



Published in final edited form as:

*Autism Res.* 2020 August ; 13(8): 1373–1382. doi:10.1002/aur.2293.

## A Six-Minute Measure of Vocalizations in Toddlers with Autism Spectrum Disorder

**Elena J. Tenenbaum<sup>#</sup>, Kimberly L.H. Carpenter<sup>#</sup>, Maura Sabatos-DeVito, Jordan Hashemi, Saritha Vermeer, Guillermo Sapiro, Geraldine Dawson**

Duke Center for Autism and Brain Development, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina (E.J.T., K.L.C., M.S.-D., J.H., S.V., G.D.); Electrical and Computer Engineering, Duke University, Durham, North Carolina (J.H., G.S.); Department of Biomedical Engineering, Duke University, Durham, North Carolina (G.S.); Department of Computer Science, Duke University, Durham, North Carolina (G.S.); Department of Mathematics, Duke University, Durham, North Carolina (G.S.)

<sup>#</sup> These authors contributed equally to this work.

### Abstract

To improve early identification of autism spectrum disorder (ASD), we need objective, reliable, and accessible measures. To that end, a previous study demonstrated that a tablet-based application (app) that assessed several autism risk behaviors distinguished between toddlers with ASD and non-ASD toddlers. Using vocal data collected during this study, we investigated whether vocalizations uttered during administration of this app can distinguish among toddlers aged 16–31 months with typical development (TD), language or developmental delay (DLD), and ASD. Participant's visual and vocal responses were recorded using the camera and microphone in a tablet while toddlers watched movies designed to elicit behaviors associated with risk for ASD. Vocalizations were then coded offline. Results showed that (a) children with ASD and DLD were less likely to produce words during app administration than TD participants; (b) the ratio of syllabic vocalizations to all vocalizations was higher among TD than ASD or DLD participants; and (c) the rates of nonsyllabic vocalizations were higher in the ASD group than in either the TD or DLD groups. Those producing more nonsyllabic vocalizations were 24 times more likely to be diagnosed with ASD. These results lend support to previous findings that early vocalizations might be useful in identifying risk for ASD in toddlers and demonstrate the feasibility of using a scalable tablet-based app for assessing vocalizations in the context of a routine pediatric visit.

### Lay Summary:

Although parents often report symptoms of autism spectrum disorder (ASD) in infancy, we are not yet reliably diagnosing ASD until much later in development. A previous study tested a tablet-based application (app) that recorded behaviors we know are associated with ASD to help identify children at risk for the disorder. Here we measured how children vocalize while they watched the movies presented on the tablet. Children with ASD were less likely to produce words, less likely to produce speechlike sounds, and more likely to produce atypical sounds while watching these

movies. These measures, combined with other behaviors measured by the app, might help identify which children should be evaluated for ASD.

### Keywords

developmental psychology; early detection; early signs; infants; language

## A 6-Minute Measure of Vocalizations in Toddlers with ASD

Although autism spectrum disorder (ASD) is a neuro-developmental condition that likely begins during the prenatal period (Elsabbagh et al., 2012; Wolff et al., 2012), we do not currently have a reliable behavioral measure for identifying children at risk for the disorder early in development. Atypical behavioral patterns across multiple domains, including response to social stimuli [Chawarska, Macari, & Shic, 2013; Hutman, Chela, Gillespie-Lynch, & Sigman, 2012; Ozonoff et al., 2010; Rozga et al., 2011], affect [Clifford et al., 2013], visual orienting [Clifford et al., 2013; Elison et al., 2013; Elsabbagh et al., 2013; Zwaigenbaum et al., 2005], and vocalizations [Oller et al., 2010; Patten et al., 2014; Sheinkopf, Iverson, Rinaldi, & Lester, 2012], have been identified as indicative of ASD in the first year of life, yet the median age of diagnosis in the United States remains at 4.3 years [Baio et al., 2018].

A gap of this magnitude between manifestation of symptoms and diagnosis (often the gate keeper to services for ASD) is a serious public health concern. Early intervention facilitates developmental outcomes in children with ASD [Dawson, 2010; Koegel, Koegel, Ashbaugh, & Bradshaw, 2014; Landa & Kalb, 2012; Rogers et al., 2014, 2019], and recommendations on pediatric screening for the disorder have been in place for more than a decade [Johnson & Myers, 2007]. Despite these recommendations, current estimates suggest that only about 30% of children are in fact being screened [Hirai, Kogan, Kandasamy, Reuland, & Bethell, 2018]. When screening does occur, it is often based on parent report questionnaires such as the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F; Robins, 2008; Robins et al., 2014). The M-CHAT-R/F has gained much traction in primary care settings and has significantly improved detection beyond clinical judgment alone [Robins, 2008]. As with any measure, the M-CHAT-R/F has limitations and recent investigation of the tool in a large pediatric network showed limited accuracy and sensitivity (Guthrie, et al. 2019). Furthermore, the measure requires follow-up by a trained professional; however, studies show that this is often not occurring as recommended, particularly in nonwhite communities and those with lower socioeconomic status [Windham et al., 2014].

The development of automated, scalable screening tools that directly assess behavior are needed. To that end, previous studies [Bovery et al., 2018; Campbell et al., 2019; Dawson & Sapiro, 2019; Hashemi et al., 2014, 2018] have evaluated a tablet-based application (app) that uses dynamic stimuli designed to elicit and measure symptoms associated with ASD. This scalable app can be administered during well-child visits in the primary care setting (and parallel efforts indicate that it is also possible to administer at home [Egger et al.,

2018)). The goal of this automated screener is to mitigate disparities in parent and provider knowledge that may contribute to delayed diagnosis in underserved populations [Windham et al., 2014]. Relying on direct observation rather than parent report, the app uses computer vision to assess social attention, affect, social referencing, and motor behaviors. To date, this approach has been shown to reliably distinguish between toddlers at high risk for ASD from toddlers with TD based on patterns of attention, facial expression, motor behavior, and response to name [Campbell et al., 2019; Dawson et al., 2018; Egger et al., 2018]. While this approach does not replace the need for formal evaluation, it does allow for the potential of unbiased identification of children at elevated risk for ASD.

In addition to capturing social attention, affect, orienting, and motor behaviors with the tablet's camera, the microphone in the tablet also records vocalizations produced while children watch the movies. Delayed onset of developmental milestones in vocalizations can be used to identify risk for ASD [Patten et al., 2014; Paul, Fuerst, Ramsay, Chawarska, & Klin, 2011]. Paul et al. [2011] measured the speechlike and non-speechlike vocalizations of infant siblings of children with ASD. They showed that infants at risk for ASD (based on the presence of an older sibling with the disorder) were more likely to be delayed in transitioning from early vocal patterns to speechlike behavior than infants who did not have a sibling with the disorder.

