Prosodic marking of information structure and contrastive focus in autistic adults

Nishtha Trivedi

Department of Cognitive Science, Rutgers University

Undergraduate Honors Thesis

Dr. Kristen Syrett

1 April 2022

Acknowledgments

I would like to acknowledge everyone who helped bring my thesis project to fruition and completion. First of all, I would like to thank my research advisor Dr. Kristen Syrett, who helped me develop my topic and provided me with the resources I needed to succeed and much need advice and knowledge to fill in the gaps in mine. I am grateful to have had the opportunity to conduct research with you and to be mentored by you. I would also like to thank Courtney Butler, Program Coordinator of the College Support Program, who also advised me on this project, and Caroline and Emma, research assistants in Courtney's lab who helped me run experiments and collect data. I also want to thank Meg Gotowski, a graduate student in Dr. Syrett's lab for helping me set up my experiments and answering questions about research writing, statistics, and more. Finally, I would also like to thank Amy Rosen, another thesis student in Dr. Syrett's lab, for being constantly supportive throughout the year as we built our own projects. Thank you, everyone, for your support!

1. Introduction

Often in verbal conversations, it's not *what* you say but *how* you say it. In the sentence "I didn't say we should kill him," the meaning changes based on which word you stress. For example, I could say, "I didn't SAY we should kill him…" or "I didn't say we should KILL him!" or even, "I didn't say we should kill HIM" (Chilton, 2012). These changes in stress or prominence need not be intentional; they are part of our implicit linguistic knowledge. For example, if someone asked you, "Who called Rutgers University a beloved community?" You might answer, "President Holloway referred to Rutgers this way," stressing "HO-lloway" and de-accenting the given information. And if someone said, "Rutgers is the state university of New York," in correcting them, you would say, "Rutgers is the state university of New JER-sey" using contrastive focus on the stressed syllable. No one taught you this, and you may not be aware that you do it.

In fact, every conversation we have, every utterance we deliver, is riddled with changes in pitch or intonation that convey meaning, our perspective as speakers, and how our mutual knowledge about the world is shared or differs. As Halliday says, when a sentence is spoken, the speaker organizes the words they are saying into an "information structure," and this structure emerges in the speaker's prosody (Halliday, 1967). These prosodic features are therefore an essential part of pragmatics (language usage), because they convey meaning in a given context. Specific stresses and pitch changes mark how we structure given and new information in a discourse. Such prosodic variations seem to come naturally to neurotypical speakers—even early in language development. By contrast, autistic individuals have observed to have "abnormal prosody." However, many of these claims are impressionistic, lacking appeal to specific prosodic features to identify how and why autistic speakers appear to deviate from neurotypical speakers.

There is also a challenge of balancing the investigation of prosody in a natural setting with eliciting consistent utterances in a controlled setting. While the former presents the best case scenario for observing speech patterns that are most natural to a given speaker, it leads to variation in productions. While the latter allows us to compare apples to apples, it can result in 'laboratory prosody,' especially from speakers who otherwise struggle with adapting to conversational settings among new speakers.

In this thesis, I perform a comparison of the prosody of neurotypical and autistic individuals, with a focus on how intonation reveals given-new information structure, using a qualitative multi-pronged approach. Rather than asking about "prosody" in general, I ask specifically how speakers of the two population compare when producing utterances that highlight a contrast between the topic and focus (terms I will define below) and given and new information. In doing so, I unpack the claim that, "Autistic people don't produce typical prosody." I instead show that they *do* share many of the prosodic hallmarks of a neurotypical speaker and not monotone, and incapable of manipulating their intonation. Rather, the difference appears to lie in how they manipulate prosody within an utterance to highlight what is known or new between speakers. I conclude that the challenge lies in pragmatics and establishing common ground, rather than prosody itself.

In the next section, I provide a brief background on prosody in general, turning to research on prosody in typical and autistic individuals. I then present the current experimental research, which consists of a sentence-elicitation task embedded in a narrative task, allowing us a baseline of normal prosody relative to a given speakers. In section 4, I present the results of the elicitation task, against the backdrop of the narration task and a naturalistic interview we have for

one of the autistic speakers. In section 5, I conclude with the main findings and propose future next steps.

2. Background

2.1. Prosody and information structure

Prosody refers to pitch or intonation, duration of vowels, consonants, syllables, the intensity or amplitude, and pauses within or in-between utterances (Gussenhoven, 2004). As I mentioned in the introduction, prosody serves many functions. It is used to distinguish speech acts such as questions, statements, and imperatives (Eigsti et al., 2011).; to convey what information is old (given) or new (DePape et al., 2012). For example, a low boundary tone at the end of sentence, denoted by L-L% indicates that the pitch ends at a low value at the end of the phrase. This contour is typically seen in simple declarative statements ("I can't attend class today."). An H-H% boundary tone indicates that the pitch ends at a high value at the end of the utterance. This contour is typically observed when we pose questions that require a yes or no answer ("Are you coming to class today?"), or even in declaratives intended this way ("You came to campus with COVID?") (Eigsti et al., 2011). *Wh*-questions have a different intonational profile than yes/no questions. (Consider when you ask, "Alexa, what is the weather like today?"). Intonation boundaries indicate clauses and items in a list (Fine et al., 1991). Thus, intonation signals how a speaker wants their utterance to be interpreted to the listener. See Beckman and Pierrehumbert (1986).

Affective prosodic cues also signal the speaker's affective state (e.g., happy versus angry) (Eigsti et al., 2011). Prosody that deviates from the expected may signal a marked stance. For example, monotonous-sounding speech due to a narrow pitch range can give a listener the impression that the speaker is unhappy or doesn't care. An unexpectedly wide pitch range on a

declarative utterance may sound inappropriate, insincere, or mocking. As a result, perceptibly "unusual" prosody can hinder social acceptance and the formation of relationships for autistic people (Peppé et al., 2006). Since these speech features are suprasegmental and above the level of the word and syntactic structure, they are independent of the word choice or grammar of the speaker's utterance.

Within prosody, "stress" refers to how we focus on a linguistic element (that is, make it more prominent). We can do this using intensity (loudness), pitch frequency (high/low pitch), pause duration, variation in prosodic rhythm, and pitch contours (McCann and Peppé, 2003). For example, saying "blue" louder than the rest of the words in the sentence and holding the vowel longer than usual in the statement, "I want blue socks," can indicate that "blue" is a focus of the sentence or that "blue" is being uttered in contrast to something else, perhaps another color (Halliday, 1967). Stress is also used to mark information structure, indicating the topic and focus of the sentence. The *topic* conveys *given* information, and the *focus* indicates *new* information. The *tune* of an utterance can be categorized into pitch accents that mark different instances of prominence in the sentence as well as the way the sentence ends (Lambrecht and Michealis, 1998). When a speaker puts focus on a word, the pitch accent is typically an L* or H*. An H* involves a pitch peak on, or near, the accented word's primary stressed vowel. An L* involves a pitch trough, on, or near, the accented word's primary stressed vowel. See Pierrehumbert (1980).

