Acoustic Analysis of Prosody in Spontaneous Productions of Minimally Verbal Children and Adolescents with Autism

Jill C. Thorson, Nicole Usher, Rupal Patel, and Helen Tager-Flusberg

1. Introduction

Variations in prosody convey lexical, grammatical, and pragmatic meaning, all essential for successful communication. Individuals with autism spectrum disorder (ASD) show deficits in communication and pragmatic use of language (Tager-Flusberg, Paul, & Lord, 2005), with mixed findings on how stress, intonation, and phrasing distinctions are employed (Paul, Augustyn, Klin, & Volkmar, 2005; McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007; Shriberg, Paul, McSweeny, Klin, Cohen, & Volkmar, 2001). Critically, research to date has focused primarily on individuals with autism who are verbal and have minimal cognitive impairment (McCann & Peppé, 2003). In addition, only a few studies have included fine-grained acoustic analyses to examine more closely the nature of expressive impairments (e.g., Baltaxe, 1984; Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009). Due to challenges in collecting data from a population with a range of behavioral and intellectual issues (Boucher, Bigham, Mayes, & Muskett, 2008), research on this group has been limited. The current investigation was designed to elucidate the prosodic abilities of minimally verbal school-aged children with ASD, a previously understudied population, in an effort to better understand communication abilities across the autism spectrum.

Relatively little research has focused on the specific acoustic properties of the speech of individuals with ASD. A study by Baltaxe (1984) compared the non-echolalic speech of children (aged 4-12) with ASD with the speech of children with language impairment as well as a control group. Interestingly, Baltaxe found that the group with ASD and the control group exhibited wider fundamental frequency (F_0) ranges than the group of children with language impairment. A second study by Jun and Fosnot (1999) found that children with ASD also showed greater F_0 variation and range than age-matched typical

---

[*] First three authors' affiliation: Northeastern University, Department of Communication Sciences and Disorders. Fourth author affiliation: Boston University, Department of Psychological and Brain Sciences. Correspondence email: j.thorson@neu.edu. Special thanks to the participants, their families, Steven Meyer, Laura Haenchen, Sarah Whalen, Brady Eggleston, Daniela Plesa-Skwerer, the CadLab at Northeastern, and the Center for Autism Research Excellence (CARE) at Boston University. Work supported by an NIH ACE grant (NIH P50 DC 13027, Dr. Helen Tager-Flusberg, PI).

controls. Neither of these studies specified the verbal, language, or intellectual abilities of the participants with ASD.

A more recent investigation by Diehl et al. (2009) included a detailed analysis of the acoustic properties of prosody in individuals with ASD. Their study controlled for language ability level and examined the speech of children and adolescents with an ASD diagnosis. They conducted two studies in which the first focused on an older group of participants aged 10 to 18 years old. All of the participants were verbal and had normal intellectual and language abilities. The authors asked how fundamental frequency (F_0) varied between verbal children with ASD and an age, gender, intellectual and language matched control group. Their primary measurement was the average standard deviation of F_0 variance, which was calculated from the pitch of a read narrative passage. Study 1 found that verbal children with ASD spoke with increased F_0 variation compared to the control group. Study 2 replicated the first but with a younger group of children aged 6 to 14 years old who had a lower IQ scores than the group in study 1. In addition, study 1 (but not study 2) revealed that their objective measure of prosody was related to clinical judgments of autism-specific communication impairments.

The current study differs from previous research in two important respects. First, this study focuses on children with an ASD diagnosis who are minimally verbal, a range of the autism spectrum that has been considerably understudied to date. Second, an effort was made to analyze a more ecologically valid sample of speech. Instead of read or elicited utterances, we analyzed spontaneous, non-echolalic utterances that were segmented from a standardized interview assessment (as part of a larger study). These two differences broaden the research on the expressive prosodic abilities of individuals with ASD and begin to fill in some of the gaps to better understand how prosody differs along the autism spectrum.

