Received: 16 September 2021 | Accepted: 4 April 2022

DOI: 10.1002/aur.2731

RESEARCH ARTICLE

Objective measurement of vocalizations in the assessment of autism spectrum disorder symptoms in preschool age children

Jacquelyn M. Moffitt[1] | Yeojin Amy Ahn[1] | Stephanie Custode[1] | Yudong Tao[2] | Emilin Mathew[1] | Meaghan Parlade[1] | Melissa Hale[1] | Jennifer Durocher[1] | Michael Alessandri[1] | Lynn K. Perry[1] | Daniel S. Messinger[1,2,3]

[1]Department of Psychology, University of Miami, Coral Gables, Florida, USA

[2]Department of Electrical and Computer Engineering, University of Miami, Coral Gables, Florida, USA

[3]Departments of Pediatrics and Music Engineering, University of Miami, Coral Gables, Florida, USA

Correspondence
Jacquelyn M. Moffitt, University of Miami, Coral Gables, FL, USA.
Email: jmoffitt@miami.edu

Funding information
Autism Science Foundation, Grant/Award Number: 18-001; National Institute on Minority Health and Health Disparities, Grant/Award Number: UL1TR002736

Abstract

Assessment of autism spectrum disorder (ASD) relies on expert clinician observation and judgment, but objective measurement tools have the potential to provide additional information on ASD symptom severity. Diagnostic evaluations for ASD typically include the autism diagnostic observation schedule (ADOS-2), a semi-structured assessment composed of a series of social presses. The current study examined associations between concurrent objective features of child vocalizations during the ADOS-2 and examiner-rated autism symptom severity. The sample included 66 children (49 male; $M = 40$ months, $SD = 10.58$) evaluated in a university-based clinic, 61 of whom received an ASD diagnosis. Research reliable administration of the ADOS-2 provided social affect (SA) and restricted and repetitive behavior (RRB) calibrated severity scores (CSS). Audio was recorded from examiner-worn eyeglasses during the ADOS-2 and child and adult speech were differentiated with LENA SP Hub. PRAAT was used to ascertain acoustic features of the audio signal, specifically the mean fundamental vocal frequency (F0) of LENA-identified child speech-like vocalizations (those with phonemic content), child cry vocalizations, and adult speech. Sphinx-4 was employed to estimate child and adult phonological features indexed by the average consonant and vowel count per vocalization. More than a quarter of the variance in ADOS-2 RRB CSS was predicted by the combination of child phoneme count per vocalization and child vocalization F0. Findings indicate that both acoustic and phonological features of child vocalizations are associated with expert clinician ratings of autism symptom severity.

Lay Summary: Determination of the severity of autism spectrum disorder is based in part on expert (but subjective) clinician observations during the ADOS-2. Two characteristics of child vocalizations—a smaller number of speech-like sounds per vocalization and higher pitched vocalizations (including cries)—were associated with greater autism symptom severity. The results suggest that objectively ascertained characteristics of children's vocalizations capture variance in children's restricted and repetitive behaviors that are reflected in clinician severity indices.

KEYWORDS
audio processing, objective measurement, vocalization

INTRODUCTION

Diagnoses of autism spectrum disorder (ASD) and ascertainment of the severity of ASD symptoms are determined behaviorally. The gold standard for ASD assessment includes the autism diagnostic observation schedule, second edition (ADOS-2), a semi-structured, play-based assessment which creates opportunities for observation of social communication skills and restricted and repetitive behaviors (RRB) or interests (Lord & Rutter, 2012). Following administration of the ADOS-2, clinicians rate children on a variety of behaviors in the social affect (SA) and RRB domains. Training in the ADOS-2 is time-intensive, typically involving a multi-day workshop and subsequent centralized reliability assessment. ADOS-2 scoring has similarities with other behavior rating systems, including other ratings of ASD-related behaviors. Obtaining adequate interrater reliability in administration and scoring of the ADOS-2 is challenging in non-research settings (Kamp-Becker et al., 2018; Zander et al., 2016) and likely not frequently assessed.

Researchers have begun using automated measures of behavior to screen for ASD risk and to objectively characterize ASD (Dawson & Sapiro, 2019; Warren et al., 2010). ASD is characterized, in part, by atypical vocal behavior including lower rates of vocalizations, less complex vocalizations, and higher pitched vocalizations (Yankowitz et al., 2019). Many of the defining vocal characteristics associated with ASD lend themselves to objective, automated detection. Attempts to objectively characterize ASD are referred to as digital behavioral phenotyping. Yet, little is known about associations between objectively identified children's vocalizations during diagnostic assessments (like the ADOS-2) and expert clinician rating of the *severity* of ASD symptoms is still unknown. The current study fills this gap by exploring objective approaches to quantifying vocalization characteristics in a community sample of young children identified at-risk for ASD. We examine associations between objective measurements of ASD-related vocal behavior during an ASD assessment and expert clinician characterization of the severity of ASD symptoms.

Digital behavioral phenotyping of ASD

Digital behavioral phenotyping encompasses automated measurement of a variety of objectively detectable ASD-related behaviors. This approach may provide economical and objective measures of ASD-related behaviors such as atypical vocalizations in a range of naturalistic and clinical settings (Dawson & Sapiro, 2019). One digital phenotyping approach is the use of deep learning to optimize linear and nonlinear associations between objectively measured behavioral indicators and ASD severity or diagnosis. Our group, for example, used a deep learning model to detect associations between audio recordings of ADOS-2 administration in a subset of 30 children from the current study sample, predicting approximately 40% of the variance in calibrated severity scores (CSS) (Sadiq et al., 2019). In similar fashion, Eni et al. (2020) utilized deep learning models of the vocal characteristics of the expert-identified child vocalizations of 72 Hebrew-speaking children, finding that a convolutional neural network predicted 67% of the variance in ADOS Total Raw Scores. A difficulty with these deep learning methods is that they are not designed to reveal interpretable associations between behaviors and outcomes. An alternative approach focuses on the detection of interpretable associations between ASD-related behavioral characteristics such as vocalization atypicalities and clinical indices of autism symptoms. Adopting this approach, the current study examined linear associations between acoustic and phonemic qualities of preschool-age child and adult vocalizations during administration of the ADOS-2 and concurrent clinician ratings of autism symptoms.

