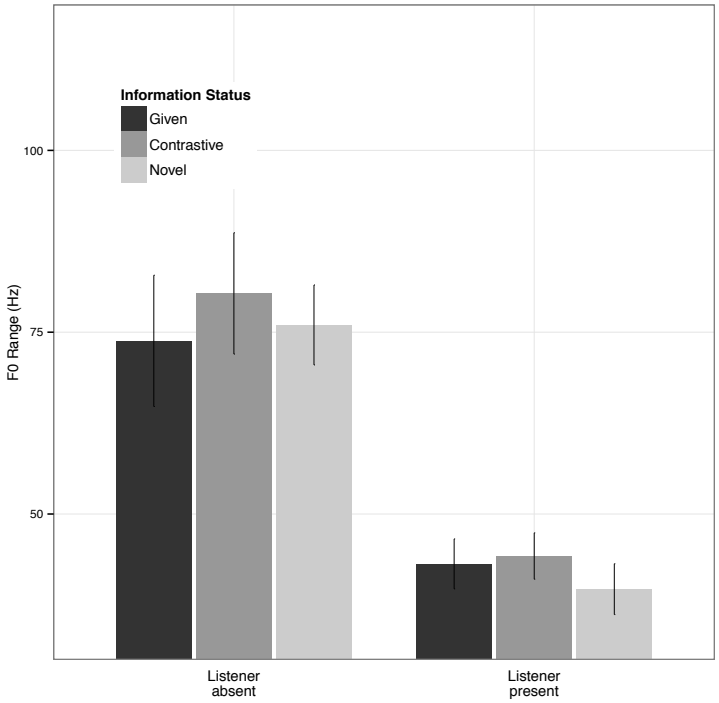

Figure 10. F0 range as a function of information status and task. Speakers used less F0 range in the listener-present task. Unlike in Experiment 1, there is no longer an interaction between information status and task. Error bars indicate standard error.

4.2.2 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.

4.3 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.

CHAPTER 5: PART 1 - 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.

CHAPTER 6: PART 2 - INTRODUCTION

An important part of parsing prosodic structure is detecting intonational boundaries, which are used to group utterances into smaller constituents that sometimes reflect the syntactic structure of spoken sentences (Cooper & Paccia-Cooper, 1980; Ferreira, 1993; Watson & Gibson, 2004). These boundaries are signaled by pauses, changes in F0 contours, and pre-boundary lengthening, among other cues (e.g., Klatt, 1975; Pierrehumbert & Hirchberg, 1990; Turk & Shattuck-Hufnagel, 2007; Ladd, 2008). Listeners, in turn, can use intonational boundaries to decipher the linguistic structure of a message, as in the case of syntactically ambiguous sentences (Schafer, Speer, & Warren, 2005; Snedeker & Trueswell, 2003).

However, few studies have explored how listeners build their representation of utterances' prosodic structure. Current models that aim to shed light on the relationship between prosody and other levels of representation tend to be unidirectional, often focusing on how prosody can guide the interpretation of other constructs such as syntax (e.g., Price et al., 1991; Kjeelgard & Speer, 1999; Schafer et al., 2000). For example, Schafer (1997) proposes the following relationship between prosody and syntax: "the prosodic representation that is constructed by the phonological component is passed on to higher-level modules in the same way that lexical information is made available to them" (p. 6) such that prosodic information is "part of the computational vocabulary of the syntactic and semantic/pragmatic processing modules" (p. 6). According to such models of prosodic parsing, listeners build prosodic representations from the acoustic cues, and then use these constructs to guide their interpretation of higher-level structures. This suggests a fully bottom-up relationship between prosody and syntactic structure. However, it is possible that prosodic parsing is more interactive, or bi-directional. In such a model, information from higher-level structures and listener expectations,

along with acoustic cues, guide the parsing of prosodic structure. This paper investigates how listeners' expectations interact with prosodic processing.

Interaction between listener expectations and processing systems is ubiquitous in language processing. For example, perception studies have found that syntax influences where listeners report hearing bursts of noise (Garrett, Bever, & Fodor, 1966), that morphological context affects the perception of ambiguous phonemes (Ganong, 1980), and that top-down knowledge of the speech signal affects whether degraded speech is perceived as speech at all (Remez et al., 1981). As such, a unidirectional bottom-up relationship between prosody and other levels of language processing would make prosodic processing stand out as inherently different from other kinds of language processing. However, it would be surprising if listener expectations did not influence their interpretation of prosody. Indeed, some studies have found that prosodic information from earlier in an utterance influences how listeners segment words (e.g., Brown et al., 2011; Dilley et al., 2010) and how they interpret lexical stress (Brown et al., 2012) later in an utterance. Also work by Bishop (2012) suggests that expectations about discourse structure can influence the perception of acoustic prominence. This is further supported by work by Cole, Mo, and Baek (2010), where untrained listeners prosodically transcribed speech from the Buckeye corpus. In their study, both vowel duration and syntactic context were correlated with boundary reports, each factor independent of the other. In fact, syntactic context was the best predictor of boundary detection, suggesting that listeners' judgments were influenced by their expectations of where boundaries should occur.

More recently, Buxó-Lugo & Watson (2016) investigated the effects of syntactic structure on prosodic parsing in an experimental task. In Buxó-Lugo & Watson (2016), words were acoustically manipulated along a boundary spectrum that ranged between sounding like

there was no boundary after the word, to strongly suggesting boundary presence after the word. Words that were manipulated to sound like they were followed by a boundary had longer durations, longer following pauses, and a descending F0 contour, while words that were manipulated to sound like they were not followed by a boundary had shorter durations and following pauses, and a non-descending F0 contour. Critically, these manipulated words were either located in a syntactically licensed location where one might expect a boundary to occur, or a syntactically unlicensed location where boundaries rarely occur and therefore should not be expected. For example, in the sentence "Put the big bowl on the tray," the word "bowl" was manipulated to sound more boundary-like in the syntactically licensed condition, since this is a likely location for a speaker to place a boundary. In the syntactically unlicensed location, the word "big" was made to sound more boundary-like. If listeners' perceptions of boundaries were determined by the acoustic cues alone, their responses should not be affected by the location in which the manipulated word was. However, listeners reported more boundaries overall at the syntactically licensed location, even when the word was not manipulated to sound like it was at a boundary at all. This further suggests that listeners make use of their knowledge and expectations when constructing a message's prosodic structure.

While recent evidence suggests that listener expectations affect the interpretation of prosody, there are still open questions as to what kinds of expectations affect prosodic parsing, as well as how flexible listeners' prosodic representations might be. The following studies attempt to shed light on some of these questions. In Experiment 1, we investigate whether asking listeners to report disfluencies changes how listeners interpret the acoustic signals from Buxó-Lugo & Watson's (2016) recordings. In Experiment 2 we investigate whether the syntactic effects from Buxó-Lugo & Watson (2016) were the result of listeners building expectations

about boundary locations across the course of the experiment. Experiment 3 includes multiple studies that investigated whether exposure to speakers that use prosody in unexpected ways could change how listeners interpret an utterance. Experiment 3a was designed to make sure different boundary locations reported in different interpretations in relative clause (RC) attachment sentences that would be ambiguous without the prosodic information. Experiment 3b investigated whether exposing listeners to irregular prosody usage affects how they interpret otherwise ambiguous RC attachment sentences. Experiment 3c investigated whether listeners can generalize what they have learned about one prosody-syntax mapping to new syntactic structures.

CHAPTER 7: PART 2 - EXPERIMENT 1

Buxó-Lugo & Watson (2016) concluded that listeners reported hearing boundaries in syntactically licensed locations because they used their knowledge from different levels of language processing in order to reach the most likely interpretation for an utterance. However, it is unknown how flexible listeners can be when accommodating an acoustic signal to their expectations and beliefs. One possibility is that listeners treat cues associated with one linguistic phenomenon as evidence for the presence of different linguistic phenomenon that is more expected or likely given the context. In fact, previous studies have found evidence that listeners perception of phonological categories is influenced by top-down knowledge such as lexical and semantic information upstream in an utterance (Connine, Blasko, & Hall, 1991) and external explanations for phonetic realizations (Kraljic, Samuel, & Brennan, 2008). Additionally, Clifton, Frazier, & Carlson (2006) found that listeners have intuitions about why speakers place intonational boundaries in certain locations, and as such they do not interpret these boundaries as conveying syntactic information if there are alternate reasons for the speaker to have produced a boundary (e.g., constituency length). These lines of research suggest that listeners will have intuitions as to where boundaries should and should not occur, and because of this, they may interpret boundaries at unexpected locations as cues indicating a different linguistic construct altogether.