Patten et al. [2014] followed up on this work by demonstrating that 9- to 12-month-olds who later received a diagnosis of ASD compared with typically developing (TD) controls produced fewer vocalizations overall and fewer speechlike “canonical babbling” sounds relative to all vocalizations. Canonical babbling has been defined as fully articulated consonant–vowel combinations with adultlike transitions between the consonant and the vowel (e.g., “mama,” “da,” “baba,” “gu”) [Lee, Jhang, Relyea, Chen, & Oller, 2018; Oller et al., 1998; Patten et al., 2014]. Unlike early vocalizations, or cooing behaviors, canonical syllables involve consonant-like closure of the vocal tract combined with vowel-like opening such that the sounds produced are recognizable as speechlike [Belardi et al., 2017]. Canonical babbling emerges in TD infants by approximately 6–7 months and is generally well established by 10 months [Oller et al., 1998]. The transition from immature vocalizations to canonical babbling is quite salient and recognizable to parents and professionals alike [Oller, Eilers, & Basinger, 2001; Oller et al., 1998]. Delays in the onset of canonical babbling have been associated with a number of non-ASD conditions including hearing impairment [Bass-Ringdahl, 2010; Eilers & Oller, 1994], Fragile X syndrome [Belardi et al., 2017], Down syndrome [Lynch, Oller, Steffens, & Levine, 1995] and cleft-palate [Chapman, Hardin-Jones, Schulte, & Halter, 2001]. Delays in vocal maturity more broadly have also been identified in disorders including specific language impairment [Rescorla & Ratner, 1996] and childhood apraxia of speech [Overby & Caspari, 2015].

Potentially more specific to ASD, atypical vocalizations (also described as nontranscribable, nonsyllabic, and non-speechlike) seem to be a distinguishing feature of ASD relative to both TD children and children with language delays but not ASD (for review see Yankowitz, Schultz and Parish-Morris, 2019). The definition of an atypical vocalization varies slightly from study-to-study, but the general term is most succinctly described by Plumb and Wetherby [2013] who define a transcribable vocalization as “a syllabic vocalization that

contains at least a vowel and may also contain a consonant” and a nontranscribable vocalization as “a nonsyllabic vocalization that does not contain a vowel or contains a vowel with atypical phonation.” Plumb and Wetherby [2013] found lower transcribable vocalization ratios among the children with ASD relative to TD but not developmentally delayed controls. Schoen, Paul, and Chawarska [2011] found similar results in comparable rates of typical vocalizations but higher rates of atypical vocalizations among children with ASD (18–36 months) compared to both age and language matched controls. Similar results were obtained by Sheinkopf, Mundy, Oiler, and Steffens [2000]. Unlike Patten et al. [2014], Sheinkopf et al. showed comparable rates of canonical babbling but again higher rates of atypical phonation among children with ASD relative to children with developmental delays. Most recently, Chenausky, Nelson III, and Tager-Flusberg [2017] found lower rates of speechlike vocalizations among high-risk toddlers with ASD (where risk was due to the presence of an older sibling with the disorder) relative to high-risk toddlers without ASD and low-risk controls, but comparable rates of atypical vocalizations.

In the current study, we explored whether 18- to 30-month-old toddlers with ASD can be reliably distinguished from those with TD or DLD on the basis of vocalizations recorded from the microphone in an iPad that was delivering movie (visual and sound) stimuli designed to assess behaviors associated with ASD. As in other studies of this age range, our sample crossed the boundary between preverbal canonical babbling and the acquisition of words. To address the varied levels of speech production in this sample, we explored a number of potential measures of speech production. These included: nonsyllabic vocalizations (consonant only vocalizations or vowel-only vocalizations that were not part of a word, e.g., “iii” or “nnn”), canonical babbling (e.g., “ba” “na,” Lee et al., 2018; Oller et al., 1998; Patten et al., 2014), syllabic vocalizations (fully formed vowels and/or consonant–vowel combinations that could be contained in words, e.g., “I” “see” “a” “bu/nny”), word types, and word tokens. We also calculated syllabic vocalization ratios (syllabic vocalizations/all vocalizations). While numerous studies have now shown that children with ASD produce fewer speechlike and more atypical vocalizations (see above), the novelty of the current study was in the scalable approach of measuring vocalizations during a 6-min app in the context of a well-child primary care visit.

## Methods

### Ethics

All study protocols were reviewed and approved by the Duke Health Institutional Review Board. All caregivers provided written, informed consent for participation.

### Participants

Participants were selected from a study of 104 toddlers aged 16–31 months who were administered the tablet-based app at their well-child visits [Campbell et al., 2019]. Of the 104 toddlers enrolled in the original study, 22 had confirmed diagnoses of ASD. This study examined the 22 children with ASD versus a comparison sample of TD children chosen from the larger group to carefully match the ASD group on age and gender ( $n = 22$ ) and all participants from the original study with confirmed language or developmental delays but

not autism ( $n = 8$ ). Known vision or hearing deficits, and lack of access to English in the home or insufficient English for consenting were exclusionary criteria in the original study.

Parents of all participants completed the M-CHAT-R/F. Those with elevated M-CHAT scores, or for whom there was clinician or parental concern for ASD were then administered the *Autism Diagnostic Observation Schedule, Second Edition* (Toddler Module) (ADOS-2) [Lord et al., 2012]. ASD diagnosis was based on clinical assessment by a licensed clinician and elevated scores on the ADOS-2. Participants with developmental or language delays (DLD) qualified for speech or developmental therapy but did not meet criteria for ASD. TD participants screened negative on the M-CHAT and no parental or clinician concerns regarding ASD had been raised. See Table 1 for demographic information.

## Procedure

Children were assessed during their well-child pediatric visits or in the laboratory. Children sat on a caregiver's lap and watched 6 min of movies on a tablet mounted on a tripod approximately 3 ft from the child. Stimuli were designed to be engaging and to elicit behaviors associated with ASD including atypical social attention, facial expression, motor behavior, and response to name. Movies included a mirror display in which the child could see him/herself (20–45 sec), cascading bubbles ( $2 \times 30$  sec), a mechanical bunny interacting with other animal puppets (66 sec), children arguing over a toy (13 sec), and clips of social and nonsocial stimuli including a female singing nursery rhymes and dynamic noise-making toys (60 and 68 sec). At three predetermined time points during the videos, a prompt on the screen alerted the examiner (positioned behind the parent and off to one side) to call the child's name for measurement of response to name. Movie presentations began with the mirror image to allow for positioning of the child in the tablet screen. During this phase, parents were able to interact with their child, but then were asked to remain quiet for all subsequent segments. Other than the name calls, examiners remained quiet throughout the administration of the movies.

## Coding

During presentation of the stimuli, a camera and microphone within the tablet recorded the child's image and vocalizations, respectively, thereby producing videos of the participant's visual and auditory responses. These recorded videos also included sounds from the stimuli themselves and the examiner's name calls. Video segments were extracted to align with stimulus presentation. There were seven video clips for each participant (responses to mirror, bubbles, bunny, puppets, social/nonsocial, children, bubbles). One primary coder, blind to diagnostic outcomes was trained in the coding procedure by the author of a study on canonical babbling specifically [Patten et al., 2014]. Coding was generally completed in a single-pass of listening to each video segment, but the coder could listen to a given segment up to three times if unsure of how to code it. Vegetative sounds (e.g., sneezes, coughs, and grunts that were associated with physical exertion), cries, and laughter were not included in the coding. As described above, some participants in this study were producing utterances containing words while others were preverbal. To be able to compare vocalizations across these levels, vocalizations were defined as nonsyllabic (consonants-only or vowel-only sounds that were not part of words), canonical (nonword consonant–vowel combinations) or

syllabic (canonical vocalizations + syllabic vocalizations from words containing at least one vowel). Words types and word tokens were tallied. A dichotomous variable, “presence of words,” was created to assess whether the child spoke any words at all during the 6-min procedure. We also calculated a syllabic vocalization ratio based on the total number of syllabic vocalizations relative to all vocalizations.