The aims of this study are two-fold. First, we examine the quality and quantity of spontaneous speech data spoken by a group of minimally verbal children and adolescents with ASD. The aim is to understand differences in verbal ability across the current group of participants. Second, we investigate the nature of F_0 variation in the spontaneous speech of this population, and then compare it to previously published data of verbal children with ASD and a group of control participants (Diehl et al., 2009). The objective is to determine if particular acoustic and prosodic features can predict the severity of diagnosis.

Based on earlier research, we predict that differences in the varied use of F_0 will increase as a result of diagnosis severity (Diehl et al., 2009; Bonneh et al., 2011). Due to the heterogeneity of children who are diagnosed with ASD and who are minimally verbal, a substantial amount of variation is also expected within and across participants. The primary objective of this study is to acoustically analyze the spontaneous speech productions of minimally verbal children with ASD in order to better identify and understand natural prosodic features of this population.

2. Method

2.1. Participants

Data were analyzed for ten American English-speaking children and adolescents (1 female). The age of participants ranged from 5 to 18 years, with a mean age of 9 years. All participants were from Boston, Massachusetts, and surrounding areas. All participants had a diagnosis of autism spectrum disorder of which nine were identified as minimally verbal and one as verbal. One verbal participant was included to provide a baseline of comparison for the minimally verbal participants. Participants are part of a larger research program investigating language, cognitive, and neural abilities of this population.

2.2. Materials

For each participant, spontaneous speech was extracted from the recordings of a standardized interview within the assessment session of a larger study. The assessment instrument from which the speech sample was collected was either the ADOS (Modules 1 or 4) or the Adapted ADOS (Autism Diagnostic Observation Schedule, Second Edition, Lord, Rutter, DiLavore, Risi, Gotham & Bishop, 2012; Adapted ADOS, Hus, Maye, Harvey, Guthrie, Liang & Lord, 2011). Each segmented ADOS session was about two hours in length. All intelligible, non-imitative utterances were extracted and analyzed. Excluded samples included imitations of the interviewer, singing, and non-words. Speech samples varied in length from one to fifteen words across participants.

The independent variables for each participant are age, gender, scores on the Peabody Picture Vocabulary Test (PPVT-4), and receptive and expressive language scores based on in-house vocabulary checklist of 350 words (Dunn & Dunn, 2007). Additional information collected from the analyzable speech included total speaking duration (out of the two hour assessment sample), average and standard deviation of speech run duration (in seconds), number of total utterances, and number of unique lexical items. Utterances and unique lexical items were each calculated based on the full analyzable speech sample for each participant. An utterance was operationally defined as a continuous speech run with no interruptions or large pauses. For the purposes of this study, a unique lexical item was defined as a lexical word. Different inflections or derivations of the same word root were not considered uniquely different from one another and were counted as one unique lexical item (e.g., dog/dogs; laugh/laughs/laughing). Age, gender, verbal scores, total duration of analyzable speech, and percent speaking time (out of total two hour assessment duration) are presented in Table 1 for each participant.

Table 1. *Individual particpant information including age (years;months), gender, verbal scores, and receptive and expressive vocabulary based on a checklist of 350 items. Some screening measures could not be collected for all participants. The final column gives total duration of meaningful speech analyzed for this study (in minutes) and the percentage of analyzed speech out of the total session duration for each participant. V = verbal, MV = minimally verbal.*

Partici-pant	Age	Gender	Verbal Level	PPVT – Raw/Standard Scores	Vocab Receptive	Vocab Expressive	Speech Duration (mins) (% of total session)
P1	15;0	Male	V	158/83	-	-	16.6 (15%)
P2	18;4	Male	MV	78/26	350	201	5.7 (8%)
P3	5;7	Male	MV	20/48	135	7	5.9 (4%)
P4	7;5	Male	MV	4/-	-	-	23.1 (19%)
P5	10;3	Female	MV	50/35	270	160	3.4 (3%)
P6	7;7	Male	MV	18/31	48	8	1.5 (1%)
P7	6;0	Male	MV	1/-	29	0	0.13 (0.1%)
P8	7;9	Male	MV	30/42	89	33	0.03 (0.02%)
P9	5;9	Male	MV	-/-	-	-	0.02 (0.02%)
P10	7;7	Male	MV	64/66	334	218	0.01 (0.01%)