For example, when the sentence "The boy is eating an apple" is uttered in response to the question "What is the boy eating?" the word "apple" is made more prominent (marked by H*). Here, the boy is the topic and "apple" is the focus (what the first speaker asked about). By contrast, in response to the question, "Who is eating the apple?" a speaker who answers, "The boy is eating an apple" would most likely stress "boy," since it is now the focus (DePape et al.,

2012). This contrast between the topic and focus is also referred to the "given-new contract" by Clark and Haviland (1977). The basic idea is that a sentence consists of information the speaker and/or listener already knows, and new information. This distinction is marked by the "information focus," which comes with a distinct intonation contour as detailed by Halliday (1967). Note, too, that we mark the given (or presupposed) information with a definite determiner (*the boy*, in the first question, and *the apple* in the second question)—another cue to the given-new contrast. Thus, in addition to prosody, we have other grammatical elements that signal information structure (definiteness, pronominal reference, word order, and other constructions).

Building upon focus, a speaker can also manipulate prosody to make apparent that there is a contrast between a specific part of the utterance and some other entity. If our example sentence, "I want blue socks," is uttered in response to "Here's your red socks," the word "blue" would most likely be spoken with a contrastive prosody: a certain type of pitch accent (L+H*) placed on the word to indicate the speaker is contrasting "blue" against the term "red" given by the speaker. For narrow contrastive focus, one looks at a lexical item or syllable in the utterance that signals meaning. For example, contrastive focus can be used to correct someone's pronunciation by accenting a particular lexical item within the utterance (Harrington, 2020).

2.2. Prosody in Autistic Individuals

All of the above observations about how prosody interacts with information structure hold for typical speakers across contexts. One might ask, then, if all of these observations hold with individuals who are known to struggle with speaker-hearer interactions and communicating clearly and efficiently. One case in question is individuals diagnosed with Autism Spectrum Disorder.

Autism Spectrum Disorder is a genetically-based neurodevelopmental disorder, with many different presentations based on varying levels of severity on the spectrum. It can be characterized by impaired social communication, causing deficits in social relationships and verbal and nonverbal communication, repetitive behaviors, hypersensitivity to various sensory stimuli, and intellectual deficits (American Psychiatric Association, 2013). Compared to neurotypical individuals, the speech of autistic individuals has been claimed to exhibit "atypical prosody," including atypical stress patterns and abnormalities in intonation and pitch range (Baron-Cohen and Staunton, 1994). People have described the prosody of autistic speech as "monotone" and "bizarre." (Baltaxe and Simmons, 1985). Autistic people seem to use a limited number of intonation contours in their sentences consistent with claims that their prosodic variation does not always align with communicative intent (Green & Tobin, 2009). Baltaxe's contrastive stress survey in autistic children using an elicitation task showed that autistic children misassigned stress: some did not use it at all, while some used stress more than necessary – stressing the focus word in the sentence, but also other lexical items in the utterances (Baltaxe, 1985).

Fine et al. (1991) collected 10-minute samples of spontaneous conversations between an investigator and three groups of children aged 8 to 18: children with an Asperger's syndrome (AS) diagnosis (this term is no longer used), children with a diagnosis of high-functioning autism (HFA), and neurotypical children. They analyzed the tone boundaries and stress phonetically and judged them inappropriate or appropriate based on 5 patterns of marked stress: stress on the final word, stress on the function word (such as prepositions and pronouns), stress on the content word (object or subject in the sentence), and stress on other words in the sentence like auxiliary verbs. Each marked stress pattern conveys something specific, usually the placement of stress marks the

new information in the sentence, or information that the speaker intends to highlight due to some presupposition or to correct something (Fine et al., 1991). Their research revealed, the HFA children demonstrate irregular use of stress in their speech, using more unmarked speech patterns and using aforementioned marked stress patterns atypically, while children in the AS group were similar to the neurotypical children in their use of stress (Fine et al., 1991). Fine et al. posited that this could be due to an inability to produce correct stress intonation patterns, or inability to assess the listener's knowledge of the given and new information in the sentence, or a lack of awareness of the value of varied stress patterns in conversations (1991).

Given divergences in prosodic production, there is an open question of whether the issue is a deficit in understanding of the various functions of prosody and how to use it appropriately or in executing it, not realizing that one's placement of stress in a sentence is excessive, misplaced, or missing unintentionally (McCann & Peppé, 2003). At the same time, there may be challenges with recognizing prosody when uttered by another speaker. Peppé et al. (2006) used the Profiling Elements of Prosodic Systems in Children (PEPS-C) procedure to analyze prosody of on autistic child, targeting receptive prosodic skills, which include understanding the effect of tone of another speaker's voice and understanding the focus or emphasis of an utterance. Their results suggest that autistic children may not have sufficient receptive prosodic skills (Peppé et al., 2006). Baron-Cohen has done seminal research on theory of mind and autism, and the social impairments that come with autism. Theory of mind refers to an understanding one has of the mental states of other people, and the ability to suppose what another person may be thinking or feeling (Meltzoff et al., 1993). Theory of mind is relevant in understanding why autistic people may have atypical prosody (Baron-Cohen, 1995). Understanding another person's prosody, their knowledge about a subject, their pragmatic intention in conversation, all of these contribute to

how one uses pitch, stress, intonation to converse with someone. Within the studies investigating prosody in autistic individuals, many more have been conducted on children than on adults (Baron-Cohen and Staunton, 1994; Baltaxe et al., 1985; Fine et al., 1991).