### Statistical Analyses

All analyses were completed in R (R Core Team, 2018). Analysis of variance (ANOVA) was used to assess group differences in number of nonsyllabic vocalizations, number of canonical vocalizations, number of syllabic vocalizations, number of word types, number of word tokens, and syllabic vocalization ratios. Chi-square was used to assess group differences in the dichotomous presence of words factor. For continuous measures that differed across groups, we plotted the receiver operating characteristic curves (ROC) using the R package “pROC” [Robin et al., 2011]. The “OptimalCutpoints” package in R [López-Ratón, Rodríguez-Álvarez, Cadarso-Suárez, & Gude-Sampedro, 2014] was used to determine optimal cut points, sensitivity, and specificity. Odds ratios were calculated to explore the likelihood of ASD diagnosis given these cut points.

### Results

Out of the 52 total participants, two (ASD:  $n = 1$ , TD:  $n = 0$ , DLD:  $n = 1$ ) did not produce any nonvegetative vocalizations during the procedure. To ensure that we were capturing risk, these cases were assumed to have syllabic vocalization rates of 0. Mean rates, standard deviations, and ranges of the participants’ nonsyllabic vocalizations, canonical syllables, syllabic vocalizations, word types, word tokens, and syllabic vocalization ratios are displayed in Table 2. ANOVA revealed no significant differences between groups on number of vocalizations:  $F(2,49) = 2.02$ ,  $P = 0.14$ ; number of canonical syllables:  $F(2,49) = 1.71$ ,  $P = 0.19$ ; total syllabic utterances:  $F(2,49) = 2.17$ ,  $P = 0.13$ ; or word tokens:  $F(2,49) = 2.99$ ,  $P = 0.06$ . This suggests that groups were not significantly different in how much they vocalized nor in the number of words they produced in the context of this procedure.

The dichotomous presence of words factor was significantly different among the three groups ( $\chi^2 = 21.59$ ,  $P < 0.001$ ). This effect was driven by the ASD versus TD ( $\chi^2 = 20.84$ ,  $P < 0.001$ ) and TD versus DLD ( $\chi^2 = 9.35$ ,  $P = 0.002$ ) contrasts. The ASD versus DLD comparison was not significant ( $\chi^2 = 0.65$ ,  $P = 0.42$ ) (Fig. 1).

ANOVA revealed significant differences in the number of nonsyllabic vocalizations,  $F(2,49) = 8.94$ ,  $P < 0.001$  (Fig. 2). Bonferroni corrected pairwise comparisons revealed that for the nonsyllabic vocalizations, contrasts were significant for ASD versus TD ( $P < 0.001$ ) and for ASD versus DLD ( $P = 0.03$ ), but not for DLD versus TD (ns). The mean raw number of nonsyllabic vocalizations among participants with ASD was 20.86 ( $SD = 22.14$ ), while among DLD and TD participants it was much lower (DLD: mean = 5.12,  $SD = 5.19$ ; TD: mean = 2.82,  $SD = 2.92$ ). ANOVA also showed significant differences in the number of word types:  $F(2,49) = 3.95$ ,  $P < 0.05$ , though Bonferroni corrected comparisons between groups were not significant ( $P > 0.07$ ). Word types was therefore dropped from the remaining analyses of potential differentiating aspects of vocalizations.

Finally, there was a significant difference by group in the ratio of syllabic vocalizations to all vocalizations,  $F(2,49) = 15.47$ ,  $P < 0.001$  (Fig. 3). For syllabic vocalization ratios, as with the dichotomous presence of words factor, the contrast was largely driven by the TD versus ASD rates ( $P < 0.001$ ). The difference in syllabic vocalization ratios was also significant for TD versus DLD,  $P = 0.005$ , but was not significant for the contrast between DLD and ASD (ns). Mean syllabic vocalization ratios among the DLD participants (mean = 0.41, SD = 0.35) fell between the ASD (mean = 0.32, SD = 0.36) and TD participants (mean = 0.84, SD = 0.25), but was not distinguishable from ASD ratios. ASD participants had reliably lower ratios of syllabic vocalizations when compared to TD participants. Based on this small sample of DLD participants, however, lower syllabic vocalization ratios were not specific to ASD.

To assess sensitivity and specificity of these measures, we explored ROC curves and the area under the curve (AUC) for the binary group contrasts (ASD vs. TD, ASD vs. DLD, DLD vs. TD) for number of nonsyllabic vocalizations and ratio of syllabic vocalizations. For nonsyllabic vocalizations, the number of nonsyllabic vocalizations produced during the 6-min presentation discriminated between the ASD and TD groups (AUC = 87.7, 95% confidence interval [CI]: 77.2–98.2) and between ASD and DLD groups (AUC = 80.4, 95% CI: 64.1–96.7), but not between the DLD and TD groups (AUC = 64.5, 95% CI: 40.7–88.3) (Fig. 4). Because the DLD group is quite small, but important for determining the utility of this measure, we collapsed across the TD and DLD groups to determine an optimal cut point. A cut point of 9 was determined to have positive predictive value of 84.21% and negative predictive value of 81.81% with AUC = 85.80 (95% CI: 74.8–96.7).

While the syllabic vocalization ratio was also highly discriminating for TD versus ASD (AUC = 85.3, 95% CI: 73.8–96.9), and TD versus DLD (AUC = 87.2, 95% CI: 73.7–100), it was not better than chance for the DLD versus ASD contrast (AUC = 57.4, 95% CI: 31.0–81.8) (Fig. 5). Once again, we collapsed across the DLD and TD groups to establish an appropriate cut point. The optimal cut point was identified at 0.50 for the syllabic vocalization ratio. Using this cut point, the positive predictive value was 80.00% while negative predictive value was 72.72%. The AUC for the binary contrast between ASD and non-ASD was 77.90 (95% CI: 65.00–90.80).

Using the already dichotomous presence of words factor and the calculated optimal cut points for the number of nonsyllabic vocalizations and the syllabic vocalization ratios, we ran logistic regressions for each of these three predictors on the ASD versus non-ASD (TD + DLD) comparisons. Participants who did not produce a word during the 6-min presentation were 11 times more likely to have an ASD diagnosis (OR = 11.2, 95% CI: 3.0–41.3,  $P < 0.001$ ). Similarly, participants with syllabic vocalization ratios less than 50% were 10 times more likely to have a diagnosis of ASD (OR = 10.7, 95% CI: 2.9–39.0,  $P < 0.001$ ). Children who produced more than nine nonsyllabic vocalizations during this period were 24 times more likely to have a diagnosis of ASD (OR = 24.0, 95% CI: 5.3–109.5,  $P < 0.001$ ).