2.3. Acoustic Analysis

For the prosodic analysis, the primary acoustic correlate of pitch, the fundamental frequency (F_0), was measured. Following Diehl et al. (2009), the two acoustic variables of interest are *average pitch* and *pitch range*. Using Praat, the average F_0 was extracted from successive 250-ms time windows, resulting in four data points per second (Boersma & Weenink, 2014). These points were taken from all analyzable utterances for each participant. *Average pitch* is the average of these data points and shows basic pitch level differences between participants. A common measurement for *pitch range* is the standard deviation of average F_0. This measure highlights the degree to which speakers utilize intonation during these spontaneous speech productions, with a larger standard deviation signaling a wider pitch range. In order to account for pitch perturbations and tracking errors in the data set, a limited number of the averaged F_0 points from the 250-ms segments were eliminated from the analysis. This did not affect the overall analysis due to the large amount of individual data points.

3. Results

An initial analysis of the speech data yielded the total number of utterances and number of unique lexical items per participant. The total number of utterances varied between 1 and 550, and the number of unique lexical items between 1 and 493. The average length of an utterance in seconds (*average run*) was calculated for each participant by taking the average duration of all analyzable utterances. Table 2 displays how many utterances and unique lexical items were produced by each participant. Average speaking run ranged from 0.4 to 2.8 seconds long (Table 2).

The two acoustic values of *average pitch* and *pitch range* were calculated for each participant. Average F_0 values for each participant are presented in Table 2, with values varying between 122 and 361 Hz. The median and interquartile ranges of the average F_0 values in hertz were additionally calculated (Figure 1). Box and whisker plots display speech sample variability and show median and interquartile ranges of average F_0 (IQR = Q3 – Q1). The IQR covers the central 50% of the data. Participants are presented partly according to their level of verbal ability in Figure 1, with the verbal participant on the left side. The verbal participant (P1) exhibited the lowest median value (121 Hz) and the narrowest IQR (16 Hz) of the participants. Remaining participants (P2 – P10) had higher median values (192 - 345 Hz). Excluding P10, who had only one data point, participants P2 through P9 exhibited larger IQRs with an average of 82 Hz. Participant P8 had the next smallest IQR after P1 with a range of only 21 Hz. The remaining minimally verbal participants (P2 to P7, and P9) had IQRs from 48 to 122 Hz.

Pearson product-moment correlation coefficients were computed to

Table 2. *Number of utterances, unique lexical items, average run duration of an utterance (in seconds), and average F_0 for each participant.*

Participant	Number of Utterances	Number of Unique Lexical Items	Average (stdev) Speaking Run (seconds)	Average F_0 (Hz)
P1	118	493	2.8 (4.5)	312
P2	550	317	3.1 (1.8)	122
P3	154	251	1.4 (4.1)	213
P4	144	182	2.0 (3.5)	361
P5	85	88	2.3 (5.4)	305
P6	31	57	1.6 (1.3)	291
P7	6	6	1.3 (0.4)	348
P8	5	5	0.4 (0.1)	264
P9	2	2	0.5 (0.2)	307
P10	1	1	0.4 (n/a)	299

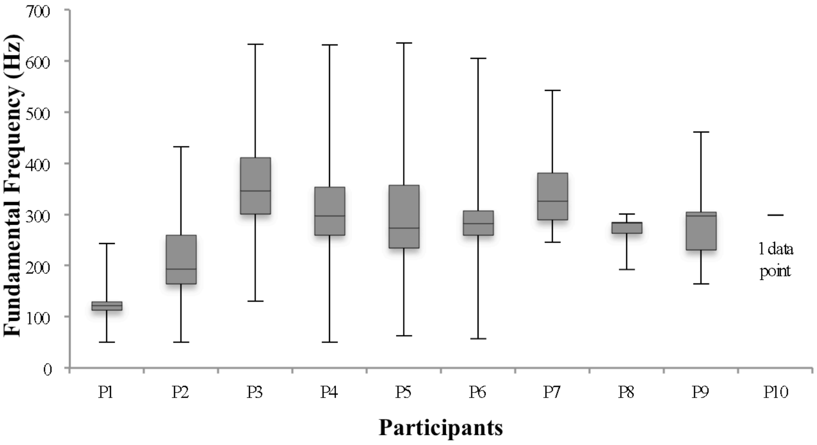

Figure 1. *Box and whisker plots show average fundamental frequency values for each participant. Boxes show median and the inner two quartiles; error bars mark minimum and maximum values. P1 is a verbal participant; P2 to P10 are minimally verbal.*