Thus, more research is necessary to understand the use of prosodic elements in ASD speech in adults. One recent exception is a study by DePape et al. (2012). In this study, both six neurotypical and twelve autistic participants (six with high language function, and six with the moderate language function) were shown images on a computer screen of an animal performing an action (e.g., a rabbit painting a bed). However, part of the scene was occluded (either the painter or the object being painted), giving rise to an opportunity to ask *who* or *what* questions that probed topic and focus, as described above. Participants were asked a question about the scene and were asked to produce an answer (and in doing so, mark topic and focus). The researchers used Praat 5.1.0.7 (Boersma and Weenink, 2009) to analyze participants' pitch range and duration (DePape et al., 2012). They found pitch to be the primary contributor to general abnormal prosody in autistic individuals: relative to the controls, the autistic participants had smaller pitch ranges. Within the autistic participant population, those with moderate language skills had smaller pitch ranges than those with stronger language skills (DePape et al., 2012). Language skills were measured based on the Peabody picture vocabulary test (Dunn and Dunn, 1997). Rise in pitch and fall in pitch are both important markers for information structure in a sentence, as discussed earlier. Neurotypical participants had significantly larger pitch rises and pitch falls than the autistic groups. Participants of the autistic group with high language skills did not vary their pitch significantly to mark information structure at any position in the sentence, even if they had a general modulation of pitch (DePape et al., 2012). This is an important finding, because it claims that it is not the case that autistic people cannot vary their pitch,

however the way they are using stress and intonation is irregular. It is not being used effectively to contrast between given and new information, thus autistic people are not appropriately signaling to their conversation partners what their pragmatic intentions are in the conversation or that they understand the mental states of their conversation partners.

While DePape et al. provided initial information about how autistic speakers mark topic and focus, there are limitations to their study. First, it was in a highly controlled setting, and may have produced artificial speech patterns. Second, they did not probe contrastive focus. Finally, they did not have a baseline against which to assess other markers of information structure beyond prosody. The current study is designed to build on their work, fill these gaps, and expand our knowledge of the prosody of autistic speakers.

2.3. Current Research

The summary above has highlighted a gap in our collective understanding of what makes prosody come across as "atypical" in autistic individuals. One could attribute the cause to lack of knowledge of how to optimally communicate intent to others, based on what they know (and how this differs from the speakers), pragmatic knowledge, or an impaired ability to represent and/or implement target prosody. That is, if we assume prosody is a surface level indicator of an underlying abstract information structure, and that information structure is crucially built upon engagement between a speaker and a hearer, then if we notice that the prosody seems atypical, it could be because the knowledge of information structure is what is at fault, or the implementation of prosody itself. In this study, we tease apart two hypotheses: one, that autistic individuals lack knowledge of information structure (and therefore fail to mark the given-new contrast or signal contrastive focus with prosody), and the other, that autistic individuals have this linguistic knowledge, but do not recruit prosody as neurotypical speakers do to signal it. We

then further ask, if autistic speakers *do* use prosody (and are not entirely monotone), where and how do these deviate from neurotypical speaker patterns. This study thus provides a clearer, more fine-grained picture of how these individuals are using or are not using prosody to signal information structure, manipulating topic and focus and contrastive focus within a task, and situating this against the backdrop of other communicative strategies.

One final note: this thesis is a collaboration between the Laboratory for Developmental Language Studies and researchers in the Graduate School of Applied Psychology. Our hope is that results from our analysis will inform speech and behavioral therapies for autistic students at Rutgers, in particular those who have noticed that they themselves "sound different" and want to know how and why, and what to do about it. We hope to provide concrete data and practical advice.

3. Experiment

3.1. Participants

Participants were three neurotypical male adults (age range: 20-22, mean age: 21) and three male adults who had been diagnosed with Autism Spectrum Disorder (age range: 20-21; mean age: 20). All participants were undergraduate students at Rutgers, The State University of New Jersey – New Brunswick. Informed consent based on an IRB-reviewed protocol was obtained before participation.

3.2. Design and Materials

The task used was an elicited production narration task. It was modeled after the well-known 'frog stories' (Berman & Slobin, 1994). In this task, participants are shown a series of scenes, and asked to produce their own narration of the plot that unfolds. Their speech is then evaluated for key features. In the current work, instead of showing participants a set of still images from a

book and then asking them to produce a narration, participants were presented with a series of short pre-recorded videos representing five distinct events, with multiple scenes therein. The narration took place in real-time, as the scenes unfolded. Second, these videos came from Syrett & Goldin (2019), and have successfully used as elicited narration tasks for typically developing populations of monolingual, bilingual, and heritage speakers ranging from early elementary to undergraduate level. They were specifically designed to investigate morphosyntactic correlates of information structure, production of perspective-taking verbs, and reference, making them ideal for the current purposes.

The materials included a series of five video scenes filmed using a Sony digital camera mounted in a tripod, then edited in iMovie to remove the sound. Each of the five scenes was divided into four to five shorter scenes of two to three minutes, with a 10-second black screen in between. Within each larger scene, a story was acted out over the course of the shorter scenes, involving a group of animals (research assistants dressed in animal costumes). These scenes included the following themes: Duck is hungry, Duck is overwhelmed, Duck is lonely, Duck is bored, Duck is invited to Frog's birthday party. The characters were recognizable animals that could be referred to by animal name (duck, pig, cow, frog). Participants were asked to narrate the scenes.

Within each of the shorter scenes, one or more characters interacted with each other. The actors used exaggerated, pantomimed gestures to indicate emotions, physiological or psychological states, or desires (e.g., being hungry or sad or overwhelmed, having an idea, consulting with another character, etc.), so that the participants could easily detect affect and intent. These gestures were normed during video creation. Specific props were chosen to highlight modes of communication (e.g., a toy flip phone, or an office phone), the

aforementioned state (e.g., an empty donut box or a messy pile of toys), direction of character motion (e.g., putting on a scarf before exiting through a door), or the kind of activity (e.g., toy cookies or cupcakes or birthday decorations, or a soccer ball).

Figure 1 illustrates a scene where the duck, concerned about a problem they're facing, calls the pig. The duck has an empty donut box next to them (which previously left them upset), while the pig is holding food. After the pig visits the duck and shares her food, she puts on her scarf and departs, and they wave good bye to each other as shown in Figure 2.

Figure 1. Video 1, Scene 2: The duck is calling the pig.

Figure 2. Video 1, Scene 7: The pig leaves.

3.3. Procedure

Participants were run one at a time. interacted with another. They were brought into the lab by a researcher, welcomed, and then seated at a table in front of a 21" iMac computer. A confederate (a second researcher, the author) was seated in front of a similar device, further down from the participant, at an adjacent table. The iMacs were angled away from each other, at approximately 90º. Both participants were then introduced to the task as the researcher that welcomed them, reading a script. See Appendix A. They were instructed to observe the videos and narrate the actions that unfolded in each video. Participants were told that the confederate would be watching the same videos and wearing noise-canceling headphones during this time, so they could not hear the narration. They were also told that the confederate's screen would be obscured so that they were unable to see all aspects of the events unfolding, including which animal was performing an action or which action was being performed. In reality, nothing was obscured.