One possible way to investigate how listeners consolidate acoustic information with their expectations and intuitions is to use the stimuli from Buxó-Lugo & Watson (2016) to see whether listeners sometimes interpret intonational boundaries in unexpected locations as something altogether different, like disfluencies. While the acoustic correlates of disfluencies are different from those of intonational boundaries, there are some cues (e.g., lengthening of

words, insertion of pauses, changes in F0) they share in common (Shriberg, 2001). Additionally, disfluencies are much more flexible as to where they appear within an utterance. Fraundorf & Watson (2014) finds that speakers are likely to produce silent pause disfluencies when they face difficulty planning the upcoming parts of the utterance. This provides an alternate explanation with softer constraints than syntactic location for listeners attempting to interpret the acoustic phenomena they are exposed to. As such, listeners may interpret intonational boundaries in syntactically unlicensed locations as disfluencies, even though they do not sound exactly like disfluencies. In other words, it might be a smaller accommodation to interpret the acoustic cues as belonging to a disfluency in a location where a disfluency might be natural, than to interpret them as a real intonational boundary produced at a highly unlikely location.

If it is the case that listeners are flexible enough to interpret boundary cues as disfluencies, listeners' responses may also shed light on what their expectations are as to where disfluencies should occur. For example, while there are specific locations where intonational boundaries are most likely to occur, speakers may face production difficulty at any point in an utterance. Listeners' expectations should reflect this by having more flexible expectations as to where disfluencies occur than where intonational boundaries occur. If this is the case, we should expect listeners to mark disfluencies at similar rates at either manipulated location, since disfluencies should occur with similar likelihoods at each of these locations.

An additional advantage to this experiment is that we can be more confident that any effects from Buxó-Lugo & Watson (2016) are not simply due to artifacts in the recorded stimuli. Since listeners are now reporting disfluencies, we should not expect them to report more disfluencies in what was the syntactically licensed location for boundaries. As such, Experiment 1 had three goals: first, to investigate whether listeners' interpretations of utterances were

flexible enough to represent intonational boundaries as disfluencies. Second, to probe listeners' expectations about where disfluencies should be more likely to occur. Lastly, Experiment 1 should provide evidence that listeners are not simply reporting more linguistic phenomena in the same location, but rather that their reports will vary based on their interpretations.

7.1 METHODS

7.1.1 Participants

Twenty English speakers from the United States of America participated in the study. Two participants were excluded due to having learned a language other than English from an early age (before 5). This resulted in 18 monolingual English speakers. They were all users of Amazon's Mechanical Turk service, and they all had at least a 95% approval rating for previous task completions. They were paid $6.00 for participating in the study.

7.1.2 Materials

Recordings used in Buxó-Lugo & Watson (2016) were reused for Experiment 1. To create these materials, a native English speaker was recorded while producing variants of 14 critical items. Each item was a unique noun-modifier pair (e.g., "green frog," "big bowl," etc.). For each of these item pairs, 2 different sentence structures were produced. One structure included a direct object with a prenominal modifier. In the other structure, the direct object was modified by a relative clause that included the same adjectival modifier. For example:

a.) Put the big bowl on the tray.

b.) Put the bowl that's big on the tray.

The purpose of the two structures was to balance the part of speech that preceded the preferred locations for boundaries. In a), a boundary is syntactically licensed after "bowl" (a noun), while in b) a boundary is syntactically licensed after "big" (an adjective). These locations were chosen

because previous work suggests that major syntactic boundaries, such as the boundary between an object phrase and a prepositional phrase, are likelier places for intonational boundaries than non-major syntactic boundaries (e.g., between a noun and a modifier: Gee & Grosjean, 1983; Watson & Gibson, 2004).

Each of these sentences was produced once with a boundary at a syntactically licensed location, and once with a boundary at a syntactically unlicensed location, as in the following:

c.) Put the big bowl | on the tray.

d.) Put the bowl that's big | on the tray.

e.) Put the big | bowl on the tray.

f.) Put the bowl | that's big on the tray.

Examples (c) and (d) have boundaries at syntactically licensed locations while examples (e) and (f) are produced with boundaries at syntactically unlicensed locations. There were 14 items, 2 sentence structures, and 2 boundary locations, resulting in 56 different recordings.

A boundary continuum was constructed by first transcribing the key nouns and modifiers ("big" and "bowl" in the previous example) in all of the items using Praat's textgrid feature (Boersma & Weenink, 2015). The duration of each word, along with the pause that followed it, were measured. In order to measure the F0 contour, the average F0 was sampled from 10 equally-spaced regions throughout the word. The measurements from the naturally produced boundary words and naturally produced non-boundary words were then used as ends of a boundary spectrum. Seven equally-spaced boundary-steps in between these 2 end points were also derived, resulting in 9 steps of boundary-likeness. The boundary steps for F0 contour were created by first smoothing the contours of the end points into the cubic functions that best fit them. The difference between the boundary and non-boundary words at each of the 10 equally

spaced points throughout the word was divided by the number of steps, which resulted in an interval by which we could change the curve at each point for each step (illustrated in Figure 13).

Two key words in each of the original 14 recordings were resynthesized so that one of the words, what we will call the target word, was the primary point of acoustic manipulation and where we varied the boundary spectrum between 1 (non-boundary) and 9 (boundary). The other word, the non-target word, always had acoustic cues that were consistent with the absence of a boundary. The non-target word was re-synthesized to balance the effects of re-synthesis on boundary detection at the target. The target word and non-target word were counter-balanced so that half the time the target word was at the syntactically licensed location and half the time the non-target word was at the syntactically unlicensed location. In order to make the recording as natural as possible, the F0 contour was resynthesized so that the initial point of the contour was matched to the F0 of the corresponding point in the original non-resynthesized word. This prevented sudden changes in F0 as the word started. The rest of the F0 contour values were derived by fitting the appropriate curve to the starting point (the beginning of the curve corresponding to the onset of the word) and calculating the values at 9 other equally-spaced points. The F0 contour was resynthesized based on these values at 10 equally-spaced points throughout the word using Praat's Manipulate function, which is based on the PSOLA algorithm. Words and pauses were lengthened (or shortened) to match the durations given by the desired boundary step. This was done using Praat's Lengthen function, which also makes use of the PSOLA algorithm. In order to control for the effects of the words surrounding the target words, we resynthesized sentences that originally had the boundary in the syntactically licensed location as well as sentences that originally had a boundary in the syntactically unlicensed location. The four most natural sounding items after resynthesis were selected for the experiment. This

resulted in a total of 272 recordings (4 items * 2 boundary locations * 9 boundary steps * 2 sentence structures * 2 source sentences – 16, since at boundary-step 1 there is no difference between boundary position).

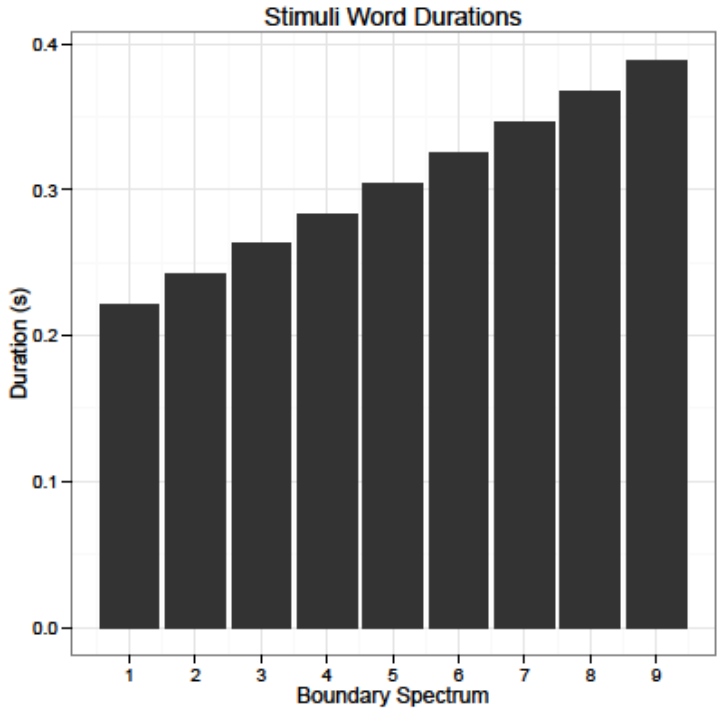

Figure 11: Word durations for critical words at each step of the boundary spectrum.