Finally, we explored whether these measures were related to ADOS-2 scores for participants in this sample. ADOS-2 scores were available for all but one of the participants in the ASD group (who had a clinical diagnosis) and also for two DLD participants and one TD

participant who were evaluated on the ADOS-2 but did not meet criteria for ASD. Of the three factors explored (presence of words, syllabic vocalization ratio, and number of nonsyllabic vocalizations), only the number of nonsyllabic vocalizations produced was significantly correlated with ADOS-2 scores. Specifically, the number of nonsyllabic vocalizations produced was significantly correlated with restricted, repetitive behaviors,  $r(22) = 0.49$ ,  $P = 0.01$  and total score,  $r(22) = 0.44$ ,  $P = 0.03$  on the ADOS-2. No other correlations were significant.

## Discussion

Reliable, accessible, and objective methods are needed to identify children at increased risk for ASD early in development. Given the heterogeneity of the disorder, no single measure is likely to provide a solution to this problem. Rather, an assessment tool that can elicit several risk behaviors identifiable outside of the laboratory, in a short period of time, and with the potential for scalable automated processing has the potential to improve early detection by identifying the children most in need of formal evaluation. In the present study, we expanded on previously published work demonstrating that a tablet-based app can reliably measure several ASD risk behaviors [Bovery et al., 2018; Campbell et al., 2019; Dawson & Sapiro, 2019; Hashemi et al., 2014, 2018]. Here, we demonstrated that vocalizations uttered during this 6-min procedure provide sufficient data for evaluating vocalizations in young children and distinguished 16- to 31-month-old toddlers with and without ASD. These measures, coupled with the demonstrated sensitivity of social attention and response to name in this procedure, have the potential to increase our ability to reliably identify children who require formal evaluation for ASD.

Our finding that syllabic vocalization ratios can reliably distinguish between children who will and will not go on to receive a diagnosis of ASD extends previous work by Patten et al. [2014] on canonical babbling ratios and Plumb and Wetherby [2013] on atypical speech production. In the Patten et al. study, home videos were solicited and carefully edited to find relevant segments for analysis of vocalizations at 9–12 and 15–18 months. Patten et al. demonstrated that TD infants were 17 times more likely than ASD participants to be in the canonical babbling stage at 9–12 months and six times more likely at 15–18 months. Consistent with the Patten et al. findings, we showed that at 16–31 months, children who had not achieved syllabic vocalization ratios of 0.50 were 10 times more likely to be diagnosed with ASD than TD participants. Inclusion of a DLD group in the current study allowed us to explore the specificity of this measure for ASD outcomes. As was true in work by Plumb and Wetherby [2013], though syllabic vocalization ratio distinguished between TD and ASD participants, it did not reliably differentiate between the ASD and DLD participants. The syllabic vocalization ratios for DLD participants fell between the TD and ASD groups. With a sample of only eight participants in the DLD group, further research will be necessary to determine whether this measure is specific to ASD.

Syllabic vocalization status seems to be functioning on par with a binary distinction between children who did and did not produce words during this protocol. Given the coding procedure employed (syllables contained within words were deemed syllabic), this is to be expected. We have retained the measure of syllabic vocalization ratio as we anticipate that it



will become more relevant in future studies including more DLD participants, and potentially younger participants in all groups.

Unlike Patten et al. [2014], we did not identify reliable differences in the raw number of utterances (canonical or otherwise) produced by TD and ASD participants in this study. This failure to replicate is likely due to methodological differences in these studies. Patten et al. selected segments from home videos, whereas the current study used videos collected in the primary care or laboratory setting using highly controlled stimuli. Infants in their home setting may be more likely to vocalize than infants in less familiar contexts. Previous evidence also suggests that the presence of a stranger may reduce volubility in infants [Iyer, Denson, Lazar, & Oller, 2016]. It is also possible that the age difference between these samples contributed to this discrepancy. However, Warren et al. tested 16- to 48-month-olds and also found low volubility in children with ASD relative to TD children [Warren et al., 2010]. Like the Patten et al., sample, Warren and colleagues based their measures on recordings of children in their natural environments. Chenausky et al. [2017] also found lower vocalization rates in toddlers with ASD, but used a much longer sample for data collection (30 min).

Our finding that nonsyllabic speech rates were higher in the ASD group than in either the TD or DLD groups is consistent with previous findings [Plumb & Wetherby, 2013; Schoen et al., 2011; Sheinkopf et al., 2000]. This measure emerged in the current study as unique in distinguishing not only between TD and ASD participants, but also between ASD and DLD participants. Given the small number of DLD participants, this result should be interpreted with caution and further study will be necessary to confirm this pattern. As noted by Sheinkopf et al. [2000], atypical vocalizations are an intriguing measure for assessing risk for ASD because unlike many existing measures (lack of attention to faces, lack of gestures), they provide a positive symptom for assessing risk that may help us predict who should be evaluated formally.

An unresolved question is why children with ASD would demonstrate such significant delays in vocal development. For speechlike vocalizations, a feature that develops by 10 months in TD children, to be absent in children up to 2.5 years is a rather robust distinction. One possibility is that initial deficits in syllabic vocalizations contribute to a social feedback loop in which parents are less likely to respond to their child's non-speechlike vocalizations, thus perpetuating the delays [Chenausky et al., 2017; Warlaumont, Richards, Gilkerson, & Oller, 2014]. An alternative interpretation is that delays in syllabic vocalizations are related to motor impairments. While motor impairments are not currently considered a core symptom of ASD [American Psychiatric Association, 2013], they are increasingly being recognized as associated with the disorder [Fournier, Hass, Naik, Lodha, & Cauraugh, 2010; Ming, Brimacombe, & Wagner, 2007], though this may not be specific to ASD [Iverson et al., 2019]. Some have argued that these motor impairments account for language delays in autism [Akhtar, Jaswal, Dinishak, & Stephan, 2016], a view that is strengthened by links between rhythmical arm movements and the onset of canonical babbling in TD infants and infants at risk for ASD [Ejiri, 1998; Iverson & Fagan, 2004; Iverson & Wozniak, 2007]. Indeed, the significant correlation between nonsyllabic vocalizations and restricted, repetitive behaviors on the ADOS-2 lend support to this interpretation.

Limitations of the current work include the small sample size (particularly in the DLD group), lack of assessment for test-retest reliability, and lack of IQ measures to assess whether these results may be related to broader developmental delays. This is a proof-of-concept study that requires further work with larger samples to confirm its utility and reliability. Future work will allow us to determine whether these results scale up in a larger sample with a more representative number of children. The age at which these participants were evaluated could be considered an additional limitation of this study. Though 16–31 months is an improvement over the current median age of diagnosis by approximately 2 years [Baio et al., 2018], it is quite late to be assessing syllabic vocalizations and may reflect the specificity of this measure to children with ASD and significant language delays. Efforts to explore the utility of this approach in younger infants are underway. The finding by Patten et al. that vocal maturity was a stronger predictor of ASD outcomes at 9–12 months than at 15–18 months suggests that the approach may hold up in younger samples.

While this work is promising as a contribution to the range of measures of risk for ASD we can assess in the primary care setting, challenges in scalability remain. For this measure to be broadly accessible, the coding process will need to be automated. Although evidence from Oller et al. [2010] suggest that automated detection of vocal maturity is feasible, methods to date have relied on full-day recordings of child vocalizations using a wearable device. Here we used vocal samples from a 6-min tablet-based app to demonstrate that we can detect differences in vocalizations in this context. We are now working to automate this process.