Vocal behaviors in ASD

Vocal atypicalities and deficits are a component of the ASD profile and an early index of the disorder (Oller et al., 2010; Ramsay et al., 2021; Sheinkopf et al., 2000). There is evidence for deficits in the phonemic complexity of the vocalizations of children with ASD. Toddlers with ASD (18–24 months) produce fewer speech sounds than both their typically developing (TD) peers and those with non-ASD DD (Plumb & Wetherby, 2013). Moreover, children with ASD tend to produce fewer phonemes per utterance (Woynaroski et al., 2017; Xu et al., 2014).

Likewise, unusual vocalization qualities, such as high pitch, are used to diagnose ASD and their severity is rated by clinicians on the ADOS-2 (Yankowitz et al., 2019). Young children with ASD (15–36 months) produce significantly more high-pitched squeals, and appear to produce cries of especially high fundamental frequency (F0; pitch), when compared to their age- and language-matched peers (Esposito et al., 2014; Esposito & Venuti, 2010; Schoen et al., 2011). Bonneh et al. (2011) found that preschool age children with ASD demonstrated a greater range and variance of vocalization F0 than their TD peers. Focusing on individual differences, Eni et al. (2020) found the mean and variance of preschool and school-age child vocalization F0 during the ADOS-2 to be significantly related to both ADOS-2 SA and RRB raw scores.

While investigators have examined diagnostic group differences in vocalizations, there is little research on the association between individual differences in vocalizations and autism symptom severity. Among children 4–9 years of age higher child vocalization pitch variability was associated with higher scores on the repetitive behavior and stereotyped patterns subscale of the Japanese autism screening questionnaire (Nakai et al., 2014). Warren and colleagues (Warren et al., 2010) utilized language environment analysis (LENA) to analyze home audio from LENA recorders worn by the child in a specially-designed vest. Children with ASD produced fewer vocalizations and engaged in fewer adult-child conversational turns than their TD peers. Further, children with ASD who produced fewer conversational turns had higher scores on an ASD screening measure, the modified checklist of autism in toddlers (M-CHAT), which is utilized by community providers to assess the presence or absence of early behavioral indicators of ASD per parent report (e.g., response to name, use of pointing). These findings suggest that objective measurements of children's vocalizations in both clinic contexts and naturalistic settings can quantify meaningful information related to symptom levels on standardized screening measures.

Vocal behavior in social partners of children with ASD

Adults may exhibit changes in their own speech when interacting with children with ASD (Bone et al., 2014; Kumar et al., 2016). The F0 of the vocalizations of parents of young children increase for children at high risk for ASD compared to other parents (Quigley et al., 2016). Machine learning analysis of acoustic features of speech during assessment of verbally fluent children with ASD (age 5–14 years) using the ADOS-2 indicated that examiners exhibited more varied pitch and volume when assessing children with higher ASD symptom levels (Bone et al., 2014). A possibility pertinent to the current investigation is that ADOS-2 examiners may change their behavior in reaction to children's ASD-related deficits. If examiner behavior contains information relevant to the severity of ASD symptoms, automated measurement of both child and clinician speech in an assessment context may be relevant to indexing ASD severity.

Current study

The current study examines objective features of vocalizations recorded during a clinical evaluation of autism symptoms (ADOS-2) in a community sample. Our goal was to quantify associations indexing common variance between objective measures of ASD-related vocal behaviors and concurrent clinician ratings of autism severity. We used audio signal processing to quantify the rate of child and adult speech (and turn-taking), as well as the acoustic (fundamental frequency) and phonemic (speech sounds per vocalization) qualities of that speech. To determine the potential association of the objective measures, we examined their zero-order and multivariate associations with clinical indices of autism severity (examiner ADOS-2 ratings). All features examined have shown previous associations with ASD-related symptoms and behaviors. However, analyses were exploratory in that we did not formulate a priori hypotheses about the relative importance of the vocalization features in predicting clinician severity ratings.

METHODS

Participants

This was a community sample of 66 preschool age children (age 24–66 months) with suspected ASD. The mean age of the children was 39.97 months (SD = 10.58) and 49 were male (74.2%). Parents indicated that 43 children were White (65.2%), 19 (28.8%) were Black, and four (6.1%) were mixed-race. With respect to ethnicity, 48 of the children were identified as Hispanic or Latino (72.7%). Participants were recruited through an autism specialty clinic housed in the Department of Psychology, which provides free ASD assessments for underserved families in the community. Based on this assessment (see Section 2.2), 61 children (92.4%) received a diagnosis of ASD and 5 (7.6%) received a developmental delay (DD) diagnosis. Table 1 provides sample demographic information. All procedures were approved by the university Institutional Review Board (IRB).

Procedure

Families were invited to participate in the study during their child's clinical evaluation at a university-based autism clinic. A research staff member described the study procedures and addressed questions. Informed consent was obtained from parents prior to initiating the study protocol. The child, parent, and a female examiner were present for the ADOS-2 administration, which was given as part of the standard clinic assessment battery. During the ADOS-2 administration, the research reliable examiner wore a pair of video-enabled eyeglasses from which audio was analyzed. Clinic visits also included administration of developmental assessments, which varied based on the child's age, and parent interviews.

