Identifying stable speech-language markers of autism in children: Preliminary evidence from a longitudinal telephony-based study

Abstract

This study examined differences in linguistic features produced by autistic and neurotypical (NT) children during brief picture descriptions, and assessed feature stability over time. Weekly speech samples from well-characterized participants were collected using a telephony system designed to improve access for geographically isolated and historically marginalized communities. Results showed stable group differences in certain acoustic features, some of which may potentially serve as key outcome measures in future treatment studies. These results highlight the importance of eliciting semi-structured speech samples in a variety of contexts over time, and adds to a growing body of research showing that fine-grained naturalistic communication features hold promise for intervention research.

1 Introduction

Natural sampling is a rich approach to investigating speech and language in autistic children. Previous studies have shown that language behavior in autism differs from neurotypical (NT) patterns in a number of ways. For example, autistic children who are more severely impacted have been shown to produce less speech, [1] slower speech, [2,3] and speech with atypical voice quality[1] compared to NT peers. It has also been observed that autistic children's prosody differs from NT children, with pitch descriptions ranging from sing-songy to monotonous. [3] In the lexical domain, prior research has shown that autistic children use nouns and cognitive words differently than NT peers when narrating a story from a picture, [4] use different patterns of filler words during clinical assessments, [5] and talk less about social topics during get-to-know-you conversations. [6] Research in this domain continues to emerge, but samples remain small and results occasionally conflict or fail to replicate.

Prior studies of natural language in autism used a variety of data collection and analysis methods that could critically affect results and may have led to conflicting findings. For example, the presence of an unfamiliar adult during in-person or remote elicitations could adversely impact the behavior of autistic children, thus reducing the quality and informativeness of their language samples. [7] Also, children's linguistic behavior might differ depending on the specifics of the elicitation task in a given study, i.e., whether natural conversations or semi-structured speech tasks are used, and the characteristics of certain elicitation stimuli.

In order to develop scalable, cost-effective, objective intervention progress monitoring systems of autistic symptoms using speech as a primary target, it is necessary to understand how contextual and testing factors affect children's behavior. Then, it will be possible to identify robust features that reliably index autism symptoms across heterogeneous testing conditions. Toward this goal, we developed a telephony protocol to examine how various factors affect speech performance in autistic children and adolescents. Telephony has particular potential to address service and monitoring gaps for autistic and NT children from historically marginalized and/or low-resource communities, and is a useful alternative to in-person data collection during the COVID-19 pandemic. The final battery of our protocol consisted of seven versions of seven tasks that a parent or legal guardian could independently facilitate. In this preliminary report from an on-going study, we assessed children's speech and language features during one of the seven tasks (picture descriptions) collected in the first and second phone sessions. Our goals were to (1) identify diagnostic group differences in automated speech and language features that are stable over time, and (2) examine potential effects of staff vs. parent administration in each diagnostic group.

2 Methods

2.1 Participants

Study inclusion and exclusion criteria are included in the Appendix. In this report, we analyzed data from 29 children who successfully completed two sessions. Participant groups were matched on age, full-scale IQ, and self-reported race (Table 1). Groups were not matched on sex (p=0.015), which is expected due to the prevalence of ASD in boys, [11] and we are currently addressing with targeted recruitment. One autistic participant identified as non-binary. Autism and NT groups differed in several clinical ratings (Table 1).

	Autism (n=13)	**NT (n=16)**	**p-value**
Age (years)	9.8 (2.5)	9.6 (2.6)	0.767
Sex (%)	10 boys (76.9%)	6 boys (37.5%)	0.015
Full scale IQ	115.1 (15.4)	119.1 (13.7)	0.469
Race (%)	4 non-whites (30.8%)	5 non-whites (31.3%)	0.69
SCQ (total)	17.0 (6.6)	1.2 (1.1)	<0.001
SRS-2 (total)	70.5 (7)	42.1 (3.5)	<0.001
CCC-2 (speech)	9.2 (2.5)	11.8 (0.8)	<0.001
CCC-2 (non-verbal)	5.5 (2.2)	11.8 (1.3)	<0.001

Table 1: Demographic and clinical characteristics of the participants. SCQ: Social communication questionnaire, [8] SRS: Social responsiveness scale, [9] CCC: Children's communication checklist. [10]

2.2 Data collection and annotation

We developed a telephony platform to support single and dual speaker modes. This platform consisted of a high-availability server, voice over internet protocol (VoIP) service by Vonage, telephony software framework (Asterisk 13.18.3), a relational database, and telephony applications.