Between each of the scenes, when the screen became black, the confederate removed their headphones and either delivered an assertion as an attempt to recap the story or asked a question about the previous scene. The participant's job was to provide the correct information, affirming or correcting the statement, and answering the question. These questions and assertions were of various types: information-seeking questions, as in (1) (a *wh*-question where the answer focuses the subject), assertions that aim to elicit an affirmation, as in (2), and assertions that elicit a correction (and corrective focus), as in (3).

(1) Question: Who called the Duck?

Answer: The Pig called the Duck.

(2) Assertion: The Pig put on a scarf.

Affirmation: Yes, the Pig put on a scarf.

(3) Assertion: The Duck called the Frog.

Correction: No, the Duck called the PIG.

The session lasted approximately 40 minutes. They were recorded using the Garageband application on the Mac computers.

4. Data Preparation and Analysis

The sound files were converted to mp3 files, then opened in Praat and saved as wav files for analysis (Boersma and Weenink, 2009). The audio segment of each participant's elicited utterance after a prompt was spliced out. We used Praat to analyze these spliced out segments of audio and extracted the pitch contours and the pitch value for stressed words or syllables in the utterance. In a workbook in Microsoft Excel, these utterances were then categorized into the prompt type, and we catalogued the expected response to the prompt and the actual response. Within question-answer pairs, we catalogued whether the focus in the response was in the subject or object position, as well as the pitch of every stressed word (F0), the boundary tone, and whether the target was accented. Within the contrastive cases, we catalogued the F0 for each stressed word, whether there was contrastive focus, the pitch type, as well as boundary tone. While we did not perform acoustic analyses, the pitch values allowed us to see relative decline of pitch in declarative statements, even with focus, and to determine a pitch range. The boundary tone allowed us to check that for both participant groups, questions and statements have the correct prosodic contour.

For the analysis, certain responses from participants were eliminated. Upon prompting them with a question or statement such as "Who was upset?" the expected response was along the lines of "The DUCK was upset" (capitalization indicates emphasis on target word). If the participant gave us no response, or if their response was unintelligible or inaudible, it was

eliminated from the analysis. Moreover, certain participants elicited audible glottalizations in their speech which made it impossible to track their pitch contour or get a pitch value in Praat. Glottalization is the production of voiceless sounds and reinforcing voiceless plosives like /p, t, k/ (Williamson, 2018). Thus, if participants produced glottalizations on the target word, we weren't able to analyze the pitch type for that word and had to eliminate those responses as well. Out of a total of 150 responses from both groups of participants, 17 were eliminated due to the issues mentioned above.

5. Results

5.1. Questions

When participants were asked a *wh*-question prompt, the neurotypical population accented the target word 95% (i.e., produced an H* or L* pitch accent on the main vowel of the target word to increase its prominence) of the time, while the autistic individuals accented the target word 87% of time. We then looked at the position of the target word in the sentence (subject or the object). When the target word was in the subject position, the neurotypical population accented it 100% of the time and the autistic population accented it 96% of the time. When the target word was in the object position, the neurotypical population accented it 82% of the time, while the autistic population accented it 64% of the time. We then looked at whether both groups' prosody was in line with what is expected for declarative sentences, an L-L% or H-L% or L% boundary contour, and found that both groups produced this 100% of the time in response to the *wh*- questions. Moreover, we counted how many times non-target words were accented in the participants' responses. For autistic participants, 61% of the responses included prominence on non-target words, while for neurotypical participants, this occurred 42% of the time.

There was no difference in the rate of target accenting between groups, as shown by a Pearson's Chi-squared test with Yates continuity correction (1) c^2= .52 (p=.47), either in subject or object position (subject: c^2= 2.6, p=1; object: c^2= .23, p=.63). There was a main effect of contrastive focus (c^2= 6.33, p=.012), and a difference in the contrastive focus on the subject (c^2= 6.13, p=.013), but no difference in the contrastive focus in the object (c^2= .26, p=.85).

5.2. Statements

In other prompts, participants heard a statement that elicited either affirmation or correction, We focus on the correction here. Production of contrastive focus involves a characteristic L*-H or L-H* pitch type on the main vowel in the word. We found that the neurotypical population produced contrastive focus 68% of the time, while the autistic population only produced contrastive focus 31% of the time. See Figure 3.

We then looked at the results for when the target corrective word was in the subject versus object position. When the target word was in the subject position, the neurotypical participants produced contrastive stress 78% of the time, while the autistic population produced contrastive stress only 28% of the time. See Figure 4.

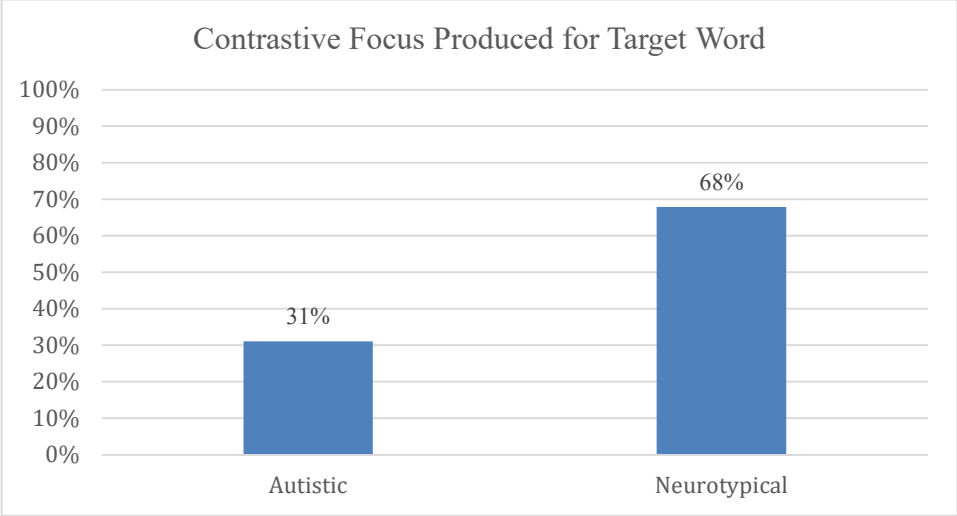


Figure 3. Percent of responses where contrastive focus was produced for the target word.