Measures

Autism Diagnostic Observation Schedule, Second Edition

The ADOS-2 (Lord & Rutter, 2012) is a semi-structured observational assessment of play, social interaction, communicative skills, and behavior that was designed to identify the presence of ASD. We administered four ADOS-2 modules—sets of developmentally appropriate activities—based on the child's age and expressive language use. Children completing module T or 1 (i.e., using no verbal language or only single words), module 2 (using phrase speech), or module 3 (fluent speech) were included in the current study. Of the sample, 14 children were evaluated using Module T, 39 with Module 1, 9 with Module 2, and 4 with Module 3 (see Table 1).

TABLE 1 Participant characteristics

	Total ($N = 66$)	Module T ($N = 14$)	Module 1 ($N = 39$)	Module 2 ($N = 9$)	Module 3 ($N = 4$)
Mean (SD)					
Age (months)	40.0 (10.58)	28.6 (1.70)	39.4 (6.54)	53.7 (10.90)	54.8 (9.43)
ADOS-2 CSS total	6.7 (2.33)	7.57 (2.59)	6.4 (2.27)	6.4 (2.07)	6.8 (2.63)
ADOS-2 SA CSS	6.0 (2.66)	7.14 (3.21)	5.4 (2.29)	6.1 (2.80)	7.3 (2.87)
ADOS-2 RRB CSS	8.0 (2.03)	7.8 (1.42)	8.5 (1.86)	7.4 (2.88)	5.5 (1.00)
N (%)					
Sex					
Male	49 (74.2)	11 (78.6)	30 (76.9)	6 (66.7)	2 (50.0)
Female	17 (25.8)	3 (21.4)	9 (23.1)	3 (33.3)	2 (50.0)
Ethnicity					
Latino	48 (72.7)	8 (57.1)	30 (76.9)	7 (77.8)	3 (75.0)
Non-Latino	18 (27.3)	6 (42.9)	9 (23.1)	2 (22.2)	1 (25.0)
Race					
White	43 (65.2)	6 (42.9)	27 (69.2)	6 (66.7)	4 (100.0)
Black	19 (28.8)	6 (42.9)	11 (28.2)	2 (22.2)	0 (0.0)
Mixed	4 (6.1)	2 (14.3)	1 (2.6)	1 (11.1)	0 (0.0)
Diagnosis					
ASD	61 (92.4)	13 (92.9)	37 (94.9)	7 (77.8)	4 (100.0)
DD	5 (7.6)	1 (7.1)	2 (51.1)	2 (22.2)	0 (0.0)

Abbreviations: ADOS-2, autism diagnostic observation schedule; ASD, autism spectrum disorder; CSS, calibrated severity scores; DD, developmental delay; RRB, restricted and repetitive behavior; SA, social affect.

The study examiner's research reliable administration of the ADOS-2 provided scores in two domains: SA (e.g., complexity of language use, conversational reciprocity) and RRB (e.g., echolalia, stereotyped motor mannerisms). The ADOS-2 provides CSS based on data-derived 14-item algorithms composed of the most discriminative items in each module (Esler et al., 2015; Hus et al., 2014; Janvier et al., 2021; Lord & Rutter, 2012). CSS range from 1 (*little to no evidence of ASD-related symptomology*) to 10 (*high level of symptoms)* and are designed to quantify severity between modules. RRB CSS cannot be assigned values of 2, 3, or 4, (Esler et al., 2015; Hus et al., 2014). We examined associations between objective measures of ADOS-2 vocalizations and SA CSS and RRB CSS.

Diagnosis

Diagnoses were provided by three ASD-experienced clinical psychologists based on DSM-5 criteria (APA, 2013) informed by the ADOS-2, parental interviews, and developmental assessments. Each child was evaluated by a single psychologist. Sixty-one children (92.4%) received a diagnosis of ASD, and five children (7.6%) received a diagnosis of DD. Three children assigned DD diagnoses received Total CSS suggestive of ASD (≥4) and five children with ASD diagnoses received Total CSS of 3.

Objective measurements of vocalizations

Speaker identification

To minimize potential distractions and disruptions to children being assessed for ASD, children did not wear vests equipped with audio recorders. Instead, audio recordings were obtained from examiner-worn eyeglasses. Both ORCA (30 children, 54.5%; AAC Mono, 32 kHz) and Pivothead SMART eyewear (36 children, 45.5%; AAC Mono, 44.1 kHz) were used across the study period. Audio files were submitted to the LENA SP Hub Version 3.3.0 software system. The LENA software distinguished between the vocalizations of adults and the vocalizations of children proximal to the recorder, the vocalizations of other children, and electronic sounds (Xu et al., 2009). LENA-identified adult vocalizations and the vocalizations of the proximal child were analyzed (utterance segments). The vocalizations of the proximal child were differentiated into child speech-like vocalizations, cry vocalizations, and vegetative sounds (e.g., burping). Speech-like vocalizations involved pre-linguistic or linguistic phonemic production (e.g., cooing, babbling, full words). LENA has been used to analyze the vocalizations of both TD children and children with developmental disabilities such as ASD (Fasano et al., 2021; Gilkerson & Richards, 2009; Warlaumont et al., 2010).