investigate the relationship between caregiver-provided verbal scores for participants and the number of lexical items in the analyzed spontaneous speech sample. Both expressive and receptive language scores from the in-house vocabulary checklist positively correlated with the number of unique lexical items (r = .86, p = .007; r = .89, p = .008, respectively). Additional correlation coefficients were computed to assess the relationship between the number of

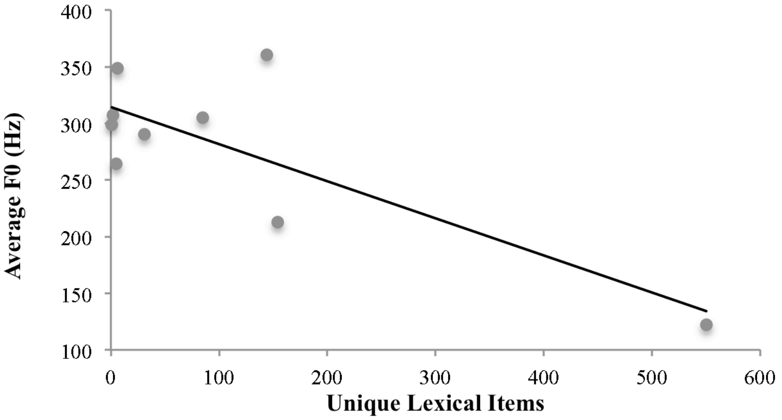

Figure 2. *Scatterplot of average F_0 and number of unique lexical items present in the analyzed speech of each participant. Each dot represents one participant.*

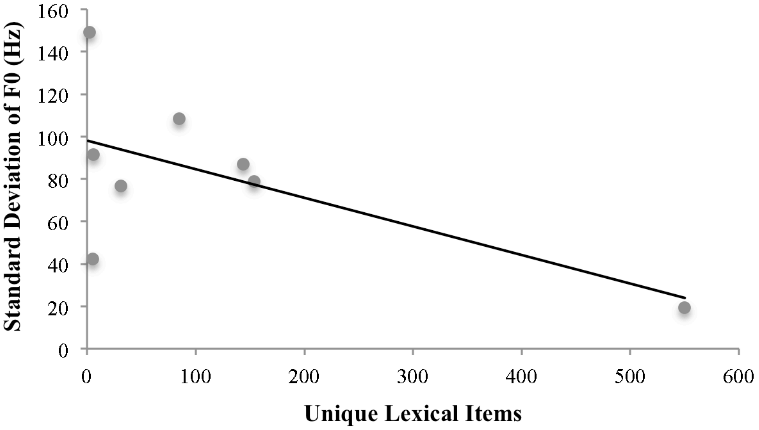

Figure 3. *Scatterplot of the standard deviation of F_0 and number of unique lexical items present in the analyzed speech of each participant. Each dot represents one participant.*

unique lexical items and *average pitch* and *pitch range* scores. There were negative correlations between the number of unique lexical items that each participant produced during their spontaneous speech sample and their average F_0 (*average pitch*) values (r = -.78, p = .008) and their standard deviation in F_0 (*pitch range*) values (r = -.62, p = .07). Two scatterplots summarize the results (Figures 2 and 3).

A group average was computed by taking the average of the standard deviation in F_0 values from each participant (*average standard deviation in F_0 variance*). In order to compare the group's average to previous research, only minimally verbal participants that were age-matched were included. This yielded six minimally verbal participants with ASD between the ages of 5.5 and 11 years old. The current study's minimally verbal group was compared to the verbal participants with ASD and the control participants from Diehl et al. (2009). The verbal and control groups had 17 participants each and ranged in age from 6 to 14 years old.