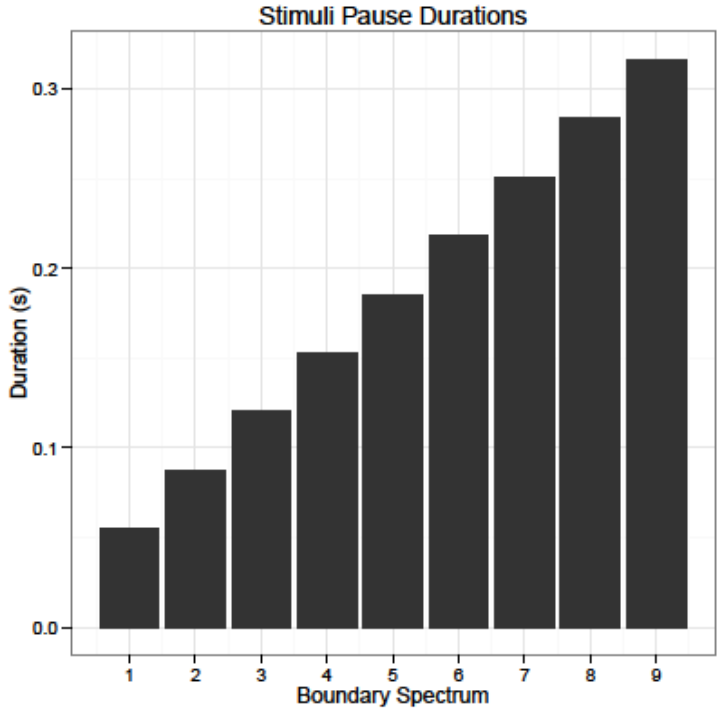

Figure 12: Following pause durations for critical words at each step of the boundary spectrum.

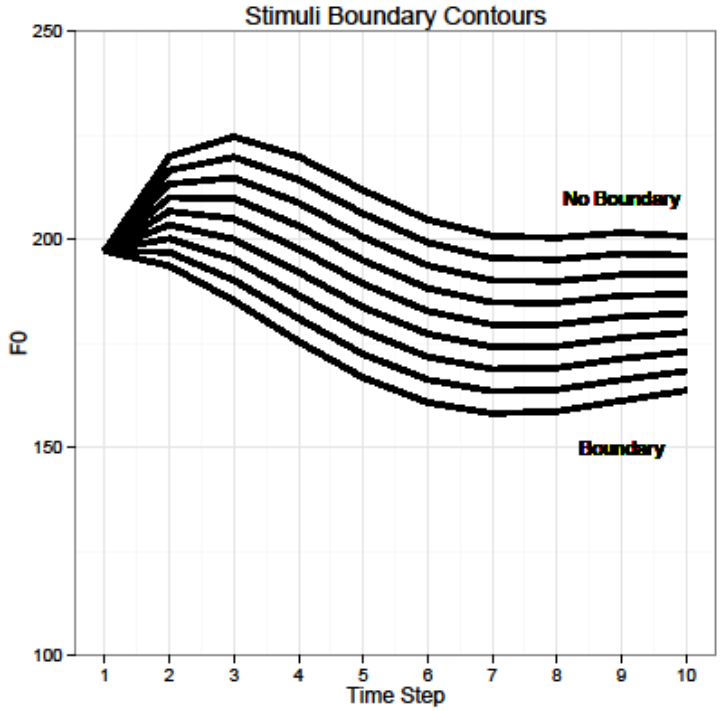

Figure 13: F0 contours for the critical words at each of the steps of the boundary spectrum.

7.1.3 Procedure

All recordings were uploaded to Qualtrics, an online survey service. The survey was posted on Amazon's Mechanical Turk website, where members were able to participate in the survey for pay. Participants were given instructions that broadly explained disfluencies as occurrences in which speakers stumble on some words or need more time to think about what to say next. They are then told that they will hear recorded sentences and they will have to report after what words they heard disfluencies.

For each question, participants saw a media player icon of the recording and under it, the sentence in written form. Next to the sentence, the question read: "There is a disfluency after:" The participants' task was to check boxes under the word(s) they felt preceded a disfluency. Recordings could be played as many times as necessary, and participants could mark as many words as they wanted. The questions were presented in a random order, and all participants heard all 272 recordings. We analyzed the perceived disfluency rate after the 2 critical words for each recording.

7.1.4 Data Analysis

We obtained binary disfluency ratings for each word of each sentence that was presented. We limit analyses to the two critical regions.

7.2 RESULTS

The data were analyzed using logistic mixed effects models to examine how disfluency reports differed as a function of the acoustic cues and critical region (i.e. the locations were there were any acoustic manipulations), as well as their interactions. All logistic mixed effect models were built using the lme4 package in R (Bates et al., 2015). Critical regions were effect coded, and random intercepts and slopes were included for subject and item. The models also included

fixed effects for source sentence and sentence structure. We report the results from the maximal model, following conventions proposed in Barr et al. (2013).

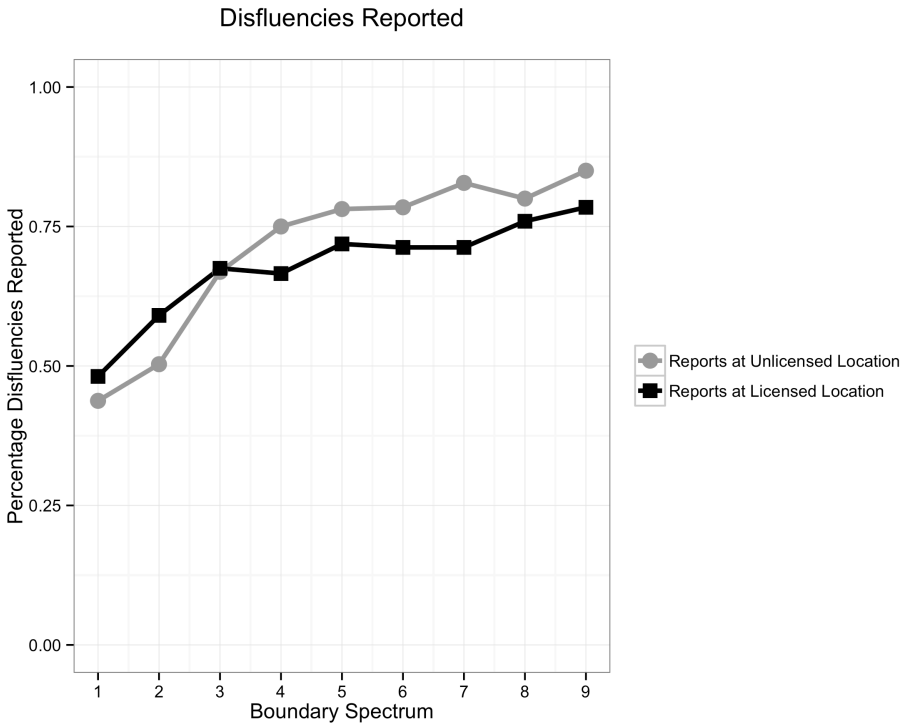

Figure 14. Disfluencies reported as a function of Boundary Spectrum and syntactic location. Licensed and unlicensed locations refer to whether or not the location is likely for an intonational boundary or not, respectively.