Vocalization quantity

LENA-based vocalization variables were expressed as rates per minute and as durational proportions of the ADOS-2 (vocal behavior duration divided by ADOS-2 recording duration). Child vocalizations were expressed as a rate per minute and child crying (CC) was expressed as a durational proportion. LENA quantifies adult-child conversational turn counts (CTC) as the number of successive responses to an adult or child vocalization (within 5 s) by the other partner. CTC was expressed as a rate per minute. Adult vocalization indices included adult word count (AWC) expressed as a rate per minute, mean length of vocalizations in seconds, and the proportional duration of adult speech. LENA does not differentiate between adults of the same sex, thus adult examiner and parent vocalizations were aggregated into a single measure of adult vocalizations.Two experts coded 326 LENA-identified vocalizations randomly sampled from all children's ADOS-2 recordings. Overall agreement (adult versus child vocalization) was 68%. For LENA-identified adult vocalizations, expert coders agreed that the vocalizations were produced by adults in 81.1% of cases. For LENA-identified child vocalizations, experts agreed that the vocalizations were produced by children in 55.6% of cases. This level of observed reliability is similar to that described in previous assessments of the reliability of LENA vocal classification (Lehet et al., 2020; Jones et al., 2019). That is, with respect to expert coders, LENA mis-identified adult speech as child speech in 44.4% of cases. It appeared, in particular, that LENA mis-identified infant-directed speech as child vocalizations (Lehet et al., 2020).

Fundamental vocal frequencies (F0)

PRAAT (Boersma & Weenik, 2020) was utilized to separately ascertain the mean fundamental vocal frequencies (F0) of LENA-identified child speech-like vocalizations, child cry vocalizations and adult speech in semitones (Esposito et al., 2014; Sheinkopf et al., 2012; Yankowitz et al., 2019). In PRAAT, the range of F0 was set to 75–600 Hz in order to be sensitive to atypical vocal characteristics at both the high and low ends of the infant vocal spectrum.

Phoneme count

LENA-identified child speech-like vocalizations and adult vocalizations were further processed using Sphinx-4 open-source software (Lamere et al., 2003), which provided estimates of child and adult phonemic richness. Phonemic richness was indexed by the average count per vocalization of consonants and vowels (Xu et al., 2014). This measure has been related to children's current and future expressive language skills (Woynaroski et al., 2017). ACPU-C + V was calculated as the sum of the consonants and vowels, including duplicated phonemes, in a vocalization.

RESULTS

Sample and variable description

IBM SPSS Statistics version 26.0.0 on a Windows computer was utilized for analyses. Table 2 presents pairwise associations of ADOS-2 severity scores with child age, sex, and the duration of the assessment. ADOS-2 scores were not associated with the duration of the recorded assessment, child age, or child sex. Table 3 shows cross-correlations among objective measurements of adult and child vocalizations during the ADOS-2. Several patterns of association ($p < 0.01$) were noteworthy. The rate of child vocalizations and the rate of adult vocalizations were both significantly associated with adult-child conversational turn-counts. Likewise, child speech-like vocalization mean F0, child cry vocalization mean F0, and adult vocalization F0 were all significantly associated. Finally, the phonemic complexity of child speech-like vocalizations and adult vocalizations were associated.

Zero-order associations

Table 4 presents pairwise univariate correlations between objective measurements of vocalizations and ADOS-2 severity scores. No objective measurements of vocalizations were significantly associated with SA CSS. Higher child vocalization mean F0, higher child cry vocalization

TABLE 2 ADOS-2 scores

	Mean	SD	Range	Total CSS	SA CSS	RRB CSS	Age	Sex
Total CSS	6.70	2.33	1–10	—				
SA CSS	6.00	2.65	2–10	0.92**	—			
RRB CSS	8.05	2.03	1–10	0.47**	0.19	—		
Age in months	39.97	10.58	24–68	0.02	0.02	0.02	—	
Sex	—	—	—	0.05	0.17	−0.12	0.14	—
ADOS-2 recording duration	40.80	13.03	7–69	0.12	0.16	−0.09	0.04	0.11

Note: Sex: male = 1, female = 2. Duration is measured in minutes. $N = 66$.

Abbreviations: ADOS-2, autism diagnostic observation schedule; CSS, calibrated severity scores; RRB, restricted and repetitive behavior; SA, social affect.

*$p < 0.05$. **$p < 0.01$.

TABLE 3 Associations among child and adult vocalizations during the ADOS-2

	1	2	3	4	5	6	7	8	9	10
1. Child vocalizations per minute	—									
2. Proportion of time crying	0.48**	—								
3. Child phonemes	0.18	0.01	—							
4. Child cry vocalization mean F0	−0.01	0.23	−0.12	—						
5. Child vocalization mean F0	−0.08	0.15	−0.43**	0.71**	—					
6. Adult word count per minute	0.03	−0.07	0.16	−0.48**	−0.38**	—				
7. Proportion of time adults speaking	0.05	−0.11	0.10	−0.25*	−0.27*	0.77	—			
8. Adult mean duration of vocalization	0.04	−0.16	0.13	−0.30*	−0.35**	0.60**	0.69**	—		
9. Adult vocalization mean F0	−0.15	−0.11	−0.17	0.53**	0.64**	−0.18	−0.06	−0.24	—	
10. Adult phonemes	0.00	−0.14	0.39**	−0.34**	−0.52**	0.52**	0.44**	0.71**	−0.41**	—
11. Adult-child CTC per minute	0.73**	0.24	0.28*	−0.25*	−0.26*	0.45**	0.42**	0.40**	−0.20	0.28*

Note: Phoneme count (both child phonemes and adult phonemes) is indexed by ACPU-C + V (Average Count Per Utterance of Consonants and Vowels); CTC (Conversational Turn Count). $N = 66$.
Abbreviation: ADOS-2, autism diagnostic observation schedule.
$*p < 0.05$. $**p < 0.01$

mean F0, and higher adult vocalization mean F0 were significantly associated with higher ADOS-2 RRB CSS. Fewer child phonemes per utterance (ACPU-C + V) and decreased AWC per minute were significantly associated with ADOS-2 RRB CSS. Examination of the association of these objective features with RRB CCS *within* ADOS-2 modules yielded results similar to those reported here for all ADOS-2 modules (see Table S1).