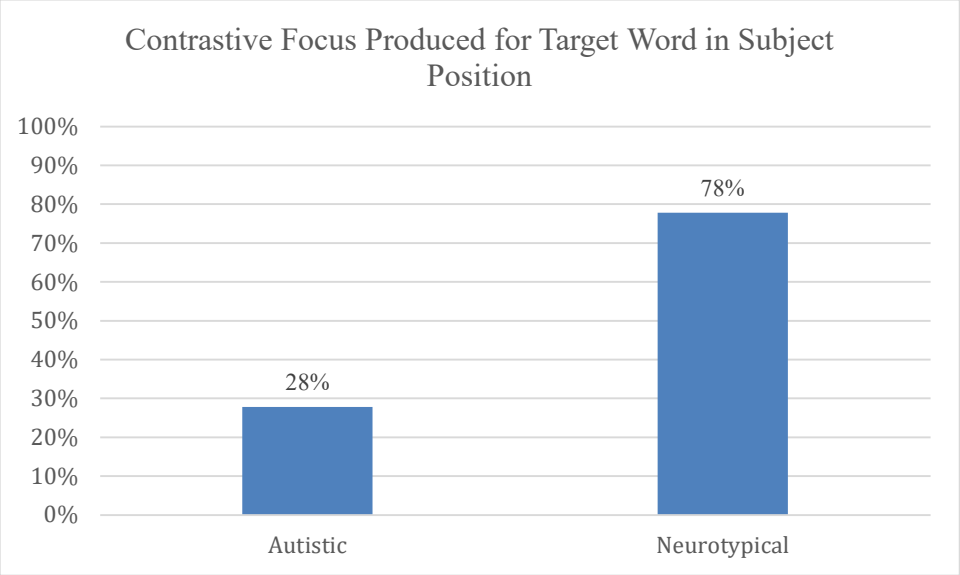


Figure 4. Percentages of responses where contrastive focus is produced for the target word when it is in the subject position, from both autistic and neurotypical participants.

For the object position, neurotypical participants produced it 50% of the time, while autistic participants produced contrastive stress 36% of the time. We analyzed how many times the participant placed prominence on the target word in a non-contrastive fashion, perhaps with regular focus L* and H* pitch accents. For autistic participants, 90% of the time that contrastive focus was absent, there was still a pitch accent placed on the target word. Moreover, autistic participants placed non-contrastive prominence on non-target words 48% of the time, while neurotypical people did this 29% of the time. Thus autistic individuals appear to have used regular focus instead of contrastive focus, and over-accented other words in the sentence, thereby not highlighting the given-new contrast. However, within the autistic individuals, there was variation. See Figure 5.

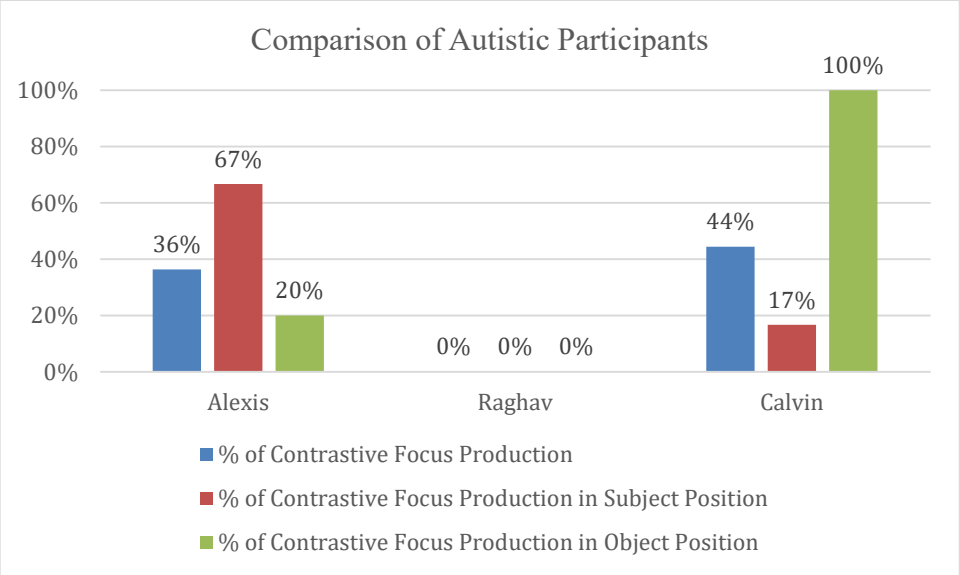
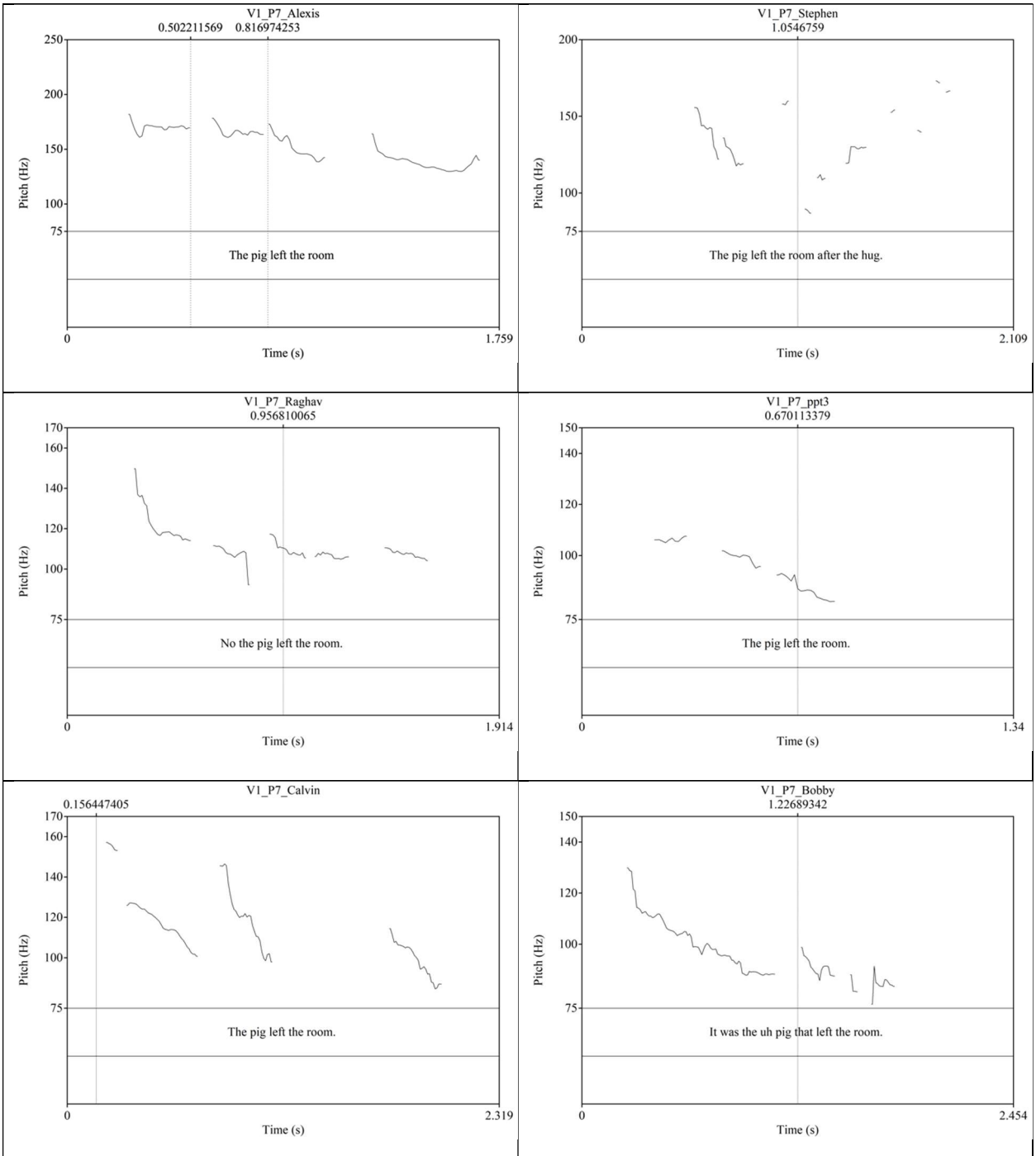