The current study adds to a growing body of evidence that early vocalizations differ among children with ASD. Further, it suggests that vocalizations can be measured during a 6-min procedure administered in the community in the context of standard pediatric care. Most intriguing, this measure seems to distinguish not only between ASD and TD, but also between ASD and DLD (though this is based on a small sample and replication will be critical). When combined with the demonstrated predictive power of this tablet-based app that utilizes carefully designed movies and computer vision analysis to assess behavior, this approach has the potential to reduce disparities in screening for ASD in the general population and may eventually help identify children at risk for the disorder at much younger ages than current methods allow.

## Acknowledgments

Funding for this work was provided by NICHD P50HD093074 (Dawson and Kollins, Co-PIs) and Duke Department of Psychiatry and Behavioral Sciences PRIDE award (Dawson, PI). G.S. is supported by DoD, National Institutes of Health, and National Science Foundation. We thank Elena Patten, PhD for the coding training and Michael Babyak, PhD for consultation.

### Conflict of Interest

Guillermo Sapiro has received basic research gifts from Amazon, Google, Cisco, and Microsoft. Geraldine Dawson is on the Scientific Advisory Boards of Janssen Research and Development, Akili, Inc., LabCorp, Inc., Tris Pharma, and Roche Pharmaceutical Company, a consultant for Apple, Inc, Gerson Lehrman Group, Guidepoint, Inc., Teva Pharmaceuticals, and Axial Ventures, has received grant funding from Janssen Research and Development. Dawson has developed technology that has been licensed and Dawson and Duke University have benefited financially. Dawson and Sapiro are CEOs of DASIO, LLC. Dawson, Sapiro, Carpenter, and Hashemi helped develop the technology that was used in this study. The technology has been licensed and Dawson, Sapiro, Carpenter, Hashemi,

and Duke University have benefited financially. Dawson receives royalties from Guilford Press, Springer, and Oxford University Press.

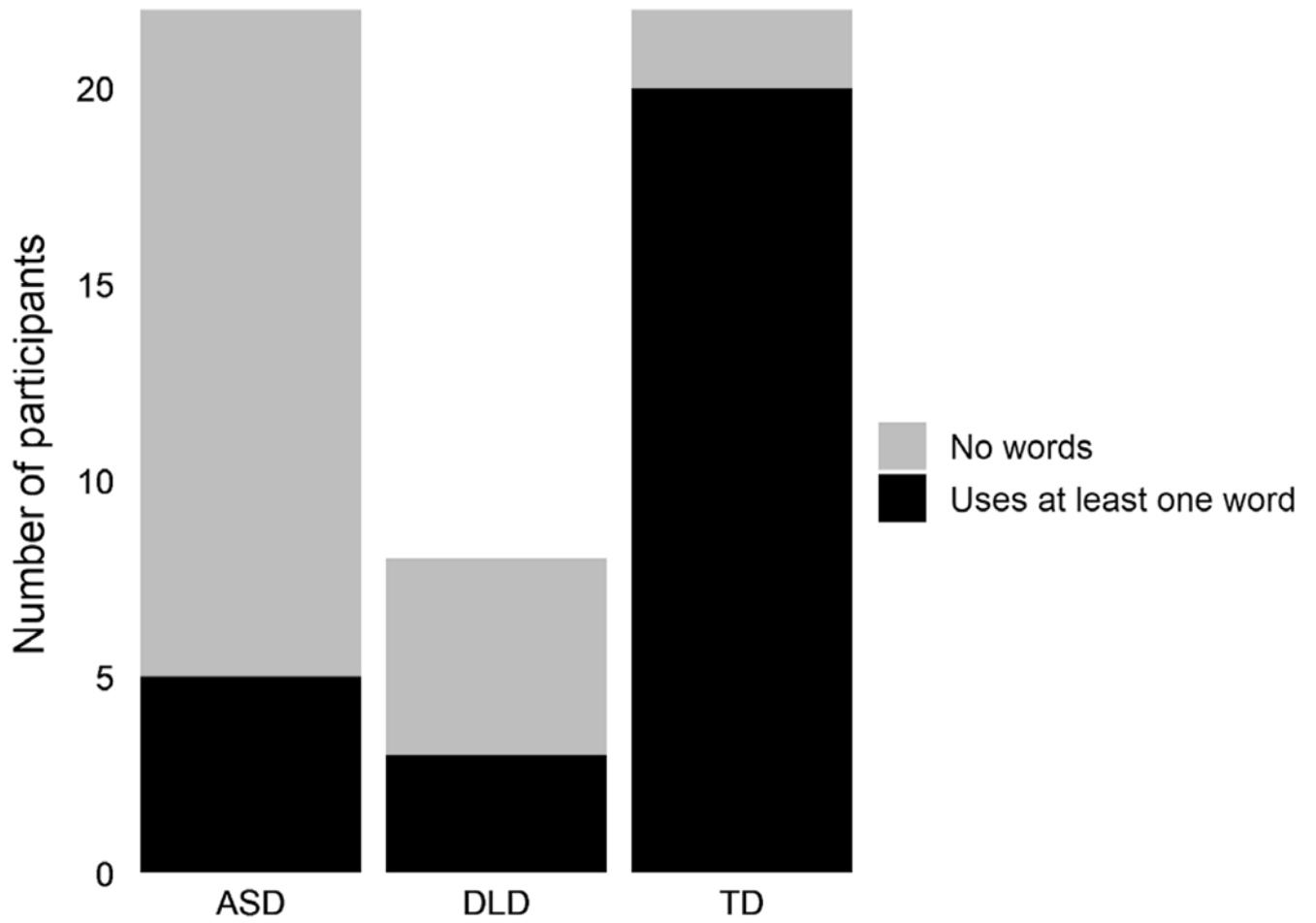
## References

- Akhtar N, Jaswal VK, Dinishak J, & Stephan C (2016). On social feedback loops and cascading effects in autism: A commentary on Warlaumont, Richards, Gilkerson, and Oller (2014). *Psychological Science*, 27(11), 1528–1530. [PubMed: 27664192]
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, ... White T. (2018). Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveillance Summaries*, 67(6), 1–23.
- Bass-Ringdahl SM (2010). The relationship of audibility and the development of canonical babbling in young children with hearing impairment. *Journal of Deaf Studies and Deaf Education*, 15(3), 287–310. [PubMed: 20457674]
- Belardi K, Watson LR, Faldowski RA, Hazlett H, Crais E, Baranek GT, ... Oller DK. (2017). A retrospective video analysis of canonical babbling and volubility in infants with Fragile X Syndrome at 9–12 months of age. *Journal of Autism and Developmental Disorders*, 47(4), 1193–1206. [PubMed: 28247019]
- Bovery MD, Carpenter KL, Hashemi J, Chang Z, Dawson G, & Sapiro G (2018). A scalable off-the-shelf framework for measuring children's direction of attention in ASD. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(10), S233.
- Campbell K, Carpenter KL, Hashemi J, Espinosa S, Marsan S, Borg JS, ... Adler E (2019). Computer vision analysis captures atypical attention in toddlers with autism. *Autism*, 23(3), 619–628. [PubMed: 29595333]
- Chapman KL, Hardin-Jones M, Schulte J, & Halter KA (2001). Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research*, 44(6), 1268–1283.
- Chawarska K, Macari S, & Shic F (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, 74(3), 195–203. 10.1016/j.biopsych.2012.11.022 [PubMed: 23313640]
- Chenausky K, Nelson C III, & Tager-Flusberg H (2017). Vocalization rate and consonant production in toddlers at high and low risk for autism. *Journal of Speech, Language, and Hearing Research*, 60(4), 865–876.
- Clifford SM, Hudry K, Elsabbagh M, Charman T, Johnson MH, & Team B (2013). Temperament in the first 2 years of life in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43 (3), 673–686. [PubMed: 22918859]
- Dawson G (2010). Recent advances in research on early detection, causes, biology, and treatment of autism spectrum disorders. *Current Opinion in Neurology*, 23(2), 95–96. [PubMed: 20216345]
- Dawson G, Campbell K, Hashemi J, Lippmann SJ, Smith V, Carpenter K, ... Baker J (2018). Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Scientific Reports*, 8(1), 17008. [PubMed: 30451886]
- 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. [PubMed: 30715131]
- Egger HL, Dawson G, Hashemi J, Carpenter KL, Espinosa S, Campbell K, ... Tepper M (2018). Automatic emotion and attention analysis of young children at home: A Research kit autism feasibility study. *NPJ Digital Medicine*, 1(1), 20. [PubMed: 31304303]
- Eilers RE, & Oller D (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *The Journal of Pediatrics*, 124(2), 199–203. [PubMed: 8301422]
- Ejiri K (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica*, 55(4), 226–237. [PubMed: 9933781]