Regression models

A series of backward, stepwise multiple regressions were undertaken to predict ADOS-2 RRB CSS. The first regression included only child vocalization indicators. The second included both child and adult indicators (including adult-child conversational turn-counts). For each regression, all child indicator variables or all child and adult indicator variables that had a significant zero-order association ($p < 0.05$) with the ADOS-2 RRB CSS were included as potential predictors.

The regression model constrained to child vocalization indicators included child phoneme count, child speech-like vocalization mean F0, and child cry vocalization mean F0. In the final model, two variables significantly explained unique variance in RRB CSS, adjusted $R^2 = 0.31$, $F(2,63) = 13.88$, $p < 0.001$ (Table 5, Figure S1). Lower child phoneme count, $\beta = -0.37$, $t(63) = -3.56$, $p < 0.01$, and higher child cry vocalization mean F0, $\beta = 0.40$, $t(63) = 3.83$, $p < 0.001$, were associated with higher RRB CSS. Next, both adult and child vocalization features—child phoneme count, adult phoneme count, child speech-like vocalization mean F0, child cry vocalization mean F0, adult vocalization mean F0, AWC per minute, adult mean duration of vocalization, and adult-child CTC—were used to predict ADOS-2 RRB CSS. The final model replicated the child model (Table 5, Figure S1). That is, lower child phoneme count and higher child cry vocalization fundamental frequency—but no adult vocal indicators—emerged as predictors of RRB CSS. Controlling for child age in the final regression model did not affect the magnitude or direction of the association between either child phoneme count or cry vocalization mean F0 and RRB CSS (see Table S1).

We are alert to the possibility that backward regression may be sensitive to small differences in the intercorrelation of variables. Consequently, we applied a Benjamini-Hochberg correction with a false discovery rate of 10% to correlations between objective measures of vocalizations and RRB CSS to account for Type 1 error (see Table S1). Using this stricter criterion, the F0 of child vocalizations, child cries, and adult vocalizations—as child phoneme count—continued to be significantly related to RRB CSS. In all regression models, restricting variables to those that survived the Benjamini-Hochberg correction yielded results equivalent to those reported here.

TABLE 4 Associations between objective measurements of vocalizations and ADOS-2 scores

	ADOS-2 total CSS	ADOS-2 SA CSS	ADOS-2 RRB CSS	Child age
Child vocalizations per minute	−0.18	−0.14	−0.03	0.05
Proportion of time crying	0.01	−0.01	0.11	0.07
Child phoneme count	−0.16	−0.01	−0.42**	0.26*
Child cry vocalization mean F0	0.18	0.03	0.44**	−0.24
Child vocalizations mean F0	0.29*	0.11	0.51**	−0.33**
Adult word count per minute	−0.07	0.07	−0.24*	0.35**
Adult proportion of time speaking	−0.12	−0.03	−0.19	0.11
Adult mean duration of vocalizations	−0.01	0.08	−0.16	0.07
Adult vocalization mean F0	0.19	0.03	0.44**	−0.04
Adult phoneme count	0.11	0.21	−0.17	0.32**
Adult-child CTC per minute	−0.22	−0.13	−0.17	0.11

Note: Phoneme count (both child phonemes and adult phonemes) is indexed by ACPU-C + V (average count per utterance of consonants and vowels); CTC (conversational turn count). $N = 66$.
Abbreviations: ADOS-2, autism diagnostic observation schedule; CSS, calibrated severity scores; RRB, restricted and repetitive behavior; SA, social affect.
*$p < 0.05$. **$p < 0.01$.

TABLE 5 Final regression model: child vocalization variables predicting ADOS-2 restricted and repetitive behaviors CSS

Model	*B*	SE	*β*	*t*	Sig.	Zero-order	Partial	*F*	R^2	Adjusted R^2	Sig.
(Constant)	1.85	3.51		0.528	0.600			15.50	0.330	0.308	0.000
Child phoneme count	−1.37	0.39	−0.37	−3.56	0.001	−0.42	−0.25				
Child cry vocalization mean F0	0.71	0.19	0.40	3.83	0.000	0.44	0.40				

Note: Child phoneme count is the average count per utterance of consonants and vowels (ACPU-C + V). Results of backward regression employing a $p < 0.05$ inclusion criterion. The addition of adult vocalization indicators as well as child prediction variables yielded the same results as the model employing only child vocalization predictors. $N = 66$.
Abbreviations: ADOS-2, autism diagnostic observation schedule; CSS, calibrated severity scores.

To investigate the potential effects of LENA misclassification on study results, we identified a subsample of 120 vocalizations clips that LENA identified as child vocalizations that were produced by eight study participants (i.e., in eight study protocols). For 71 of these vocalizations, expert coders agreed these were child vocalizations; for 49 of the clips, experts coded these as adult vocalizations. Using a paired samples *t* test, the mean F0 of vocalizations in which there was agreement ($M = 22.25$, SD = 2.32) and disagreement ($M = 21.80$, SD = 3.25) did not differ, $t(7) = -0.288$, $p = 0.782$, $d = 0.159$. Likewise, the phoneme count of vocalizations in which there was agreement ($M = 4.42$, SD = 1.53) and disagreement ($M = 4.23$, SD = 1.53) on whether the vocalization was a child vocalization did not differ, $t(7) = -0.346$, $p = 0.740$, $d = 0.126$.

DISCUSSION

We conducted objective quantification of ASD-related characteristics of vocalizations occurring during administration of the ADOS-2 to determine their associations with clinician-rated ASD severity. Both phonemic and acoustic features of vocalizations were associated with clinician-rated indices of autism symptom severity in the RRB domain. Less phonemically complex child vocalizations were associated with higher RRB severity scores, as were higher pitched child and adult vocalizations.