Prior to the first official data collection call, study staff held an "informational call" with the participating parent to review standard elicitation methods to be utilized across sessions. During the first session with the child, study staff remained on the line and facilitated tasks with the parent and child. During the second session, children and parents independently completed all seven tasks on their own. Children described different pictures during the first and second sessions, and the second session was collected approximately one week after the first session was completed.

Recordings were transcribed by trained annotators using a web-based transcription tool with a built-in speech activity detector (SAD) function. For dual speaker mode recordings, SAD ran on each channel separately. Annotators also corrected speech segment boundary errors.

2.3 Acoustic and text features

Words were automatically tagged for part-of-speech (POS) categories using spaCy. [12] POS categories, fillers, partial words, repetitions, and "hm" were counted separately and converted to counts per 100 words. Content words were rated for word frequency, [13] concreteness, [14] ambiguity, [15] age of acquisition (AoA), [16] and familiarity. [16] We also ran the Language Inquiry and Word Count program [17] to calculate additional word-level measures found to be useful in clinical populations.

For acoustic processing, stereo recordings were split into single channels for precise audio processing. We extracted low-level descriptors of pitch, jitter, shimmer, harmonic-to-noise ratio (HNR), and four spectral moments (1st order: centroid, 2nd order: standard deviation, 3rd order: skewness, 4th order: kurtosis) from participants' picture descriptions per 10 ms using openSMILE with the ComParE13 configuration file. [18] Pitch values in hertz were converted to semitones (st) using individuals' 10th percentiles to normalize physiological differences among participants ($St = \log_2(f0 / 10^{th}\ \text{percentile}) \times 12$). Several durational measures were computed from SAD timestamps.

2.4 Statistical considerations

Preliminary analyses revealed that our variable distributions met the assumptions of parametric tests, so we employed analysis of covariance models. Speech/language features were included as dependent variables, with group, session, and the interaction of group and session as independent variables. Sex was covaried in all models.

3 Results

3.1 Acoustic measures

Median shimmer and jitter values were higher for autistic children than NT children (shimmer: $F(1,52)=4.17$, $p=0.046$; jitter: $F(1,52)= 3.96$, $p= 0.052$, Figure A-B). Mean, standard deviation (SD), and interquartile range (IQR) of jitter and shimmer did not differ by group. Autistic children also had higher mean (skewness: $F(1,52)=13.46$, $p<0.001$; kurtosis: $F(1,52)=12.98$, $p<0.001$), median (skewness: $F(1,52)=6.17$, $p=0.016$; kurtosis: $F(1, 52)=4.7$, $p=0.035$, Figure C-D), SD (skewness: $F(1,52)=9.89$, $p=0.003$; kurtosis: $F(1, 52)=13.86$, $p<0.001$), and IQR values (skewness: $F(1,52)=7$, $p=0.011$; kurtosis: $F(1,52)=8.26$, $p= 0.006$) of spectral skewness and kurtosis than NT children. Groups did not differ in pitch and HNR, and Session had no significant effect on any acoustic variables.