Figure 5. Contrastive focus production compared among autistic participants.

The contrast between a neurotypical and autistic individual with respect to contrastive focus can be captured by looking at the pitch track itself. Unfortunately, since the names of the animals in these videos have obstruents (/d/, /p/) and fricatives (/f/), the pitch track is disrupted. However, the comparisons are most clear in particular instances, as with the correction to the assertion prompt, “The duck left the room” to “No, the PIG left the room.” See Figure 6.



Autistic individuals	Neurotypical individuals

Figure 6s. Pitch tracks for responses for Video 1 Prompt 7: “The duck left the room.”

Given the ways in which focus differed between neurotypical and autistic individuals, one might wonder if these participants were marking information structure at all. Here, the transcribed narratives are important. The autistic individuals successfully marked given information with definite determiners and pronouns, and new information with indefinite determiners, as shown by the following narrations. Moreover, they are descriptive in their reference, highlighting quantities of sets.

(4) video 2, scene 3 (Alexis, autistic)

The Frog came with **a bag** to pick up **the Duck's mess**.

They're dancing with **the bag**.

The Duck the Frog leans over.

And gives the Duck **a bag**.

The Duck holds **the bag** .

While **the Frog** picks up **the mess**.

And puts it in **the bag**.

(5) video 4, scenes 4 and 5 (Alexis, autistic)

The Duck, the Duck is waiting.

And hears **the animals** coming behind the door.

He opens **the door** and sees the three animals.

They come inside with their balls.

They all hug successively with the Duck.

And then they, then **they**, **all the three animals** take off **their scarves**.

No, **only** the Pig off his scarf and **they** play.

They, the Pig and the Frog play around with **the beach ball.**

The Duck play and the Cow play with **the soccer ball**.

Complementing these elicited prompts and narration, one of the autistic participants was invited to participate in a series of 5-minute unscripted interviews with a researcher, in which they were asked about topics that had been pre-selected as being either 'preferred' or 'non-preferred' topics to that speaker. Both individuals communicated via a video platform and viewed each other side-by-side on the screen. The sessions were recorded for later viewing, and then transcribed, and prepared for analysis. We appealed to this naturalistic interview to see whether this speaker demonstrated normal prosody outside of contrastive focus, and whether there, too, he marked information structure in other ways.

Our analysis showed that while he had the right 'shape' to his intonations (rising or falling when called for), they were depressed relative to a neurotypical norm. While he did ask questions of the interviewees ("Is that because your Jewish heritage?") and occasionally make jokes that picked up on a previous statement by the interviewer, he does not appear to use prosody to do things like fail to accommodate a presupposition, as shown in (6).

(6)

Interviewer: Is that why you like shopping and clothes so much?

Autistic individual: I don't like clothes so much. It's kind of boring actually. Just once or twice.

In addition, there are times when the autistic speaker does not wait for a pause to take turns and ask a question, as shown in (7), which has more declarative prosody and no rise, then followed by a joke.

(7)

Interviewer: Um, she [mother] actually did not go to college, fun fact for you-

Autistic individual: High school?

[probing question, did not have H%, sounded more like a statement]

Interviewer: She went to high school in...

Autistic individual: 1900? (laughs) sorry.

The same individual when describing a process and generating lists uses continuation rises at the end of prosodic and syntactic chunks, and accompanies his descriptions with co-speech gestures, as shown in (8).

(8)

Interviewer: So, what is the motherboard?

Autistic individual: The motherboard is the main chip of the computer.

[H on 'mo' but doesn't seem to signal topic/focus or drawing on given information, co-gestures with hands parallel, like grabbing a motherboard on 'main'; doesn't seem conversational, seems explanatory]*

It has the processor, it does instructions, it has…and it has paths connecting the processor to the peripherals…and to the different like plugs.

[list intonation on 'pro', 'stru', 'paths' a, co-gestures with left thumb then with hand as listing items, co-gestures with left hand in 'O' formation on 'plugs'.]

On a plug, you could put like memories, uh, GPU uh video, power source, like to bring the power from the from the outlet to the computer to the motherboard.

[co-gestures with pinching formation and downward motion with left hand for 'memories' 'GPU' and 'video'; 'power source'; co-gestures with left hand up and then brings down when describing bring the power to the computer; list intonation

6. Discussion

This study presented a quantitative and qualitative analysis of autistic and neurotypical prosody with the goal of investigating autistic participants' pragmatic knowledge and use of prosody to differentiate given and new information, as well as contrastive focus to correct a speaker's utterance. We found that autistic speakers emphasized the target word less than neurotypical people. Comparing the subject and object positions, autistic participants focused the target word less when it was in the object position, especially when compared to neurotypical participants who did so for significant majority of the time. One may posit that deaccenting or lack of pitch accent is because the target in the object position often occurs at the end of the sentence, such as in the response: "The duck called the FROG," and the end of a declarative sentence in English often requires a falling or flat contour producing a low (L%) boundary tone (Ladd, 2008).

As noted above, autistic participants always produced the L% boundary tone required of declarative sentences. On the other hand, we also noted autistic participants overly pitch accenting non-target words more than the neurotypical participants. It is clear from these results that the autistic participants are capable of accenting the target word and recognizing that new information requires prominence. However, they are not deaccenting the rest of the information, especially what is already known to the listener. In English, normally non-focus words are deaccented, in order to provide prosodic emphasis to the focused words (Kim, 2018). It is

possible that autistic individuals are not deaccenting non-focused words consistently, and this is contributing to a perception that their speech is “bizarre” or “stilted” (DePape et al., 2012).