Preliminary observations of the data show a linear trend in the pitch range usage of these three populations. Figure 4 shows that the minimally verbal group had the highest amount of F_0 variance, while the control children exhibited the least amount. Statistical tests were not performed due to limited data availbility and differences in how the data were collected.

4. Discussion

This study explored how minimally verbal school-aged children with ASD express prosodic information acoustically in spontaneous speech. The prosodic factor of interest was pitch, and specifically the average and standard deviation

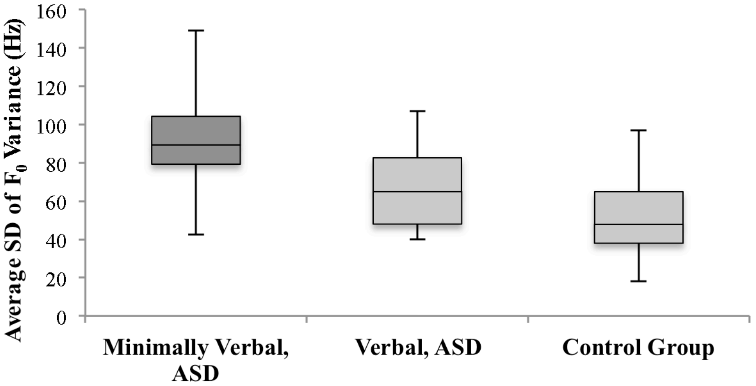

Figure 4. *Box and whisker plots of the average standard deviation of F_0 variance. The leftmost box shows the median and the inner two quartiles; error bars mark minimum and maximum values. The right two boxes show the median and 95% of the range of the standard deviation of F_0. The box on the left shows the data from the age-matched minimally verbal children and adolescents in this study (6 children, 5.5 – 11 y.o.). The two boxes on the right display data from Diehl et al.'s (2009) study (17 children in each group, age 6 – 14 y.o.).*

of the fundamental frequency (F_0) (i.e., *average pitch* and *pitch range*, respectively). Meaningful spontaneous speech samples were segmented and analyzed for their F_0 qualities to better understand how this population regulates pitch during running speech. We studied the quality and quantity of spontaneous speech samples and analyzed the fundamental frequency patterns in comparison to previous research. As predicted, substantial variation was observed across the ten participants for both prosodic and verbal abilities.

The first aim was to capture qualitative and quantitative measures of verbal ability from the spontaneous speech sample. Verbal ability ranged from one participant with ASD who was considered verbal to a participant who only produced one lexical item over the course of the two-hour interview sample. The remaining participants showed substantial variability between those two extremes for both number of utterances and number of unique lexical items produced. When observing the amount of time that a participant spoke overall, the percentage of time spoken varied from less than one percent to almost twenty percent of the two-hour assessment. Also, positive correlations between both the receptive and the expressive in-house vocabulary checklists revealed that the higher the scores on these caregiver questionnaires, the more unique lexical items were produced during the speech sample. Taken together this variation highlights the heterogeneity of this population, and begins to quantify differences among individuals with ASD.

The second aim was to investigate how pitch varied across this population and to test if *average pitch* and *pitch range* varied as a function of verbal ability. Number of unique lexical items was used as the predictor of pitch modulation due to its high correlations with both caregiver questionnaires that report receptive and expressive language abilities. Both *average pitch* and *pitch range* correlated with number of unique lexical items uttered. Pitch average and pitch range both decreased as the number of lexical items increased. Important to note is that these correlations were driven largely by the presence of one or two data points. Future work will include more participants in order to better discover the nature of the relationship between verbal ability and prosodic variance.

In order to compare our data to previous work, individual standard deviations were collapsed to form a single averaged pitch range value for the minimally verbal group with ASD. These data from our minimally verbal group were then compared to previous data from Diehl et al. (2009). As predicted, the data together form a linear trend, with greater pitch ranges found for the least verbal group (Figure 4). Although these data are from two different studies with unique experimental protocols, this early investigation into group differences provides a promising first look at pitch modulation across the autism spectrum.