Results are presented in Figure 14. There was a main effect of disfluency spectrum ($b = 0.527$, $Z = 4.097$, $p < .001$), such that listeners reported more disfluencies when the acoustic manipulation was stronger. Critically, there was no longer a main effect of critical region ($b = 0.322$, $Z = 0.176$, $p = .068$), indicating that listeners were just as likely to report disfluencies in either of the two manipulated locations. Lastly, there was an interaction between critical region and the disfluency spectrum ($b = -0.107$, $Z = -3.725$, $p < .001$). This suggests that the acoustic manipulation had a larger effect when the manipulated word was in the earlier critical region than in the later region.

7.3 DISCUSSION

Experiment 1 was designed for a few reasons. Firstly, to investigate whether listeners interpretations of prosodic cues would result in the perception of a more likely linguistic category given its context. The results suggest that this is the case: listeners heard the same stimuli as the subjects from Buxó-Lugo & Watson (2016), but their response patterns were drastically different when they are given reason to interpret the prosodic cues as disfluencies. Surprisingly, listeners were more accurate reporting "boundaries" when they thought they were reporting disfluencies than when they were asked to report intonational boundaries. Related to this finding, another goal of Experiment 1 was to shed light on how listener's expectations about disfluencies differed from their expectations about intonational boundaries. The findings suggest that listeners are more flexible as to where they think disfluencies may occur versus where intonational boundaries occur. This assumption likely emerges from listeners' day to day experience with language; intonational boundaries may appear in a few locations, determined by an utterance's syntactic structure and constituent length, but a speaker may face production difficulty at any point in the utterance.

Lastly, Experiment 1 works as a way to make sure that the original results from Buxó-Lugo & Watson (2016) were not due to the recordings having stronger acoustic cues or artifacts in the syntactically licensed location. In fact, when listeners are asked to report disfluencies they are more likely to report them in what would have been the syntactically unlicensed location in Buxó-Lugo & Watson (2016). This suggests that the differences in responses are more likely due to listeners' interpretations of the utterances, which are influenced by their knowledge and expectations about intonational boundaries and disfluencies.

CHAPTER 8: PART 2 - EXPERIMENT 2

An unexpected aspect of the results from Experiment 1 is that listeners are fairly likely to report disfluencies in locations where there should be no acoustic evidence indicating the presence of a disfluency. For example, listeners reported hearing a disfluency over 40% of the time at locations that had been manipulated to have no boundary/disfluency (step 1 in the spectrum). Interestingly, this pattern was also seen for the experiments in Buxó-Lugo & Watson (2016), where listeners still reported more boundaries than expected in the absence of the relevant cues to boundaries even when the manipulated word was at a syntactically unlicensed location. One possible explanation for this pattern of results is that these effects are the result of learning across the experiment. Because listeners only ever hear boundaries or disfluencies at two locations in the sentence, they should be more likely to report hearing a boundaries or disfluencies there, and may even be more likely to do so in a canonical location for whatever phenomena they were listening for. The goal of Experiment 2 was to demonstrate that these expectations were not generated across the course of the experiment. In Experiment 2 we rule this possibility out by exposing each participant to a boundary detection task composed of only two trials.

8.1 METHODS

8.1.1 Participants

Three-hundred English speakers from the United States of America participated in the study. Thirty-eight participants were excluded due to having learned a language other than English from an early age (before 5). This resulted in 262 monolingual English speakers. They were all users of Amazon's Mechanical Turk service, and they all had at least a 90% approval rating for previous task completions. They were paid $0.75 for participating in the study.

8.1.2 Materials

Materials were a subset of those used in Experiment 1. Only recordings that had been manipulated so that both critical words were at boundary step 1 were used (16 total). This means that neither of the critical words had acoustic cues that should signal the presence of a boundary. Each participant heard a random subset of 2 from these 16 recordings.

8.1.3 Procedure

The procedure was similar to that of Experiment 1. However, the instructions now explained that speakers often group utterances into chunks, and that these chunks are often divided by what we call boundaries. They were then told that words that precede boundaries sound "different" than words that do not. Instructions were phrased in this way so that listeners would not explicitly look for cues such as pauses to determine whether there was a boundary or not. There were 2 recordings of sentences with naturally produced boundaries so that listeners could hear them, followed by a sentence indicating where they were likely to have heard a boundary in the examples. The speaker in these recordings was not the same as the speaker who recorded the sentences used in the study.

For each question, participants saw a media player icon of the recording and under it, the sentence in written form. Next to the sentence, the question read: "There is a boundary after:" The participants' task was to check boxes under the word(s) they felt preceded a boundary. Recordings could be played as many times as necessary, and participants could mark as many words as they wanted.

8.1.4 Data Analysis

Data analysis was identical to Experiment 1. Because there was no longer a boundary spectrum manipulation, logistic mixed effect models only tested for the main effect of boundary position.

There were a total of 524 sentences, resulting in 1048 data points.

8.2 RESULTS

Experiment 2 replicated the main results from Buxó-Lugo & Watson (2016). Participants reported hearing a boundary at syntactically licensed locations 59.4% of the time, as opposed to 42% of the time at syntactically unlicensed locations. This main effect of critical region was significant ($b = 0.506$, $Z = 4.185$, $p < .001$).

8.3 DISCUSSION

Experiment 2 was designed to determine whether listeners from Buxó-Lugo & Watson's (2016) Experiments 1 & 2 were developing expectations across the experiment that were driving their reports of hearing an intonational boundary. The results suggest this is not the case. In Experiment 2, listeners only heard 2 sentences. This is unlikely to have been enough exposure for them to develop new expectations about likely locations for intonational boundaries. In addition, none of the recordings they heard were supposed to have signaled boundary presence. The two critical words were identical in terms of duration, following pause duration, and F0 contour. Thus, any difference in reports between the two critical regions was likely due to the syntactic position at which the boundary occurred.

One unexplained puzzle is the relatively high rates of boundaries reported at the syntactically unlicensed location. We think it is likely that these are the result of the acoustic manipulations of the recordings. Although recordings were resynthesized so that they sounded as natural as possible, there were sometimes noticeable changes in speaking rate or F0 from one word to the next due to these manipulations. This could have resulted in the detection of a boundary even for words that were resynthesized to sound like non-boundary words. However, it is important to note that this only explains the overall base rate of hearing boundaries. It does

not explain why listeners report hearing a boundary more often in the syntactically expected location than syntactically unexpected location, where the acoustic signal is exactly the same.

CHAPTER 9: PART 2 - EXPERIMENT 3

Experiment 2 shows that listeners' responses in some of our previous boundary detection studies were not due to them having built expectations as to where boundaries tended to happen within the experiment. However, this presents what appears to be a contradiction. On the one hand, listeners must develop expectations based on the language they are exposed to, and then use these expectations to decipher messages in the future. On the other hand, listeners in these experiments are exposed to numerous sentences with boundaries in (eventually) predictable locations, yet it does not seem like listeners are learning this.

These instances of listeners not adapting to the linguistic realizations they are exposed to is striking, as it seems to set prosody apart from other aspects of language comprehension. There have been various studies that have found listeners adapt to foreign accents (Bradlow & Bent, 2008), dialectic differences in the pronunciation of words (Maye, Aslin, & Tanenhaus, 2008), lexical choices (Brennan & Hanna, 2009), and preferences in syntactic usage (Fine, Jaeger, Farmer, & Qian, 2013). This ability for adapting to changes in the statistics of the language input is beneficial, as it gives listeners a mechanism for coping with the inherent variability of language. It would be surprising if this type of adaptation did not apply to prosody, since the prosodic signal is highly variable and subject to a variety of factors (e.g., Ladd, 2008), and efficient adaptation to prosody can provide the listener with a wealth of information.

In fact, some recent studies suggest that listeners can in fact adapt to new uses of prosody in other realms of language. For example, Kurmada, Brown, & Tanenhaus (2012) found that listeners keep track of the reliability of a speaker's prosodic use and adapt their pragmatic inferences to take speaker reliability into account. However, so far there has been no evidence that suggests that listeners adapt to new mappings between intonational boundary processing and

syntactic structures. Nevertheless, if adaptation is a general mechanism in the language processing system, the mappings between prosodic and syntactc structures should be sensitive to changes in the statistics of the signal. For example, some theories propose that prosodic boundaries function as probabilistic cues to syntactic boundaries (Watson & Gibson, 2005). If this is the case, listeners should be sensitive to changes in how these cues are used and they should learn how they relate to syntactic structures. On the other hand, there are also reasons to believe that listeners will not adapt to these new mappings between prosody and syntax. For example, some accounts argue that intonational boundaries occur where they do because they must divide an utterance into meaningful chunks, and these chunks often correspond with syntactic constituents (e.g., Selkirk, 1984; Steedman, 1991). According to such views, prosodic and syntactic structures truly are inherently linked so that the eccentric use of prosody listeners are exposed to in this study might be treated as noise in the system, since it would not be possible to naturally have intonational boundaries in arbitrary locations. In Experiment 3, we will investigate whether listeners can adapt to speakers using prosody in novel ways, and how they apply this knowledge in order to decipher the meaning behind a message.