Autistic individuals also appear to be producing contrastive focus significantly less consistently than neurotypical individuals. We saw that almost all of the responses given by autistic participants included focus on the target word, but it wasn’t always contrastive focus. Previous research with autistic children has shown significant differences between typical and autistic children when marking contrastive focus, with autistic children doing it less than neurotypical children (Rapin et al., 2015). DePape et al.’s study showed that autistic individuals used narrower pitch ranges than neurotypical ones, and thus they were not marking information structure correctly and not differentiating appropriately between topic and focus in a sentence (2012). Their results thus showed that problems autistic children have with prosodic markers in childhood persist through adulthood (DePape et al., 2012). Thus, our study reinforces this claim, especially in regard to contrastive focus.

We also (briefly) compared our results to the narrations and the naturalistic interview we had for one of the autistic participants that we had annotated with qualitative analysis of prosody. From those interviews, we gathered that participant one is definitely not monotone or stilted in his speech. He often uses list intonation when listing things, rising intonation when he emphasizes something. However, when asking a question, he often does not have the characteristic H% rising boundary tone at the end. Often, when he answers a question asked by the interviewer, he wants to correct the interviewer or reject his assertion, but his prosody does not indicate that he failed to accommodate the interviewer’s presupposition. This qualitative analysis again shows that autistic people are perhaps inconsistent in their use of prosodic elements, which contributes to a perception of atypical prosody.

7. Future Directions

Admittedly, we were only able to collect and analyze data from six participants: three autistic and three neurotypical, thus the sample size is too small to make valid generalizable conclusions. A future study that replicates this study's experimental procedure with a larger sample size would be effective in assessing whether the conclusions made in this paper are valid. It would also be optimal to elicit productions using prompts with more sonorants (/l/, /m/, /n/, etc.). It would also be interesting to run a perception study in which participants listen to the elicited responses from the autistic and neurotypical subjects and answer whether the speaker was asked a question or given a statement as a prompt. That is, can listeners infer what the previous discourse was from the statement made? This study can also include questions on whether the participant thinks that the subject whose response they heard is autistic or neurotypical. Such a follow-up study would help pinpoint what elements of autistic prosody may be apparent to the listener and contribute to a perception of "abnormal" speech.

Finally, we hope that the current results contribute to behavioral skills endeavors for autistic students at Rutgers that are a part of the College Support Program. All three of the participants in this study volunteered because they felt that their speech was different from neurotypical people. For autistic students such as those who perceive that their prosody is atypical, and want support towards better social conversations, results from this study can help create behavioral skills interventions. One suggestion is that students could listen to their own speech recordings from this study, as well as recordings from neurotypical people, and see if they can perceive differences. If they do, and when asked, say they want to work on these differences, they can take a look at the pitch contours in Praat and see the various instances

where they are doing something different than neurotypical people. Further analysis and further steps in the study can then create targeted solutions.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing**.**

Baltaxe, C. A., & Simmons, J. Q. (1985). Prosodic development in normal and autistic children. *Communication Problems in Autism*, 95–125. https://doi.org/10.1007/978-1-4757-4806-2_7

Barth-Weingarten, Dehé, N., & Wichmann, A. (2009). *Where prosody meets pragmatics* (First edition.). Emerald.

Baron-Cohen, S., and Staunton, R. (1994). Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language* 14, 241–248.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Beckman, M.E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3, 15–70.

Berman, R., & Slobin, D.. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Boersma, P., and Weenink, D. (2009). Praat: Doing Phonetics by Computer [Computer Program]. Version 5.1.07. Available at: http://www.praat.org/

Chilton, G. (2012, May 18). *"I didn't say we should kill him!" – intonation, sentence stress and murder.* designer lessons. Retrieved April 1, 2022, from https://designerlessons.org/2011/12/10/intonation-sentence-stress-esl/

Clark, H. H. (1977). Comprehension and the given-new contract. In S. E. Haviland (Ed.), *Discourse Production and Comprehension* (Vol. 1, pp. 1–21). Ablex Publishing Corporation.

DePape, A.-M. R., Chen, A., Hall, G. B., & Trainor, L. J. (2012). Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00072

Eigsti, I.-M., Schuh, J., Mencl, E., Schultz, R. T., & Paul, R. (2012). The neural underpinnings of prosody in autism. *Child Neuropsychology*, *18*(6), 600–617. https://doi.org/10.1080/09297049.2011.639757

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*, *32*(5), 771–782. https://doi.org/10.1111/j.1469-7610.1991.tb01901.x

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult … but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315. https://doi.org/10.1080/17549500903003060

Gussenhoven, C. (2004). The phonology of tone and intonation . Cambridge University Press.

Halliday, M. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics, 3*(2), 199-244. doi:10.1017/S0022226700016613

Harrington, J. (2020, March 12). *Intonation - tobi introduction*. Macquarie University. Retrieved from https://www.mq.edu.au/about/about-the-university/our-faculties/medicine-and-health-sciences/departments-and-centres/department-of-linguistics/our-research/phonetics-and-phonology/speech/phonetics-and-phonology/Intonation-tobi-introduction

Kim, J.-Y. (2019). Heritage speakers' use of prosodic strategies in focus marking in Spanish. International Journal of Bilingualism, 23(5), 986–1004. https://doi.org/10.1177/1367006918763139

Kirkwood, H. M. A. (1967). *Intonation and grammar in British English*. Mouton.

Ladd. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press.

Lambrecht, & Michaelis, L. (1998). Sentence Accent in Information Questions: Default and Projection. Linguistics and Philosophy, 21(5), 477–544. https://doi.org/10.1023/A:1005327212709

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

Meltzoff, A., Gopnik, A., Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. (1993). Understanding other minds: Perspectives from autism.

Patel, S. P., Kim, J. H., Larson, C. R., & Losh, M. (2019). Mechanisms of voice control related to prosody in autism spectrum disorder and first-degree relatives. *Autism Research*, *12*(8), 1192–1210. https://doi.org/10.1002/aur.2156

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*, *38*(10), 1776–1791. https://doi.org/10.1016/j.pragma.2005.07.004

Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT. Reproduced by Indiana University Linguistics Club, Bloomington.

Rapin, Trudeau-Fisette, P., Bellavance-Courtemanche, M., & Ménard, L. (2015). Production of contrastive focus in children with autistic spectrum disorder. *The Journal of the Acoustical Society of America*, *137*(4), 2431–2431. https://doi.org/10.1121/1.4920869

Syrett, K., & Goldin, M. (2019). Duck and friends: Videos designed for elicited narration studies. Rutgers University Laboratory for Developmental Language Studies.