Due to the administration of several different modules and versions of the ADOS, it was not possible to compare scores across individuals or to make any predictions about how ADOS scores relate to fundamental frequency measures. Although there was a range of versions administered, six of the participants were tested on the same module (ADOS, module 1). No correlations between ADOS scores and pitch were found. ADOS scores were clustered in a small range from 20 to 26, with four of the participants receiving a score of 23. In Diehl et al.'s (2009) study, a relationship between ADOS scores and F_0 variance was found, but only in one of their two studies. Future research that includes data from individuals with a variety of scores on the ADOS is necessary in order to further investigate the relationship between pitch and autism-specific assessment strategies.

5. Conclusion

The current study serves as an initial exploration into the prosodic patterns of pitch in minimally verbal school-aged children with ASD and provides further insight into how prosody varies along the autism spectrum. Future work will expand the participant pool to better understand group differences. Additionally, data analyses of other acoustic correlates of prosody, including duration and intensity are currently underway. A better understanding of prosodic characteristics in ASD has implications for early intervention strategies and provides a baseline for melodic-based therapies. A corollary study to this project is devoted to the creation of a prosodic assessment that evaluates receptive and expressive prosodic abilities in individuals who are minimally verbal with ASD. This work is part of a larger study that aims to more fully understand the cognitive, linguistic, neurological, and genetic underpinnings of minimally verbal children and adolescents with ASD.

References

Baltaxe, Christiane. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research, 27,* 97-105.

Boersma, Paul, & Weenink, David. (2014). *Praat: A system for doing phonetics by computer* [Software]. Available from http://www.praat.org

Bonneh, Yoram. S., Levanon, Yoram, Dean-Pardo, Omrit, Lossos, Lan, & Adini, Yael. (2011). Abnormal pitch spectrum and increased pitch variability in autism spectrum. *Frontiers in Human Neuroscience, 4*(237), 1-7.

Boucher, Jill, Bigham, Sally, Mayes, Andrew, & Muskett, Tom. (2008). Recognition and language in low functioning autism. *Journal of Autism and Developmental Disorders, 38*, 1259-1269.

Diehl, Joshua J., Watson, Duane, Bennetto, Loisa, McDonough, Joyce, & Gunlogson, Christine. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics, 30*, 385-404.

Dunn, Lloyd M., & Dunn, Douglas M. (2007). *Peabody Picture Vocabulary Test (4th ed.).* Minneapolis, MN: Pearson Assessments.

Fosnot, Suzi, & Jun, Sun-Ah. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. Paper presented at the *14th International Congress of Phonetic Sciences (ICPhS).*

Hus, Vanessa, Maye, Melissa, Harvey, L, Guthrie, Whitney, Liang, J., & Lord, Catherine. (2011). The Adapted ADOS - Preliminary findings using a modified version of the ADOS for adults who are nonverbal or have limited language. Poster presented at the *International Meeting for Autism Research*, San Diego, CA.

Lord, Catherine, Rutter, Michael, DiLavore, Pamela C., Risi, Susan, Gotham, Katherine, & Bishop, Somer. (2012) *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1-4*. Torrance, CA: Western Psychological Services.

McCann, Joanne, & Peppé, Susan. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders, 42,* 682-702.

McCann, Joanne, Peppé, Susan, Gibbon, Fiona E., O'Hare, Anne, & Rutherford, Marion. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language and Communication Disorders, 42,* 682-702.

Paul, Rhea, Augustyn, Amy, Klin, Ami, & Volkmar, Fred R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35,* 205-220.

Shriberg, Lawrence D., Paul, Rhea, McSweeny, Jane L., Klin, Ami, Cohen, Donald J., & Volkmar, Fred R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44,* 1097-1115.

Tager-Flusberg, Helen, Paul, Rhea, & Lord, Catherine. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. J. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Diagnosis, development, neurobiology, and behavior* (Vol. I, pp. 335-364). Hoboken, NJ: Wiley.