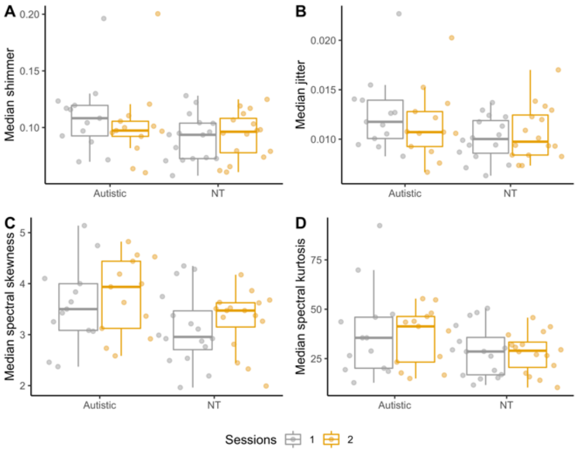

Figure 1: Acoustic features during picture description tasks.

3.2 Durational measures

Autistic children produced longer ($F(1,52)=7.79$, $p=0.007$) and more variable ($F(1,52)= 8.49$, $p=0.005$) speech segment durations than NT children (Figure A-B). The difference in total speech duration between the first and second sessions was larger for autistic children than NT children ($F(1,52)=4.34$, $p=0.042$). Total pause duration was shorter in autistic participants than NT children ($F(1,52)=5.14$, $p=0.028$, Figure C-D), and children paused longer during the first session compared to the second ($F(1,52)=4.82$, $p=0.033$). Autistic children paused less frequently than NT children ($F(1,52)=6.33$, $p=0.015$).

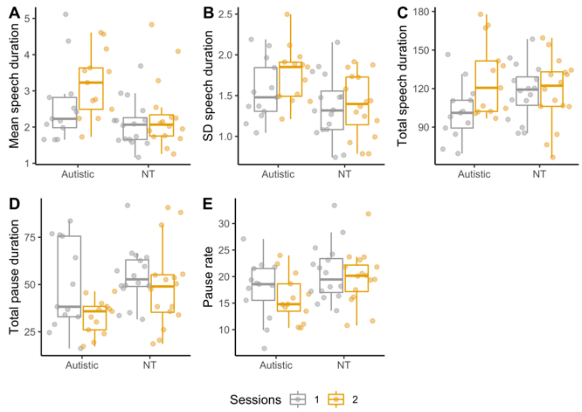

Figure 2: Durational measures during picture descriptions. The units of the y-axis are seconds, except the pause rate, where pause rate per minute was plotted.

3.3 Textual measures

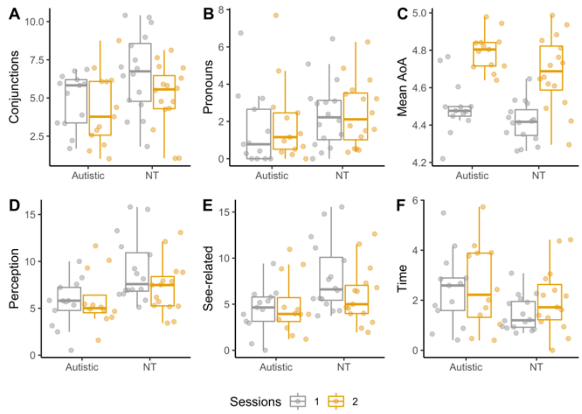

Figure 1: Lexical measures during picture description tasks. All POS counts are per 100 words, and mean AoA was averaged across all content words. LIWC categories were normalized.

Autistic participants produced fewer conjunctions ($F(1,52)=5.06$, $p=0.029$) and pronouns ($F(1,52)= 4.75$, $p=0.034$) than NT children, and their content words had a higher AoA than those of NT children ($F(1,52)=6.35$, $p=0.015$, Figure 1A-C). Also, autistic children produced fewer perception ($F(1, 52)=9.17$, $p=0.004$) and see-related words ($F(1,52) =7.1$, $p=0.01$) and more time-related words ($F(1, 52)=4.79$, $p=0.033$) than NT children (Figure 1).