Thurber, C., & Tager-Flusberg, H. (1993). Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism and Developmental Disorders*, *23*(2), 309–322. https://doi.org/10.1007/bf01046222

Williamson, G. (2018, April 30). *Glottalization*. SLT info. Retrieved from https://www.sltinfo.com/allo101-glottalization/

Appendix A: **Instructions in the Elicited Narration Task**

Instructions for the Participant

In this study, you will see a series of short videos involving some animals. In each video, some events are going to happen, separated by some very brief black screens.

You'll be watching these videos alongside of and in sync with another person, who is your partner. I'm going to give you each a set of instructions. Please listen carefully to your set. Please take your masks off when we start the experiment.

Your job will be to narrate these videos as the actions unfold. You will be recorded. Your partner will be watching a similar video. However, they cannot hear you narrate the event, because they will be wearing noise-canceling headphones.

In your partner's version, certain visual aspects of the events are blocked out. For example, they may not see *which animal* is performing an action or *what action* is being performed.

The videos are brief, and within the videos, there are also segments separated by black screens. During the brief black screens, your partner is going to briefly remove their headphones. They are going to ask you some questions and make some statements to make sure they understand what is going on.

Your job is to watch each video carefully, and help your partner understand what is happening, even though they don't see everything. When they ask a question, please answer it. When they

make a statement, please let them know whether they are right or wrong. No matter what, always answer in complete sentences.

Once you have finished this brief exchange with your partner and before the next scene begins, your partner will put their headphones back on, and try to follow along. When the next black screen appears, they will do the same thing again: remove their headphones, ask a question and/or make a statement, and then replace their headphones and continue watching.

This will continue for each of the subevents for the 5 main videos. Remember, your job is to narrate the videos, and help make sure your partner understands what happened. Always use complete sentences.

You can take a break at any time.

Instructions for the Confederate

Your job will be to watch a video similar to the one your partner is viewing. They will be narrating the video and will be recorded doing so, but you will not be able to hear them, because you will be wearing noise-canceling headphones.

Your partner's version of the video is complete and uninterrupted. In your version, however, certain visual aspects of the events will be obscured. For example, you may not be able to see *which animal* is performing an action or see entirely *what action* is being performed.

The videos are brief, and within the videos, there are also segments separated by black screens. During these brief black screens, your job is to briefly remove your headphones and either ask questions or make statements to make sure you understand what is going on.

Your partner should let you know whether you are right or wrong. If they do not respond in a complete sentence, please encourage them to do so.

Once you have finished this brief exchange with your partner and before the next scene begins, please put your headphones back on, and try to follow along. When the next black screen appears, do the same thing again: remove your headphones, ask a question and/or make a statement, and then replace your headphones and continue watching.

This will continue for each of the subevents for the 5 main videos. Remember, your job is to watch, and ask questions or make statements to make sure you understand is happening in the videos.

Appendix B: Prompts from Confederate to Participant in the Elicited Narration Task

For each subevent, you will elicit an answer, an affirmation, or a correction with one or more of these prompts. An anticipated response is in [ ] following your prompt.

Q: information-seeking questions (balancing subject and object *wh*-questions)

A^c: correct assertions prompting affirmation

A_x: incorrect assertions prompting correction

Video 1 – Duck is hungry

1

Q: Why was the Duck upset?

[The Duck was hungry.]

2

Q: Who did the Duck call?

[The Duck called the Pig.]

A_x: The Pig put on a hat.

[No, the Pig put on a scarf.]

3

Q: Who visited the Duck?

[The Pig visited the Duck.]

A^c: The Pig gave the Duck some food.

[Yes, the Pig gave the Duck some food.]

4

A^c: The Duck and the Pig hugged each other.

[Yes, the Duck and the Pig hugged each other.]

A_x: The Duck left the room.

[No, the Pig left the room.]

Video 2 – Duck is overwhelmed

1

Q: Who was upset?

[The Duck was upset.]

2

A_x: The Duck called the Pig.

[No, the Duck called the Frog.]

Who picked up a bag?

[The Frog picked up a bag.]

3

A^c: The Duck opened the door.

[Yes, the Duck opened the door.]

A_x: The Duck put the ball in the bag.

[No, the Frog put the ball in the bag.]

4

A^c: The Duck and the Frog cleaned up the mess.

[Yes, the Duck and the Frog cleaned up the mess.]

Q: Who took the bag?

[The Frog took the bag.]

Video 3 – Duck Is lonely

1

A^c: The Duck is alone.

[Yes, the Duck is alone.]

Q: How do you think the Duck is feeling?

[The Duck is sad/lonely/upset.]

2

Q: Who did the Duck call?

[The Duck called the Dog.]

3

A^c: The Dog put on a scarf.

[Yes, the Dog put on a scarf.]

A_x: The Dog picked up some ice cream.

[No, the Dog picked up some cupcakes.]

4

A^c: The Dog brought the Duck cupcakes.

[Yes, the Dog brought the Duck cupcakes.]

A_x: The Duck ate all of the cupcakes.

[No, the Duck ate one of the cupcakes.]

5

A^c: The Dog and the Duck hugged good-bye.

[Yes, the Dog and the Duck hugged good-bye.]

Q: Who left the room?

[The Dog left the room.]

Video 4 – Duck is bored

1

Q: Who is bored?

[The Duck is bored.]

2

A_x: The Cow answered the phone.

[No, the Frog answered the phone.]

A^c: All of the animals have a ball.

[Yes, all of the animals have a ball.]

3

A_x: Two of the animals put on a scarf.

[No, ALL/THREE of the animals put on a scarf.]

4

A^c: The Duck is happy to see the other animals.

[Yes, the Duck is happy to see the other animals.]

A_x: All of the animals have a ball.

[No, all of the animals do not have a ball/The Duck does not have a ball.]

5

Q: Who was tossing the ball back and forth?

[The Pig and the Frog were tossing the ball back and forth?]

6

Q: Who did the Duck give the ball to?

[The Duck gave the ball to the Cow.]

A_x: The Cow left first.

[No, the Frog left first / The Cow left last.]

Video 5 – Duck is invited to Frog's birthday

1

Q: Who is at the birthday party?

[The Cow, the Pig, and the Frog are at the birthday party.]

A_x: The Cow picked up the phone.

[No, the Frog picked up the phone.]

2

A_x: Who did the Frog call?

[The Frog called the Duck.]

A_x: The Duck put on gloves.

[No, the Duck put on a scarf.]

3

Q: Who opened the door?

[The Frog opened the door.]

A^c: The Frog hugged the Duck.

[Yes, the Frog hugged the Duck.]

4

A_x: The Cow left the party first.

[No, the Duck left the party first.]