No vocalization variables were associated with the SA domain. This was unexpected as SA domain algorithm items include direct rating of reciprocal vocal communication (Eni et al., 2020). In fact, the pragmatic use of language falls in the social communication domain in the DSM-5 description of ASD symptoms. Instead, we found meaningful patterns of association in the RRB domain. Although a direct rating of the acoustic aspects of vocalizations ("intonation of vocalizations or verbalizations") is only included for Modules T and 1, the DSM-5 description of ASD symptoms lists repetitive and stereotyped qualities of vocalizations under the RRB domain (American Psychiatric Association, 2013). One possibility is that these acoustic and phonemic vocalization qualities covary with concurrent ratings of overall level of ASD-related impairment that are captured by the RRB construct.

A growing literature indicates that children with ASD use fewer phonemes (ACPU-C + V; (Xu et al., 2014) and a less diverse set of consonants (Wetherby et al., 2007) than their TD peers. Extending these findings, we found that child speech involving lower numbers of phonemes was associated with higher ratings of RRB severity within the ADOS-2. We employed an estimate of the number of phonemes per utterance that included repeated phonemes, a coarse measure of phonemic richness. Nevertheless, this coarse measure of phonemically rich child speech was associated with lower RRB CSS (which are designed to index severity despite differences in expressive language). The results suggest that child phonemic richness is a key index of the severity of repetitive behavior in ASD. In addition, the use of diverse phonemes predicts later language skills in children with ASD (Saul & Norbury, 2020; Woynaroski et al., 2017), providing additional support to the clinical importance of measures of phonemic production at diagnosis.

We found associations between ADOS-2 scores and the fundamental frequency of child cries, child speech-like vocalizations, and adult vocalizations. Notably, these three measures of fundamental frequency were highly correlated and had similar zero-order associations with RRB severity. This is consistent with previous findings. Eni et al. (2020) found significant positive zero-order associations between the F0 of child vocalizations during the ADOS-2 and ADOS-2 RRB Raw Scores. More generally, young children with ASD demonstrate cries of significantly higher pitch than their TD peers (Esposito et al., 2014; Sheinkopf et al., 2000), and parents of infants at risk for ASD demonstrate more variable pitch in their infant-directed vocalizations than other parents (Quigley et al., 2016). Bone et al. (2014) found significant associations between ADOS-2 total scores and the pitch variability of both examiner and child vocalizations during the emotion interview activity of the ADOS-2 Module 3. There is thus reason to believe both that the fundamental frequency of child vocalizations is associated with ASD symptoms, and growing evidence that the F0 of child vocalizations occurring during the ADOS-2 may be a promising index of RRB severity.

Child vocalization rates were not significantly associated with child RRB CSS in the current study. Likewise, others have found that the quantity of children's classroom speech is not associated with their autism symptom severity (Dykstra et al., 2013). Use of a measure of vocal responsivity that focuses on the child's responsiveness to adults and reduces the influence of chanced vocalization pairings may serve as a better future indicator in this area (Harbison et al., 2018). Consideration of the *quantity* of vocalizations may be less informative in indexing autism symptom severity than consideration of the acoustic and phonemic *quality* of the vocalizations produced.

We examined vocal characteristics previously found to be associated with ASD symptoms using readily available automated measures to identify child and adult vocalizations. Deep learning of the vocal characteristics of vocalizations has shown strong association with ASD symptoms (Eni et al., 2020; Sadiq), but deep learning approaches do not shed appreciable light on the ASD phenotype. However, Eni et al.'s examination of linear associations between child vocalization F0 and ADOS-2 scores parallels our results. The current study combined the use of both automated detection of vocalizations and considered the direct relationship between objectively measured vocalization features and clinical ratings of behavior.

The current study has noteworthy limitations and strengths. The study design examined concurrent objective and subjective measures of vocalizations during administration of the ADOS-2 but did not allow for prediction of future symptoms. In addition, the current approach should be applied to larger samples of behavior in different contexts to further test the association between acoustic and phonological characteristics of children's vocalizations and their clinically ascertained autism severity. The sample size was limited and included a range of ages, necessitating the use of four ADOS-2 modules. Consequently, we examined associations between objective measurements of vocalizations and CSS, which allowed for comparison across modules. Moreover, results did not change when adding age as a covariate in final models. In addition, results characterized a diverse community sample, suggesting the ecological validity of digital phenotyping focused on children's vocalizations.

LENA algorithms were trained and tested using child-worn recorders (Gilkerson & Richards, 2009; Warren et al., 2010; Woynaroski et al., 2017), while the current study utilized audio from examiner-worn eyeglasses. Although the reliability of speaker classification between adults and children in the current sample was low it was comparable to LENA reliability in which children wore specially designed vests (Lehet et al., 2020). Moreover, characteristics of correctly and incorrectly identified child vocalizations appeared to be similar. In fact, Eni et al. (2020) found similar results with respect to vocalization F0 using expert-identified child vocalizations that were further isolated computationally based on their energy levels. The acoustic and phonemic characteristics of vocalizations identified by LENA as child vocalizations appeared to be similar regardless of whether human experts coded those vocalizations as child or adult vocalizations. Nevertheless, a substantial proportion of LENA-identified child vocalizations in the full reliability sample were misclassified (that is, they were coded by experts as adult vocalizations). Thus, there is a possibility that the acoustic and phonemic qualities of these vocalizations were influenced by characteristics of adult vocalizations in these protocols. Finally, the examiner-worn eyeglasses used AAC, which compresses files for storage and has the potential to reduce the accuracy of acoustic feature estimation (Gabrieli et al., 2019). Future studies utilizing a non-lossy audio format may improve accuracy. Thus, our vocal measurements and speaker identification may provide only coarse indices of speech during the ADOS-2. However, the consistent pattern of results obtained suggests the potential of readily accessible technologies to characterize ADOS-2 vocalizations and index ASD severity.