9.1 PRELIMINARY EXPERIMENT 3A

One possibility for the apparent lack of adaptation in our previous experiments is that listeners are not learning anything meaningful or useful about the prosody presented. For example, in the sentence "Put the big bowl on the tray," learning that boundaries might happen after "big" does not really help listeners disambiguate the message or complete some sort of task. It is possible that if the utterance were potentially ambiguous, with the prosody being the main disambiguating cue, then listeners would be more likely to adjust their expectations as to where intonational boundaries are located, and what to conclude from that. As such, we use sentences

with relative clause (RC) attachment ambiguities that can be prosodically disambiguated to probe whether listeners adapt their interpretations based on the prosodic manifestations they are exposed to. Before we investigate whether exposure to novel prosody usage can change listeners' expectations as to where intonational boundaries occur, we want to make sure that listeners regularly use prosody to disambiguate the meaning of sentences. In Experiment 3a, we present listeners with test sentences to investigate whether listeners could reliably use prosodic cues to arrive at the intended meaning. The results from this study will also give us a baseline for how listener responses might change throughout the experiment.

9.2 METHODS

9.2.1 Participants

Fifty monolingual speakers of American English were recruited from the University of Illinois psychology subject pool.

9.2.2 Materials

Twenty-four sentences with RC attachment ambiguities were recorded twice, once with the prosody biasing the interpretation toward each possible meaning. For example, for the sentence:

g.) I met the sister | of the actress who was on the balcony.

h.) I met the sister of the actress | who was on the balcony.

example (g) has a boundary after "sister," indicating that the actress was on the balcony. Example (h) has a boundary after "actress," suggesting that it was the sister of the actress who was on the balcony. The sentences were naturally produced, following the experimenters' instructions.

These 48 recordings (along with 24 filler recordings) were uploaded to a Qualtrics survey. Each recording was paired up with a comprehension question. Critically, all critical

trials involve a comprehension question that required the attachment ambiguity to be solved in order to be answered correctly (e.g. “Who was on the balcony?” for the sentence above). The survey was programmed such that each participant was exposed to one instance of the 24 critical trials, as well as all 24 filler trials.

9.2.3 Procedure

Participants saw instructions that told them they would have to answer questions about a few recordings. Before they start the survey, they were given two example exercises that they had to answer correctly in order to continue. Participants were instructed to answer each question with only one or two words. The example sentences did not have the attachment ambiguities that the critical trials contained. Then, participants went through 48 pages, one per recording. For each question, participants saw a play button that they could press in order to listen to the recording. They were asked to listen to the recording only once, and their number of clicks and time spent in each screen was tracked in order to confirm that they heard each recording only once. Below the media player, participants saw a question about the sentence in the recording, with a text box they could fill in. Participants had to input an answer into the text box in order to continue the survey.

9.2.4 Data Analysis

Responses were coded based on whether listeners’ answers followed a low-attachment or high-attachment interpretation. If listeners clicked more than 3 times on a screen, the trial was excluded, as it suggested that they heard the recording multiple times (participants should only need to click twice: once to start the recording, and a second time to input their answers into the text box).

9.3 RESULTS

The data were analyzed using logistic mixed effect models to investigate whether listeners' interpretations of ambiguous RC attachment sentences varied as a function of prosody (in this case, boundary location) and trial order, as well as their interactions. Boundary locations were effect coded, and random intercepts and slopes were included for subject and item. Since the maximal model did not converge, we use the maximal random effects structure that converges, following conventions proposed in Barr et al. (2013).

Models found no effects for trial order ($p = 0.877$), boundary location ($p = 0.590$), or their interaction ($p = 0.252$). A summary of the results can be seen in Figures 15 and 16. These results suggest that listeners were not changing the likelihood they would reach a certain interpretation throughout the experiment. Instead, they answered according to the prosody about 59% of the time. Additionally, there was no noticeable difference in accuracy dependent on whether the boundary was at an early location (indicating low attachment) or at a late location (indicating high attachment).

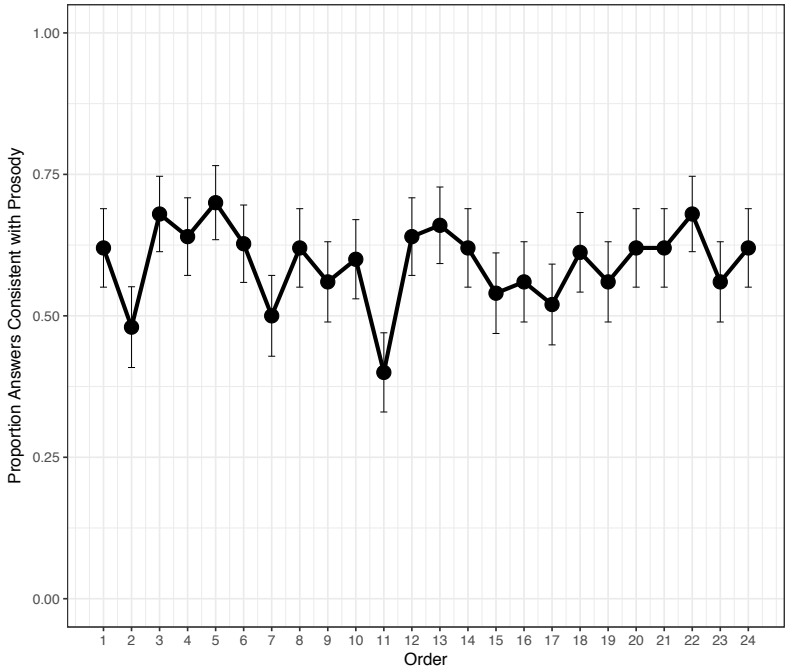

Figure 15. Proportion of answers consistent with the interpretation indicated by the prosody as a function of trial order.

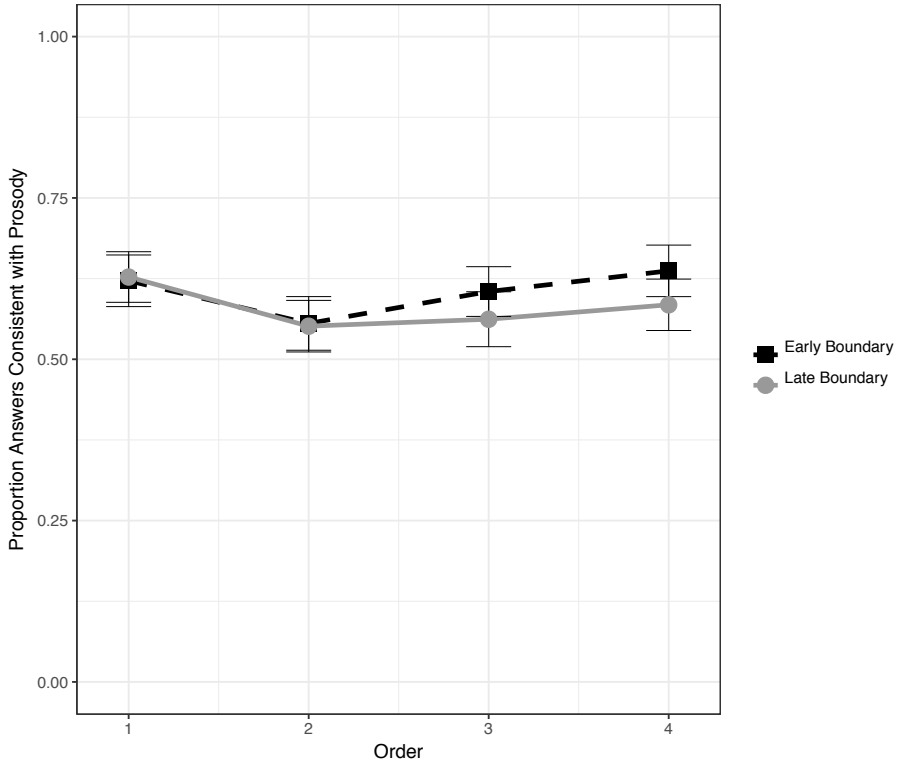

Figure 16. Proportion of answers consistent with the interpretation indicated by the prosody as a function of trial order and boundary location. Trial order is divided into quarters for illustrative purposes.