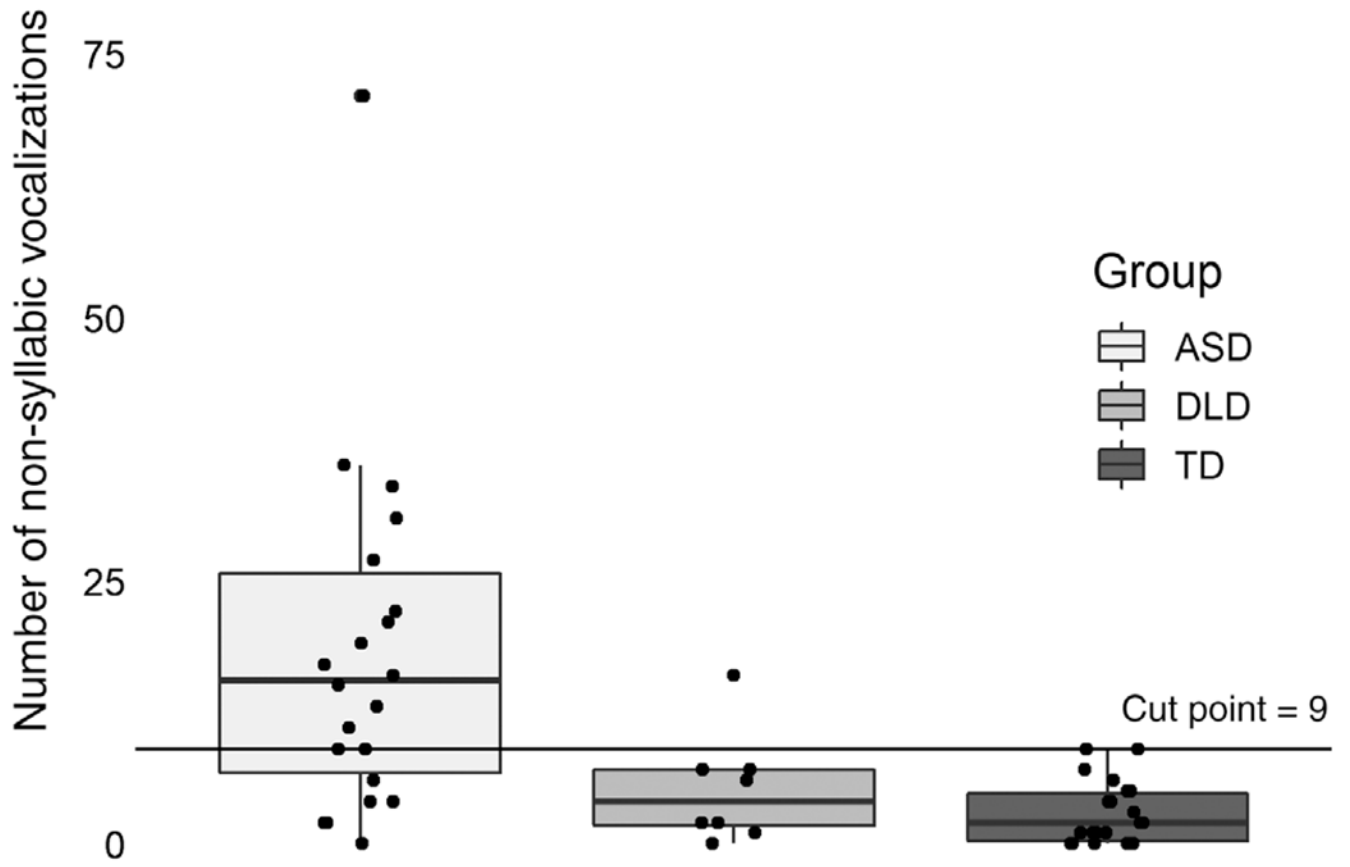
- Elison JT, Paterson SJ, Wolff JJ, Reznick JS, Sasson NJ, Gu H, ... Evans AC. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *American Journal of Psychiatry*, 170(8), 899–908.
- Elsabbagh M, Mercure E, Hudry K, Chandler S, Pasco G, Charman T, Pickles A, Baron-Cohen S, Bolton P., Johnson MH. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, 22 (4) , 338–342. [PubMed: 22285033]
- Elsabbagh M, Fernandes J, Webb SJ, Dawson G, Charman T, Johnson MH, & British Autism Study of Infant Siblings Team. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, 74(3), 189–194. [PubMed: 23374640]
- Fournier KA, Hass CJ, Naik SK, Lodha N, & Cauraugh JH (2010). Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *Journal of Autism and Developmental Disorders*, 40(10), 1227–1240. [PubMed: 20195737]
- Guthrie W, Wallis K, Bennett A, Brooks E, Dudley J, Gerdes M, ... & Miller JS. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*, 144(4), e20183963. [PubMed: 31562252]
- Hashemi J, Dawson G, Carpenter KL, Campbell K, Qiu Q, Espinosa S, ... Sapiro G (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing*. 10.1109/TAFFC.2018.2868196.
- Hashemi J, Tepper M, Vallin Spina T, Esler A, Morellas V, Papanikolopoulos N, ... Sapiro G (2014). Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism Research and Treatment*, epub. 10.1155/2014/935686.
- Hirai AH, Kogan MD, Kandasamy V, Reuland C, & Bethell C (2018). Prevalence and variation of developmental screening and surveillance in early childhood. *JAMA Pediatrics*, 172(9), 857–866. [PubMed: 29987317]
- Hutman T, Chela MK, Gillespie-Lynch K, & Sigman M (2012). Selective visual attention at twelve months: Signs of autism in early social interactions. *Journal of Autism and Developmental Disorders*, 42(4), 487–498. [PubMed: 21519953]
- Iverson JM, & Fagan MK (2004). Infant vocal-motor coordination: Precursor to the gesture–speech system? *Child Development*, 75(4), 1053–1066. [PubMed: 15260864]
- Iverson JM, Shic F, Wall CA, Chawarska K, Curtin S, Estes A, ... Levin AR. (2019). Early motor abilities in infants at heightened versus low risk for ASD: A Baby Siblings Research Consortium (BSRC) study. *Journal of Abnormal Psychology*, 128(1), 69–80. [PubMed: 30628809]
- Iverson JM, & Wozniak RH (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders*, 37(1), 158–170. [PubMed: 17191097]
- Iyer SN, Denson H, Lazar N, & Oller DK (2016). Volubility of the human infant: Effects of parental interaction (or lack of it). *Clinical Linguistics & Phonetics*, 30(6), 470–488. [PubMed: 27002533]
- Johnson CP, & Myers SM (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, 120(5), 1183–1215. [PubMed: 17967920]
- Koegel LK, Koegel RL, Ashbaugh K, & Bradshaw J (2014). The importance of early identification and intervention for children with or at risk for autism spectrum disorders. *International Journal of Speech-Language Pathology*, 16(1), 50–56. [PubMed: 24328352]
- Landa RJ, & Kalb LG (2012). Long-term outcomes of toddlers with autism spectrum disorders exposed to short-term intervention. *Pediatrics*, 130(Suppl. 2), S186–S190. [PubMed: 23118250]
- Lee C-C, Jhang Y, Relyea G, Chen L-M, & Oller DK (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development*, 50, 140–153. [PubMed: 29289753]
- López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, & Gude-Sampedro F (2014). *OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests*. *Journal of Statistical Software*, 61(8), 1–36.
- Lord C, Rutter M, DiLavore P, Risi S, Gotham K, & Bishop S (2012). *Autism diagnostic observation schedule—2nd edition (ADOS-2)*. Los Angeles, CA: Western Psychological Corporation.
- Lynch MP, Oller DK, Steffens ML, & Levine SL (1995). Onset of speech-like vocalizations in infants with Down syndrome. *American Journal on Mental Retardation*, 100(1), 68–86. [PubMed: 7546639]