CONCLUSIONS

The current study builds on a growing effort to characterize autism using objective measures of behavior. We document meaningful patterns of associations between objective characteristics of vocalizations during the ADOS-2 and clinician ratings of autism severity in a community sample, predicting over a quarter of the variance in these ratings. Less phonemically complex child speech and higher pitched speech were associated with higher levels of severity. Replication of these exploratory findings will be necessary. Future studies with larger samples should consider the consistency of prediction of ADOS-2 scores for children of various ethnic and cultural groups. These results suggest that objective measures of vocalizations reflect underlying ASD-related characteristics, and that these measures have the potential to inform clinician severity ratings of young children being assessed for ASD.

ACKNOWLEDGEMENTS

The project described was supported in part by Grant Number UL1TR002736. Miami Clinical and Translational Science Institute, from the National Center for Advancing Translational Sciences and the National Institute on Minority Health and Health Disparities. Its contents are solely the responsibility of the authors. This project was supported in part by an Autism Science Foundation Predoctoral Fellowship (18-001) entitled "Automating and accelerating the autism diagnostic process," awarded to Y. Ahn.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on the Open Science Framework at https://osf.io/g3dc7/.

ORCID

Jacquelyn M. Moffitt https://orcid.org/0000-0002-1873-5326

Stephanie Custode https://orcid.org/0000-0002-5078-6350

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

Boersma, P. & Weenik, D. (2020). Praat: Doing phonetics by computer [computer program]. https://www.fon.hum.uva.nl/praat/ or www.praat.org

Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. (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. https://doi.org/10.1044/2014_JSLHR-S-13-0062

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech Spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, *4*, 237. https://doi.org/10.3389/fnhum.2010.00237

Dawson, G., & Sapiro, G. (2019). Potential for digital behavioral measurement tools to transform the detection and diagnosis of autism spectrum disorder. *JAMA Pediatrics*, *173*(4), 305–306. https://doi.org/10.1001/jamapediatrics.2018.5269

Dykstra, J. R., Sabatos-Devito, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L. (2013). Using the language environment analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, *17*(5), 582–594. https://doi.org/10.1177/1362361312446206

Eni, M., Dinstein, I., Ilan, M., Menashe, I., Meiri, G., & Zigel, Y. (2020). Estimating autism severity in young children from speech signals using a deep neural network. *IEEE Access*, *8*, 139489–139500. https://doi.org/10.1109/ACCESS.2020.3012532

Esler, A., Bal, V. H., Guthrie, W., Wetherby, A. M., Weismer, S. E., & Lord, C. (2015). The autism diagnositic observation schedule, toddler module: Standardized severity scores. *Journal of Autism and Developmental Disorders*, *45*(9), 2704–2720. https://doi.org/10.1007/s10803-015-2432-7

Esposito, G., Del, M., Rostagno, C., Venuti, P., Haltigan, J. D., & Messinger, D. S. (2014). Brief report: Atypical expression of distress during the separation phase of the strange situation procedure in infant siblings at high risk for ASD. *Journal of Autis*, *44*, 975–980. https://doi.org/10.1007/s10803-013-1940-6

Esposito, G., & Venuti, P. (2010). Developmental changes in the fundamental frequency (f0) of infants' cries: A study of children with autism spectrum disorder. *Early Child Development and Care*, *180*(8), 1093–1102. https://doi.org/10.1080/03004430902775633

Fasano, R. M., Perry, L. K., Zhang, Y., Vitale, L., Wang, J., Song, C., & Messinger, D. S. (2021). A granular perspective on inclusion: Objectively measured interactions of preschoolers with and without autism. *Autism Research*, *14*(8), 1658–1669. https://doi.org/10.1002/aur.2526

Gabrieli, G., Scapin, G., Bornstein, M., & Esposito, G. (2019). Are cry studies replicable? An analysis of participants, procedures, and methods adopted and reported in studies of infant cries. *Acoustics*, *1*(4), 866–883. https://doi.org/10.3390/acoustics1040052

Gilkerson, J., & Richards, J. A. (2009). The power of talk. In *LENA technical report (LTR-01-2)* (2nd ed., pp. 1–30).Boulder, CO: LENA Research Foundation.

Harbison, A. L., Woynaroski, T. G., Tapp, J., Wade, J. W., Warlaumont, A. S., & Yoder, P. J. (2018). A new measure of child vocal reciprocity in children with autism spectrum disorder. *Autism Research*, *11*(6), 903–915. https://doi.org/10.1002/aur.1942

Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, *44*, 2400–2412. https://doi.org/10.1007/s10803-012-1719-1

Janvier, D., Choi, Y. B., Klein, C., Lord, C., & Kim, S. H. (2021). Brief report: Examining test-retest reliability of the autism diagnostic observation schedule (ADOS-2) calibrated severity scores (CSS). *Journal of Autism and Developmental Disorders*, *0123456789*, 1–7. https://doi.org/10.1007/s10803-021-04952-7

Jones, R. M., Plesa Skwerer, D., Pawar, R., Hamo, A., Carberry, C., Ajodan, E. L., Caulley, D., Silverman, M. R., McAdoo, S., Meyer, S., Yoder, A., Clements, M., Lord, C., & Tager-Flusberg, H. (2019). How effective is LENA in detecting speech vocalizations and language produced by children and adolescents with ASD in different contexts? *Autism Research*, *12*(4), 628–635. https://doi.org/10.1002/aur.2071

Kamp-Becker, I., Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., Poustka, L., Weber, L., Schmidt, H., Smidt, J., Stehr, T., Roessner, V., Kucharczyk, K., Wolff, N., & Stroth, S. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child and Adolescent Psychiatry*, *27*(9), 1193–1207. https://doi.org/10.1007/s00787-018-1143-y

Kumar, M., Gupta, R., Bone, D., Malandrakis, N., Bishop, S., & Narayanan, S. (2016). Objective language feature analysis in children with neurodevelopmental disorders during autism assessment. In *Proceedings of the annual conference of the international speech communication association, INTERSPEECH, 08-12-Sept* (pp. 2721–2725). https://doi.org/10.21437/Interspeech.2016-563

Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., Warmuth, M., & Wolf, P. (2003). The CMU SPHINX-4 speech recognition system. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong (Vol. *1*, pp. 2–5).