9.4 DISCUSSION

Experiment 3a was intended as a preliminary experiment to investigate how listeners responded to prosodically disambiguated RC attachment sentences throughout an experiment. The results show that listeners do make use of prosody to disambiguate these sentences, reaching the intended interpretation about 59% of the time. This is useful, as it gives us a baseline to compare performance to in Experiment 3b. Additionally, there did not seem to be any sort of adaptation occuring throughout the experiment. Although participants did not get feedback on their answers, it was a possibility that as they heard more examples of sentences with RC attachment ambiguities, they would be more likely to consider alternate interpretations. However, this was not the case. Lastly, listeners were just as accurate in answering questions about sentences with low attachment (late boundary) as they were sentences with high attachment (early boundary), and this remained the case throughout the experiment as a whole. This is somewhat surprising, as English is said to have a low-attachment bias (e.g., Cuetos & Mitchell, 1988). This would suggest that listeners should be more experienced with early boundary structures and therefore should be more accurate answering questions about it (or at least, have a bias towards that interpretation). This does not seem to be the case in this experiment. Nevertheless, we still explore effects of boundary location in Experiment 3b, as even if there is no difference in response accuracy, there could be an effect on adaptation rates once there is feedback provided.

9.5 EXPERIMENT 3B

Next, we investigate whether exposure to irregular prosody usage results in listeners adapting to the irregular prosody. Specifically, if listeners notice that a speaker has been using prosody in a different way than they are used to, they might learn a new mapping between

prosodic structure and syntactic structure. Although recent studies have shown that listeners can adapt to new mappings between prosody and pragmatics (e.g., Kurumada, Brown, & Tanenhaus, 2012), there is very little evidence that a similar adaptation mechanism is at work for mappings between prosody and syntax. In fact, previous research suggests that intonational phrase structure is highly entrenched and unlikely to be subject to short-term changes in production and comprehension e.g., priming (Tooley, Konopka, & Watson, 2014; Jun & Bishop, 2015). If this is the case, the close relationship between syntactic and prosodic structures may make it so that learning mappings where prosodic boundaries' locations do not correspond with syntactic boundaries too difficult. On the other hand, proposals have been made that this relationship is inherently probabilistic and hence potentially adaptive (Watson & Gibson, 2005). If this is the case, then listeners should have little difficulty adapting to novel mappings between prosodic and syntactic structure, as long as the exposure has been sufficient. In Experiment 3b we use a similar task to that of Experiment 3a in order to investigate whether listeners can use feedback to form new mappings between prosodic and syntactic structures.

9.6 METHODS

9.6.1 Participants

One-hundred English speakers from the United States of America participated in the study. Any participants that have learned a language other than English from an early age (before 5) were excluded, resulting in 78 participants. All participants were users of Amazon's Mechanical Turk service, and they had at least a 90% approval rating for previous task completions.

9.6.2 Materials

The same recordings used in Experiment 3a were used for Experiment 3b. All participants heard an equal number of critical sentences with prosody indicating high attachment and low attachment. As in Experiment 3a, these sentences were accompanied by comprehension questions that probed listeners' parsing of prosodic structure. Sentences were presented in randomized order.

9.6.3 Procedure

Participants saw instructions that told them they had to answer questions about a few recordings. Before they started the survey, they were given two example exercises that they had to answer correctly in order to continue. Participants were instructed to answer each question with only one or two words. The example sentences did not have the attachment ambiguities that the critical trials contained. Participants were also told that in some cases it might be hard to determine what the correct answer to the question was, that they should answer with whatever they think is the correct response, and that they would be given feedback on their performance throughout the experiment. Then participants went through 48 pages, one per recording. For each question, participants saw a play button that they could press in order to listen to the recording. They were asked to listen to the recording only once, and their number of clicks and time spent in each screen was tracked in order to confirm that they had heard each recording only once. Below the media player, participants saw a question about the sentence in the recording, with a text box they could fill in. Participants had to input an answer into the text box in order to continue the survey. After each answer, participants were given feedback as to whether their answer was correct or incorrect. In the congruent prosody condition, participants were told they were correct whenever they answered what the prosody would normally bias listeners to answer

(e.g., interpreting high attachment from a later boundary). In the incongruent prosody, participants were correct when they went against the normal interpretation of the prosody. When participants answer "incorrectly" (based on what condition they are in), they were told they were incorrect and given the "correct" answer for their condition.

9.6.4 Data Analysis

Responses were coded based on whether or not participants answered according to the feedback they got throughout the experiment. Trials were also coded based on the order in which participants saw them. This way, we could track how listeners' responses changed throughout the experiment as they got more feedback. Additionally, we coded trials based on boundary location (early vs. late). Although there was no effect of boundary location in Experiment 3a, RC attachment constructions are not equally likely in the English language (e.g. Cuetos & Mitchell, 1988). As such, we should investigate whether adaptation rates are affected by listeners' familiarity with each construction.

9.7 RESULTS

The data were analyzed using logistic mixed effect models to investigate whether listeners' interpretations of ambiguous RC attachment sentences varied as a function of the type of feedback they got, how much feedback they have received (i.e. trial order), and boundary location, as well as their interactions. Boundary locations and feedback type were effect coded, and random intercepts and slopes were included for subject and item. Because the maximal model did not converge, we use the maximal random effects structure that converged, following conventions proposed in Barr et al. (2013).

A summary of the results can be seen in Figures 17 and 18. There was a main effect of trial order ($p < 0.001$), suggesting that listeners were more likely to match whatever feedback

they got as they progressed through the experiment. There was also a main effect of feedback type ($p < 0.001$), indicating that listeners were more likely to match the feedback in the congruent prosody condition than the incongruent prosody condition. There was no effect of boundary location ($p = 0.458$). However, there was an interaction between boundary location and feedback type ($p < 0.001$), suggesting that listeners responded to the feedback differently depending on what structure the feedback applied to. This interaction suggests that there was a greater effect of feedback type for trials with an early boundary (usually indicating low attachment) than trials with a late boundary (usually indicating high attachment). Lastly, there was a three-way interaction between trial order, feedback type, and boundary location ($p = 0.010$), suggesting that adaptation rates throughout the experiment varied based on a sentence's prosodic structure, as well as whether the feedback was congruent with everyday prosody or not.

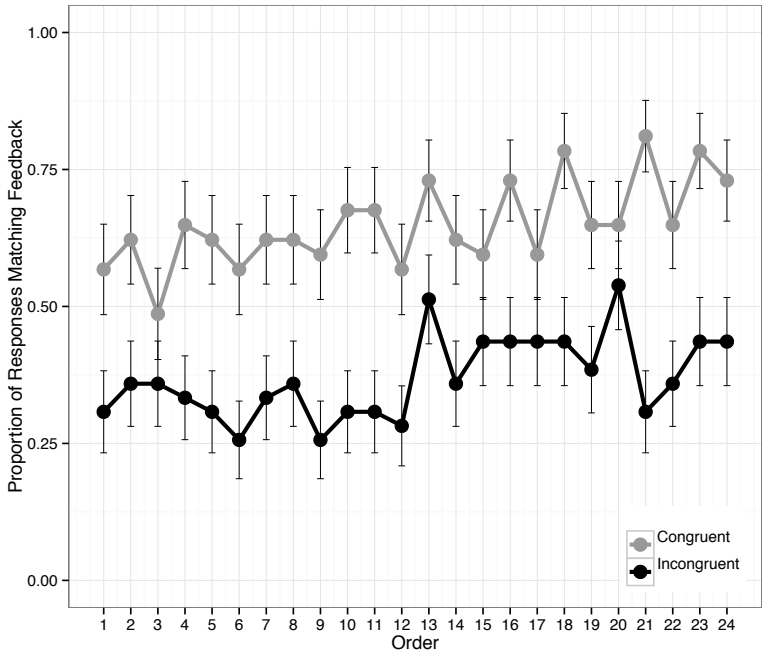

Figure 17. Proportion of listener answers matching the feedback they got throughout the experiment. Listeners in the congruent condition got feedback consistent to everyday prosody usage (early boundaries indicated low attachment, late boundaries indicated high attachment). Participants in the incongruent condition got the opposite feedback.