- Ming X, Brimacombe M, & Wagner GC (2007). Prevalence of motor impairment in autism spectrum disorders. *Brain and Development*, 29(9), 565–570. [PubMed: 17467940]
- Oller D, Eilers RE, & Basinger D (2001). Intuitive identification of infant vocal sounds by parents. *Developmental Science*, 4(1), 49–60.
- Oller D, Levine S, Cobo-Lewis AB, Eilers RE, Pearson BZ, & Paul R (1998). Vocal precursors to linguistic communication: How babbling is connected to meaningful speech. In *Exploring the speech-language connection*. Baltimore, MD: Paul H. Brookes Publishing Co pp. 1–23.
- Oller D, Niyogi P, Gray S, Richards J, Gilkerson J, Xu D, ... Warren S (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, 107(30), 13354–13359.
- Overby M, & Caspari SS (2015). Volubility, consonant, and syllable characteristics in infants and toddlers later diagnosed with childhood apraxia of speech: A pilot study. *Journal of Communication Disorders*, 55, 44–62. [PubMed: 25934424]
- Ozonoff S, Iosif A-M, Baguio F, Cook IC, Hill MM, Hutman T, ... Sigman M (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(3), 256–266.e252. [PubMed: 20410715]
- Patten E, Belardi K, Baranek GT, Watson LR, Labban JD, & Oller D (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders*, 44(10), 2413–2428. [PubMed: 24482292]
- Paul R, Fuerst Y, Ramsay G, Chawarska K, & Klin A (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry*, 52(5), 588–598. [PubMed: 21039489]
- Plumb AM, & Wetherby AM (2013). Vocalization development in toddlers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 56 (2), 721–734.
- Rescorla L, & Ratner NB (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*, 39(1), 153–165.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, & Müller M (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(1), 77. [PubMed: 21414208]
- Robins DL (2008). Screening for autism spectrum disorders in primary care settings. *Autism*, 12(5), 537–556. [PubMed: 18805946]
- Robins DL, Casagrande K, Barton M, Chen C-MA, Dumont-Mathieu T, & Fein D (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*, 133(1), 37–45. [PubMed: 24366990]
- Rogers SJ, Estes A, Lord C, Munson J, Rocha M, Winter J, ... Vismara L (2019). A multisite randomized controlled two-phase trial of the early Start Denver Model compared to treatment as usual. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(9), 853–865. [PubMed: 30768394]
- Rogers SJ, Vismara L, Wagner A, McCormick C, Young G, & Ozonoff S (2014). Autism treatment in the first year of life: A pilot study of infant start, a parent-implemented intervention for symptomatic infants. *Journal of Autism and Developmental Disorders*, 44(12), 2981–2995. [PubMed: 25212413]
- Rozga A, Hutman T, Young GS, Rogers SJ, Ozonoff S, Dapretto M, & Sigman M (2011). Behavioral profiles of affected and unaffected siblings of children with autism: Contribution of measures of mother-infant interaction and nonverbal communication. *Journal of Autism and Developmental Disorders*, 41(3), 287–301. [PubMed: 20568002]
- Schoen E, Paul R, & Chawarska K (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research*, 4(3), 177–188. [PubMed: 21308998]
- Sheinkopf SJ, Iverson JM, Rinaldi ML, & Lester BM (2012). Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Research*, 5(5), 331–339. [PubMed: 22890558]
- Sheinkopf SJ, Mundy P, Oller DK, & Steffens M (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, 30(4), 345–354. [PubMed: 11039860]

- R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria  
Retrieved from: R Foundation for Statistical Computing <https://www.R-project.org/>
- Warlaumont AS, Richards JA, Gilkerson J, & Oller D (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, 25(7), 1314–1324. [PubMed: 24840717]
- Warren SF, Gilkerson J, Richards JA, Oller DK, 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. [PubMed: 19936907]
- Windham GC, Smith KS, Rosen N, Anderson MC, Grether JK, Coolman RB, & Harris S (2014). Autism and developmental screening in a public, primary care setting primarily serving hispanics: Challenges and results. *Journal of Autism and Developmental Disorders*, 44(7), 1621–1632. [PubMed: 24408091]
- Wolff JJ, Gu H, Gerig G, Elison JT, Styner M, Gouttard S, Botteron KN, Dager SR, Dawson G, ... Piven J (2012). Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. *American Journal of Psychiatry*, 169(6), 589–600.
- Yankowitz LD, Schultz RT, & Parish-Morris J (2019). Pre- and paralinguistic vocal production in ASD: Birth through school age. *Current psychiatry reports*, 21(12), 126. [PubMed: 31749074]
- Zwaigenbaum L, Bryson S, Rogers T, Roberts W, Brian J, & Szatmari P (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, 23(2), 143–152. [PubMed: 15749241]