Lehet, M., Arjmandi, M. K., Houston, D., & Dilley, L. (2020). Circumspection in using automated measures: Talker gender and addressee affect error rates for adult speech detection in the language ENvironment analysis (LENA) system. *Behavior Research Methods*, *53*, 113–138. https://doi.org/10.3758/s13428-020-01419-y

Lord, C., & Rutter, M. (2012). *Autism diagnostic observation schedule* (2nd ed.). Torrance, CA: Western Psychological Services.

Nakai, Y., Takashima, R., Takiguchi, T., & Takada, S. (2014). Speech intonation in children with autism spectrum disorder. *Brain and Development*, *36*(6), 516–522. https://doi.org/10.1016/j.braindev.2013.07.006

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(30), 13354–13359. https://doi.org/10.1073/pnas.1003882107

Plumb, A. M., & Wetherby, A. M. (2013). Vocalization development in toddlers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, *56*, 721–734. https://doi.org/10.1044/1092-4388(2012/11-0104)

Quigley, J., McNally, S., & Lawson, S. (2016). Prosodic patterns in interaction of low-risk and at-risk-of-autism spectrum disorders infants and their mothers at 12 and 18 months. *Language Learning and Development*, *12*(3), 295–310. https://doi.org/10.1080/15475441.2015.1075405

Ramsay, G., Edwards, M., Bailey, J., & Ghai, S. (2021). *Developmental cascades Linkign infant-directed speech, infant vocal behavior, and social contingency in typical development and ASD*. INSAR.

Sadiq, S., Castellanos, M., Moffitt, J., Shyu, M. L., Perry, L., & Messinger, D. (2019). Deep learning based multimedia data mining for autism spectrum disorder (ASD) diagnosis. In *IEEE international conference on data mining workshops, ICDMW, 2019-November* (pp. 847–854). https://doi.org/10.1109/ICDMW.2019.00124

Saul, J., & Norbury, C. (2020). Does phonetic repertoire in minimally verbal autistic preschoolers predict the severity of later expressive language impairment? *Autism*, *24*(5), 1217–1231. https://doi.org/10.1177/1362361319898560

Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research: Official Journal of the International Society for Autism Research*, *4*(3), 177–188. https://doi.org/10.1002/aur.183

Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., & Lester, B. M. (2012). Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Research*, *5*, 331–339. https://doi.org/10.1002/aur.1244

Sheinkopf, S. J., Mundy, P., Oller, D. K., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, *30*(4), 345–354 http://www.ncbi.nlm.nih.gov/pubmed/11039860

Warlaumont, A. S., Oller, D. K., Dale, R., Richards, J. A., Gilkerson, J., & Xu, D. (2010). Vocal interaction dynamics of children with and without autism. In *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 121–126).

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*, *40*(5), 555–569. https://doi.org/10.1007/s10803-009-0902-5

Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders*, *37*(5), 960–975. https://doi.org/10.1007/s10803-006-0237-4

Woynaroski, T., Oller, D. K., Keceli-Kaysili, B., Xu, D., Richards, J. A., Gilkerson, J., Gray, S., & Yoder, P. (2017). The stability and validity of automated vocal analysis in preverbal preschoolers with autism spectrum disorder. *Autism Research*, *10*(3), 508–519. https://doi.org/10.1002/aur.1667

Xu, D., Richards, J. A., & Gilkerson, J. (2014). Automated analysis of child phonetic production using natrualistic recordings. *Journal of Speech, Language, and Hearing Research*, *57*, 1638–1650.

Xu, D., Richards, J. A., Gilkerson, J., Yapanel, U., Gray, S., & Hansen, J. (2009). Automatic childhood autism detection by vocalization decomposition with phone-like units general terms. *Proceedings of the 2nd Workshop on Child, Computer and Interaction*. https://doi.org/10.1145/1640377.1640382

Yankowitz, L. D., Schultz, R. T., & Parish-Morris, J. (2019). Pre- and paralinguistic vocal production in ASD: Birth through school age. *Current Psychiatry Reports*, *21*(12), 126. https://doi.org/10.1007/s11920-019-1113-1

Zander, E., Willfors, C., Berggren, S., Choque-Olsson, N., Coco, C., Elmund, A., Hedfors Moretti, Å., Holm, A., Jifält, I., Kosieradzki, R., Linder, J., Nordin, V., Olafsdottir, K., Poltrago, L., & Bölte, S. (2016). The objectivity of the autism diagnostic observation schedule (ADOS) in naturalistic clinical settings. *European Child & Adolescent Psychiatry*, *25*, 769–780. https://doi.org/10.1007/s00787-015-0793-2

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Moffitt, J. M., Ahn, Y. A., Custode, S., Tao, Y., Mathew, E., Parlade, M., Hale, M., Durocher, J., Alessandri, M., Perry, L. K., & Messinger, D. S. (2022). Objective measurement of vocalizations in the assessment of autism spectrum disorder symptoms in preschool age children. *Autism Research*, 1–10. https://doi.org/10.1002/aur.2731