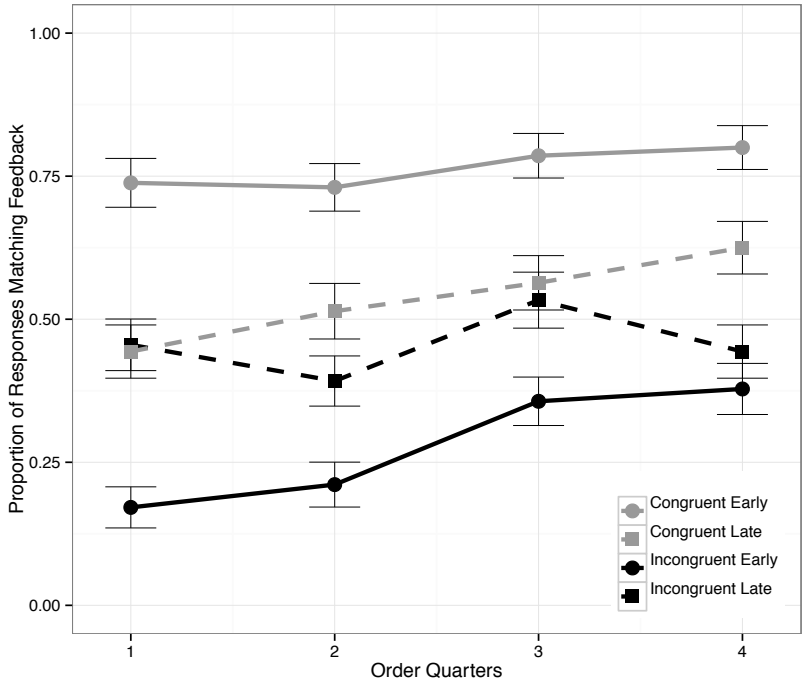

Figure 18. Proportion of listener answers matching feedback as a function of trial order, feedback type, and boundary location. Trial order is divided into quarters for illustrative purposes.

9.8 DISCUSSION

The goal of Experiment 3b was to investigate whether listeners could adapt to novel mappings between prosodic and syntactic structures. The results from this experiment suggest that listeners can in fact learn these new correspondences. Participants were more likely to match the feedback as the experiment progressed, regardless of whether they were in the congruent or incongruent feedback condition.

Critically, there was a three-way interaction suggesting that adaptation rates were modulated by feedback type and prosodic structure. When listeners were exposed to feedback that was consistent with their usual use of prosody, and they heard a sentence in a more familiar construction (early boundary/low attachment), they did not change their responses much throughout the experiment, as they were always relatively accurate about these sentences.

However, when participants heard a less familiar construction (late boundary/high attachment) in the prosody congruent condition, they did adapt such that they increasingly matched the feedback as the experiment progressed. It is unclear as to why participants seem to be less accurate throughout the early trials in this condition when they are getting feedback, compared to when participants answered questions about the same sentences with no feedback provided. One possibility is that the feedback makes listeners more aware of the potential ambiguities in the sentences, and perhaps they default to the more common syntactic structure because of this knowledge. Nevertheless, by the end of the experiment, participants answer the questions correctly at a higher level than the baseline, indicating that listeners can learn how to correctly interpret these less familiar constructions.

For the incongruent feedback conditions, we see no adaptation for the late boundary trials. It is possible that this is because the structure is less familiar and therefore harder to learn a new mapping for it. Alternatively, it is possible that listeners are just learning to discard the prosodic information, and as such stay at around chance levels of answering the question correctly. For the early boundary trials in the incongruent feedback conditions, we do see adaptation. In this case, listeners tend to answer the questions incorrectly for that condition at the beginning of the experiment. However, as the experiment progresses, they are more likely to match the feedback. It is likely that the reason why participants match the feedback so rarely at the beginning of the experiment is because they are familiar with these constructions, and so they usually know how to interpret the meaning of them. However, since they are in the incongruent condition, the interpretation should be the opposite of that which they would normally reach. As the experiment progresses, they are more likely to learn the new mapping, and therefore more likely to answer according to the feedback. By the end of the experiment, participants have not

yet reached the baseline level of accuracy, though. This leaves open the possibility that, again, listeners might be learning to discard the prosody and aiming for chance levels of accuracy.

One tempting explanation for the pattern of results seen in Experiment 3b is that listeners are just learning the rules of the experiment at a higher level. In other words, they are not adapting to new mappings between prosody and syntax, but rather learning to answer the opposite of what they would usually say (in the incongruent condition). However, this does not explain why participants in the congruent condition significantly improve in accuracy for the late boundary condition. In this case, the prosody is consistent with everyday usage, and listeners are learning that prosody is a reliable indicator as to the intended meaning of these potentially ambiguous sentences. This does not eliminate the possibility that listeners are learning to discard the prosody in the incongruent feedback condition. Future studies should investigate how different forms of feedback and manifestations of prosody affect listener adaptation. Nevertheless, it is clear that listeners are sensitive to changes in the usage of prosody, and can change their patterns of responses to result in higher levels of accuracy.

CHAPTER 10: PART 2 - GENERAL DISCUSSION AND CONCLUSIONS

The series of studies presented in Part 2 were meant to shed light on how listeners process prosodic information. Experiment 1 shows evidence that listeners' expectations can influence how listeners interpret a set of prosodic cues, even going so far as to being interpreted as a different linguistic phenomenon: disfluencies. Experiment 2 conceptually replicated previous findings from Buxó-Lugo & Watson (2016): listeners were more likely to report having heard an intonational boundary at a more plausible syntactic location than a less plausible one, even when the major cues for boundary presence were held constant across the two locations. Together, Experiments 1 and 2 suggest that prosodic processing is highly susceptible to the influence of other levels of language processing. These could be parallel processes, like syntactic processing attracting prosodic interpretation, or top-down processes, such as listeners' expectations guiding their interpretation as to the utterance as a whole. In both cases, it seems like listeners reach interpretations that are more likely as a whole, even if this goes against the acoustic evidence they have at hand. In Experiment 1, they hear cues to a boundary, but are likely to report these as disfluencies because of the task they are performing. Interestingly, they are more likely to report disfluencies at the syntactically implausible location for a boundary (although this is not significant). In Experiment 2, listeners' answers result in what would be the most likely manifestation of the utterance at hand (with a boundary at a plausible place for a boundary). The fact that the same types of materials were used for both experiments (as well as Buxó-Lugo & Watson, 2016), suggests that listeners are not just searching for acoustic anomalies or manipulations, but rather they are making inferences as to how that information fits in with the rest of the message, and what it is most likely they heard as a whole.

In Experiments 3a and 3b we explored whether listeners are able to adapt to new mappings between prosody and syntax. Experiment 3a provided us with a baseline rate of

accurate answers for RC attachment sentences that are disambiguated through prosody. Experiment 3b tested whether listeners' interpretations were sensitive to feedback indicating that a new mapping between prosody and syntax was being employed. Our results suggest that they are, although it is still an open question via what mechanism they reach these changes. Listeners were more likely to match the feedback they were provided with as the experiment progressed. Additionally, there were interesting interactions suggesting that the type of structure listeners had to learn to adapt modulated this rate of adaptation.