Regardless of diagnostic status, children produced more adverbs ($F(1,52)=9.08$, $p=0.003$) and prepositions ($F(1,52)=6.47$, $p=0.014$) during the second session than the first (not shown in the figure). Children also produced content words that were more ambiguous ($F(1,52)=10.82$, $p=0.002$), later acquired ($F(1,52)=54.9$, $p<0.001$), and familiar ($F(1,52)=14.85$, $p<0.001$) during the second session than the first session. Finally, several LIWC categories, including anger ($F(1,52)=4.69$, $p=0.035$), difference ($F(1,52)=5.55$, $p=0.023$), feeling ($F(1,52)=4.06$, $p=0.049$), bio ($F(1,52)=4.99$, $p=0.03$), and ingestion ($F(1,52)=19$, $p<0.001$), showed significant effects of Session.

4 Discussion

In this study, we elicited picture descriptions from autistic and NT children using a telephony platform, and tested for the presence of diagnostic group differences in a variety of acoustic and lexical features over two sessions. Results showed that autistic children produced greater local jitter, shimmer and the third and fourth orders of spectral moments, as well as shorter and less frequent pauses compared to NT children, across two sessions and with different stimuli. Autistic children produced more speech during the second session when parents administered the task without study staff, compared to the first session, while NT children's speech duration did not differ by session. Lexically, autistic children produced fewer conjunctions and pronouns than NT children, and used later-acquired content words compared to NT peers. Our results also showed that autistic children used fewer see- or perception-related words and more time-related words than NT children. However, many other lexical features differed by session without group differences, suggesting that the picture stimuli may have had more influence than diagnostic group on lexical production.

Given that the acoustic features described here remained stable from the first to the second telephony session, and also distinguished the groups, they might hold potential as reliable speech markers of autism. High jitter (variability in pitch) and shimmer (variability in intensity) are perceived as harsh, hoarse, or breathy voice. [19] The observation that autistic children's jitter and jitter variability were higher than NT peers is consistent with prior research that also showed positive correlations between jitter and autism symptomology. [1] However, prior research also found lower HNR values for autistic children compared to NT peers, with no significant differences in shimmer; this differs from our pattern of results. Spectral moments in autism have rarely been studied, even though these measures are known to characterize individuals' voice timbre. [20] We plan to study these features further in a larger sample, to explore whether they could serve as validated speech markers of autism.

Autistic participants spoke longer and paused less frequently during the second session than the first session, whereas NT children's duration measures did not differ by session. This might be because autistic individuals experience social-communicative challenges which might have hindered their willingness to speak freely in the presence of unfamiliar study staff. In this case, they may have spoken longer in the second session because their parent administered the task. Thus, it is important to consider the presence of study staff when interpreting studies of speech in autism.

Finally, our study also found that autistic children produced fewer conjunctions, pronouns, see- and perception-related words with high AoA than NT children. We also observed that many word-level features differed by session in both the autistic and NT groups, suggesting that picture selection has an outsized effect on lexical features. In this study, we selected seven different pictures to prevent boredom and practice effects across multiple sessions. However, since different pictures include unique objects that children are likely to list in their descriptions, this will result in significant session-based differences in word-level features. As data collection continues in the current study, we will investigate whether group differences in more abstract lexical features (e.g., pronoun use) might remain stable across sessions.

5 Conclusion

Telephony carries great potential as a low-cost and scalable platform for monitoring intervention responses from afar, as well as measuring longitudinal developmental changes in individual children. Acoustic features extracted from data collected using a telephony system, which delivered consistent, high-quality recordings, could be important tools for identifying speech markers of autism.

Acknowledgments

We thank the children and parents who participated in this study. This study was funded by Roche, Ltd (PI: Parish-Morris), and R01DC018289 (PI: Parish-Morris).

References

Daniel Bone, Chi-Chun Lee, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan. 2014. The Psychologist as an interlocutor in Autism Spectrum Disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research*, 57(4):1162–1177.

Julia Parish-Morris, Mark Liberman, Neville Ryant, Christopher Cieri, Leila Bateman, Emily Ferguson, and Robert T. Schultz. 2016. Exploring autism spectrum disorders using HLT. In *Proceedings of North American Association of Computational Linguistics, Comp Ling and Clin Psych*, pages 74–84.

Bonneh Yoram, Levanon Yoram, Dean-Pardo Omrit, Lossos Lan, and Adini Yael. 2011. Abnormal Speech Spectrum and Increased Pitch Variability in Young Autistic Children. *Frontiers in Human Neuroscience*, 4. https://www.frontiersin.org/article/10.3389/fnhum.2010.00237

Jaclin Boorse, Meredith Cola, Samantha Plate, Lisa Yankowitz, Juhi Pandey, Robert T. Schultz and Julia Parish-Morris. 2019. Linguistic markers of autism in girls: Evidence of a “blended phenotype” during storytelling. *Molecular Autism*, 10:14. https://doi.org/10.1186/s13229-019-0268-2

Julia Parish-Morris, Mark Liberman, Christopher Cieri, John D. Herrington, Benjamin E. Yerys, Leila Bateman, Joseph Donaher, Emily Ferguson, Juhi Pandey, and Robert T. Schultz. 2017. Linguistic camouflage in girls with autism spectrum disorder. *Molecular Autism*, 8:48. https://doi.org/10.1186/s13229-017-0164-6

Amber Song, Meredith Cola, Samantha Plate, Victoria Petrulla, Lisa Yankowitz, Juhi Pandey, Robert T Schultz, and Julia Parish-Morris. 2021. Natrual language markers of social phenotype in girls with autism. *Journal of Child Psychology and Psychiatry*, 62(8): 949-960.

Mihaela Barokova and Helen Tager-Flusberg. 2020. Commentary: Measuring language change through natural language samples. *Journal of autism and developmental disorders*, *50*(7): 2287-2306.

Michael Rutter, Anthony Bailey, and Catherine Lord. 2003. *SCQ: The Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

John N. Constantino. 2011. *Social Responsiveness Scale, Second Edition.* Los Angeles, CA: Western Psychological Services.

Dorothy Bishop. 2006. *Children’s Communication Checklist-2 U.S. Edition.* San Antonio, TX: Psychological Corporation.

Baio J. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ. 2018;67 Available from: https://www.cdc.gov/mmwr/volumes/67/ss/ss6706a1.htm.

Matthew Honnibal and Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In *EMNLP 2015: Conference on empirical methods in natural language processing*, pages 1373-1378.

Mark Brysbaert and Boris New. 2009. Moving beyond Ku cera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4): 977-990.

Mark Brysbaert, Amy B. Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3): 904-911.

Paul Hoffman, Matthew A. Lambon Ralph, and Timothy Rogers. 2013. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3): 718-730.

Mark Brysbaert, Paweł Mandera, Samntha F. McCormick, and Emmanuel Keuleers. 2018. Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, 51(2): 467-479.

Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. 2013. Recent developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. In *Proceedings of ACM Multimedia*, pages 835–838.

Athanasios Tsanas, Max A. Little, Patrick E. McSharry, and Lorraine O. Ramig. 2011. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson’s disease symptom severity. *Journal of the Royal Society*, 8: 842–855.

Alexander Lerch. 2012. *An introduction to audio content analysis: Applications in signal processing and music informatics.* John Wiley & Sons.

A Appendix: Inclusion and Exclusion Criteria

Inclusion criteria for participants were the following:

- Subjects age 6 – 17.99
- English is participant’s first language
- Verbally fluent – language on grade level/consistent with chronological age
- Strongly suspected/confirmed diagnosis of autism or typical development
- Full-scale and verbal IQ > 75
- For autistic children, current SCQ score >= 11
- For the NT group, current SCQ scores < 11

Exclusion criteria for participants were the following:

- Known genetic condition that impacts neurodevelopment or vocal production/language
- History of persistent language deficits that are currently affecting child’s language abilities such that it impacts their ability to have a conversation
- Extreme prematurity (<32 weeks)
- History of severe neurological injury likely to affect expressive language and communication behavior
- If NT, no first-degree family members with autism
- Plan to begin or change medication during study duration
- Plan to begin or change an intervention during study duration.
- Diagnosis of hearing impairment or cochlear implant