Altogether, the experiments from Part 2 suggest that prosodic processing in general is highly flexible, as it is subject to changes in interpretation based on the surrounding context, expectations, and feedback. An adaptable system is crucial for such a context, as it provides a good way for listeners to cope with the variability present in prosodic manifestations, especially as we see that prosodic interpretations can be shifted by a variety of factors. Although the present evidence cannot shed light as to what specific mechanisms would be responsible for this type of interactivity and adaptation, it does provide some goals and constraints for future models of prosodic processing. Such a model would have to account for the multiple inputs or influencers that other levels of language and expectations provide, and propose a way in which these different types of information are weighed and integrated in order to reach the most likely global interpretation of an utterance. Additionally, the model must allow for adaptability and learning to occur. Future studies will aim to further shed light on the nature of prosodic processing and the possibility of adaptation in the prosodic domain.

CHAPTER 11: CONCLUSIONS

The present studies were carried out in order to accomplish one main goal: to investigate the interactivity between prosody and the context in which prosody occurs in both production and comprehension. Although the areas in which this was studied varied considerably between studies (e.g., prominence and information status in production, and intonational boundaries and syntax in comprehension), together they reveal some common themes as to how prosody functions and how it fits in with the rest of the language system. Across these experiments, we see evidence that prosodic production and comprehension is highly dependent on context, not unlike many other areas of language (e.g., in production: Schober & Clark, 1989; Brown-Schmidt, 2005; in comprehension: Ganong, 1980; Remez et al., 1981; Kim & Osterhout, 2005). This allows us to speculate as to what mechanisms are responsible for prosodic production and comprehension. One possibility is that similar mechanisms to those used to explain the production and comprehension of syntax, semantics, and/or phonology are also responsible for prosody. However, if this is the case, prosody offers a special scenario in which to observe these mechanisms, as it offers an opportunity to explore how they function when dealing with cues that are noisy, highly confusable, and continuous, and that have to communicate a wealth of information about many other levels of language.

There are still significant questions to be answered about the processes investigated here. For example, now that we know that prosody is influenced by a variety of factors, it is important to investigate how these factors are weighed and how they are integrated in both production and comprehension. For example, in the case of production, how does each aspect of the greater communicative context contribute to the final prosodic production? In the case of comprehension, how do listeners weigh and consolidate information from multiple levels of language processing in order to reach a conclusion about the prosody they heard? The present

studies can serve as a foundation for pursuing the answers to these questions.

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Bates D., Maechler M., Bolker B., & Walker S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-8, http://CRAN.R-project.org/package=lme4.

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.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P . Prieto (Eds.), *Prosody and Meaning* (pp. 239-270). Berlin: Walter de Gruyter.

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/

Bradlow, A. R. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition.* 106, 707-729.

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

Brennan, S. E. & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science, 1,* 274-291.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2012). Metrical expectations from preceding prosody influence spoken word recognition. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review 18,* 1189-96.

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.

Buxó-Lugo, A. & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic

parsing. *Journal of Memory and Language, 90,* 1-13.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23,* 567-598.

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.

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes, 25,* 1141-1177.

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.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234-250.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Cuetos, F. & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of late-closure strategy in Spanish. *Cognition*, 30, 73-105.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63,* 274-294.

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.

Ferreira, F. (1993). Creation of prosody during sentence prosody. *Psychological Review, 100,* 233-253.

Fine, A. B., Jaeger, T. F., Farmer, T. A, & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*.

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.

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic dimensions of variation in disfluency production. *Language, Cognition and Neuroscience, 29,* 1083-1096.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765-768.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110-125.

Garrett, M., Bever, T. G., & Fodor, J. (1966). The active use of grammar in speech perception. *Perception and Psychophysics, 1,* 30-32.

Gee, J., & Grosjean, F. (1983). Performance structures: A psycholinguistic appraisal. *Cognitive Psychology, 15,* 411-458.

Gibson, E., Bergen, L., & Piantadosi, S. T., (2013). The rational integration of noise and prior semantic expectation: Evidence for a noisy-channel model of sentence interpretation. *Proceedings of the National Academy of Sciences*, *11,* 8051-8056.

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 Psychol*ogy*, 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.

Jun, S. -A. & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier and E. Gibson (Eds.): *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor.*

Juslin, P. N., & Laukka, P. (2003). Communication of emotion in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129,* 770-814.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205-225.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40,* 153-194.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3,* 129-140.

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.

Kraljic, T., Samuel, A.G., & Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science, 19(4),* 332-338.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It *looks like* speech adaptation. *The proceedings of the 35th annual meeting of the Cognitive Science Society*, Sapporo, Japan, August.

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.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud witch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32,* 543-562.

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

Price, P. J., Ostendorf, S., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 9,* 2956-2970.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 22,* 947-949.

Schafer, A. J. (1997). Prosodic Parsing: The Role of Prosody in Sentence Comprehension. Doctoral dissertation, University of Massachusetts, Amherst, MA.

Schafer, A. J., Speer, S. R., & Warren, P. (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In M. Tanenhaus & J. Trueswell (Eds.) *Approaches to Studying World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Tradition* (pp. 209-225). Cambridge: MIT Press.

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

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.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Shriber, E. (2001). To "err" is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association, 31,* 153-169.

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.

Steedman, M. (1991). Structure and intonation. *Language, 67,* 260-296.

Tanenhaus, M. K. (2013). All P's or mixed vegetables? *Frontiers in Psychology*, *4*, 234. doi:10.3389/fpsyg.2013.00234

Tooley, K. M., Konopka, A. E., & Watson, D. G. (2014). Can intonational phrase structure be primed (like syntactic structure)? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 348-363.

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.

Turk, A., & Shattuck-Hufnagel, S. (2007). Phrase-final lengthening in American English. *Journal of Phonetics 35,* 445-472.

Wagner, M. & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*, 905-945.

Watson, D. G., & Gibson, E. (2004). The relationship between intonational phrasing and

syntactic structure in language production. *Language and Cognitive Processes*, *19,* 713-755.

Watson, D. G. & Gibson, E. (2005). Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica*, *59,* 279-300.

Yoon, S. O., Koh, S., & Brown-Schmidt. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review*, *19*, 699-707.

APPENDIX A: PART 1 TABLES

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 = *more communicative*	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 = *less communicative*	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?
High-engagement	0.82	0.12	0.33	0.43	0.42	Yes ($p<0.001$)
Low-engagement	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 × 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 = *listener-present*	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 = *listener-absent*	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?
Listener-present	0.60	0.22	0.08	0.13	0.26	Yes (p=0.011)
Listener-absent	0.19	0.19	0.11	0.06	0.14	No (p=0.723)

APPENDIX B: PART 2 TABLES

Table 5. Summary of LMEM results (Part 2, Experiments 1 & 2).

	Experiment 1	Experiment 2
Intercept	b = -0.957 SE = 0.319 Z value = -3.001 p < 0.01	b = 0.047 SE = 0.076 Z value = 0.622 p = 0.534
Critical Region	b = 0.322 SE = 0.176 Z value = 1.828 p = 0.068	b = 0.506 SE = 0.121 Z value = 4.184 p < 0.001
Spectrum	b = 0.527 SE = 0.129 Z value = 4.097 p < 0.001	
*Critical Region * Spectrum*	b = -0.107 SE = 0.029 Z value = -3.725 p < 0.001	

Table 6. Summary of LMEM results (Part 2, Experiments 3a).

	Experiment 3a
Intercept	b = 0.437 SE = 0.178 Z value = 2.456 p = 0.014
Order	b = 0.001 SE = 0.010 Z value = 0.154 p = 0.877
Boundary Location	b = 0.089 SE = 0.165 Z value = 0.539 p = 0.590
*Order*Boundary Location*	b = -0.011 SE = 0.010 Z value = -1.146 p = 0.252

Table 7. Summary of LMEM results (Part 2, Experiment 3b).

	Experiment 3b
Intercept	b = 0.025 SE = 0.069 Z value = 0.354 p = 0.723
Order	b = 0.033 SE = 0.008 Z value = 4.318 p < 0.001
Prosody Congruent	b = 0.651 SE = 0.059 Z value = 10.959 p < 0.001
Boundary Location	b = -0.050 SE = 0.067 Z value = -0.743 p = 0.458
*Prosody Congruent*Boundary Location*	b = -0.473 SE = 0.064 Z value = -7.342 p < 0.001
*Order*Prosody Congruent*Boundary Location*	b = 0.020 SE = 0.008 Z value = 2.566 p = 0.010