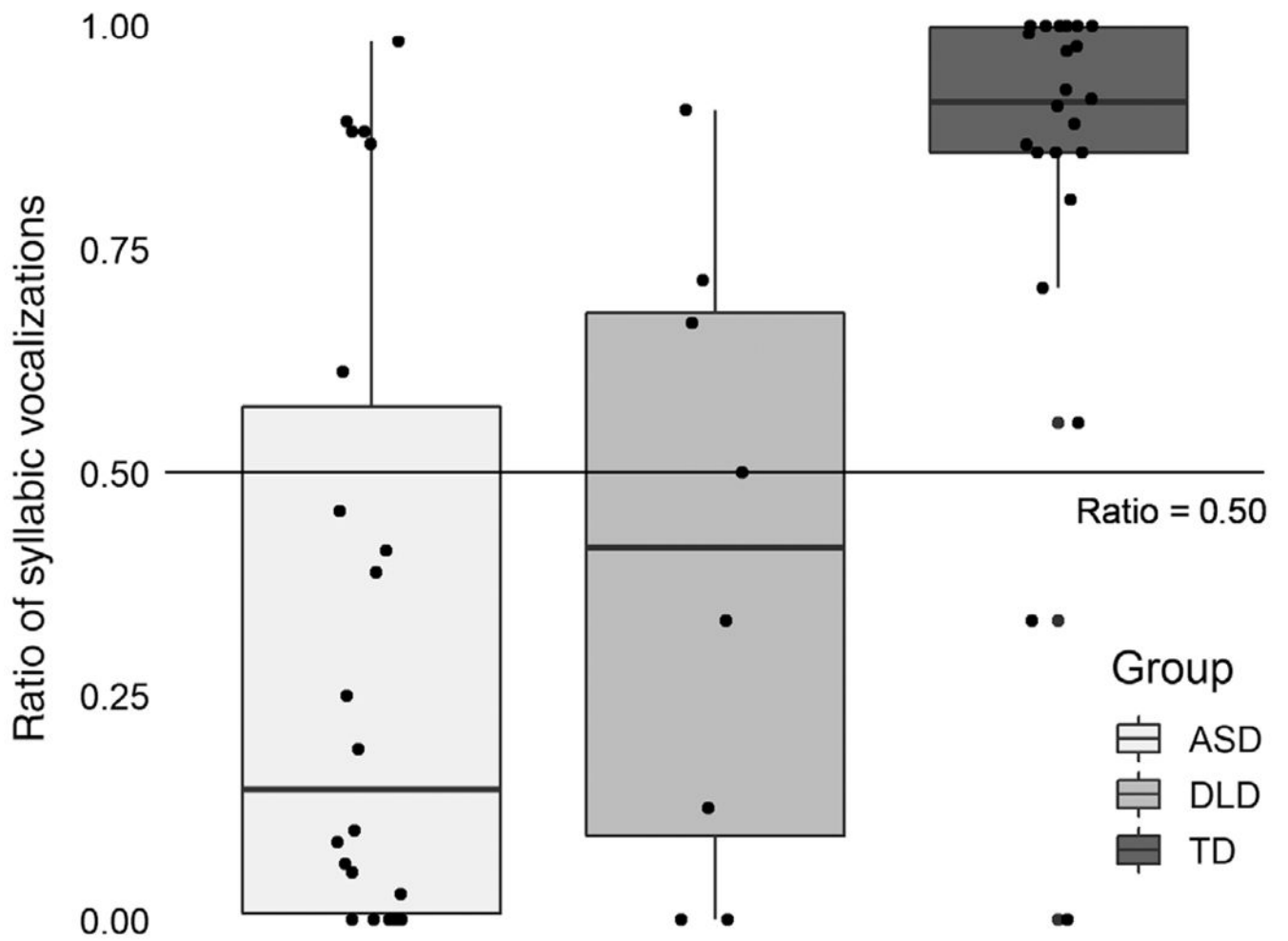


**Figure 1.** Presence of words. Count of participants producing at Least one word during the administration of the app.

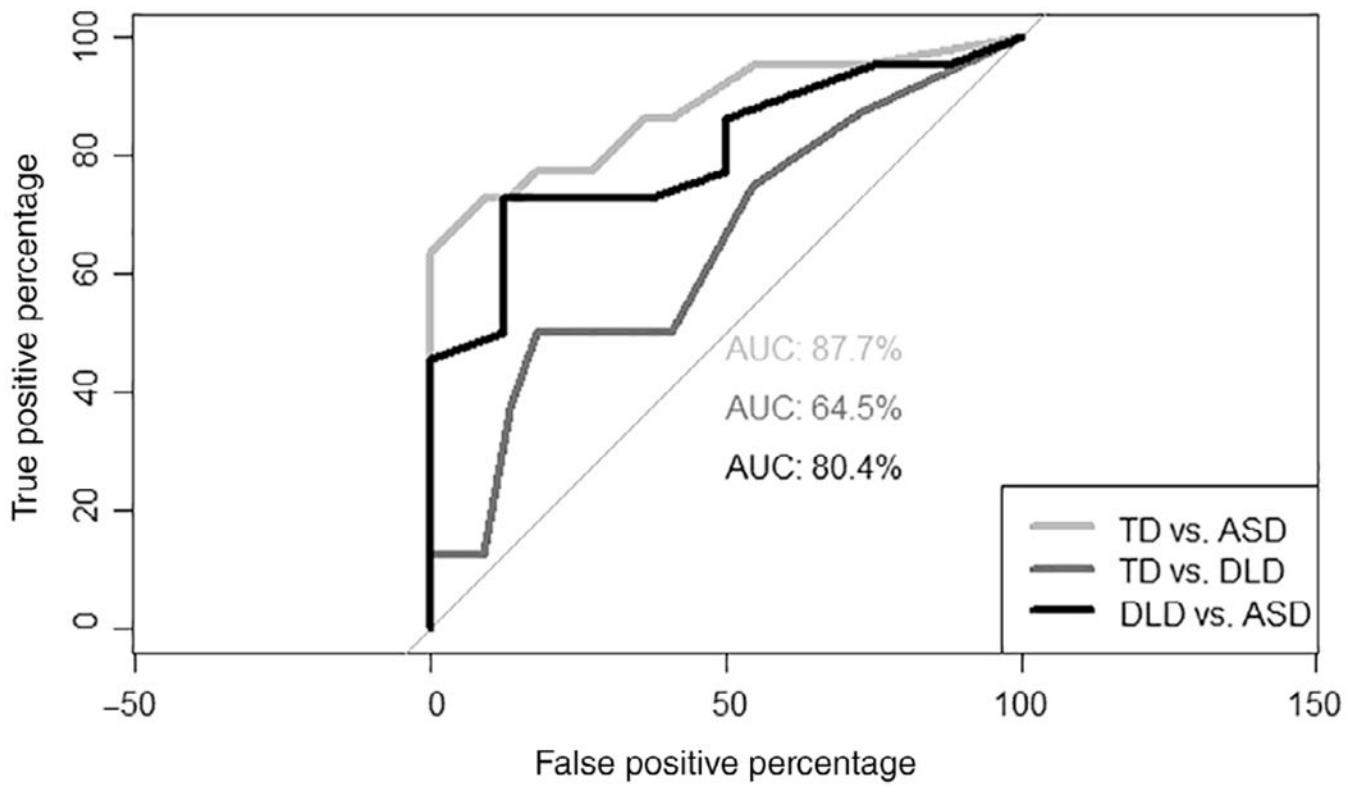


**Figure 2.**  
Number of nonsyllabic vocalizations by group.





**Figure 3.**  
Ratio of syllabic vocalizations to aLL vocalizations by group.



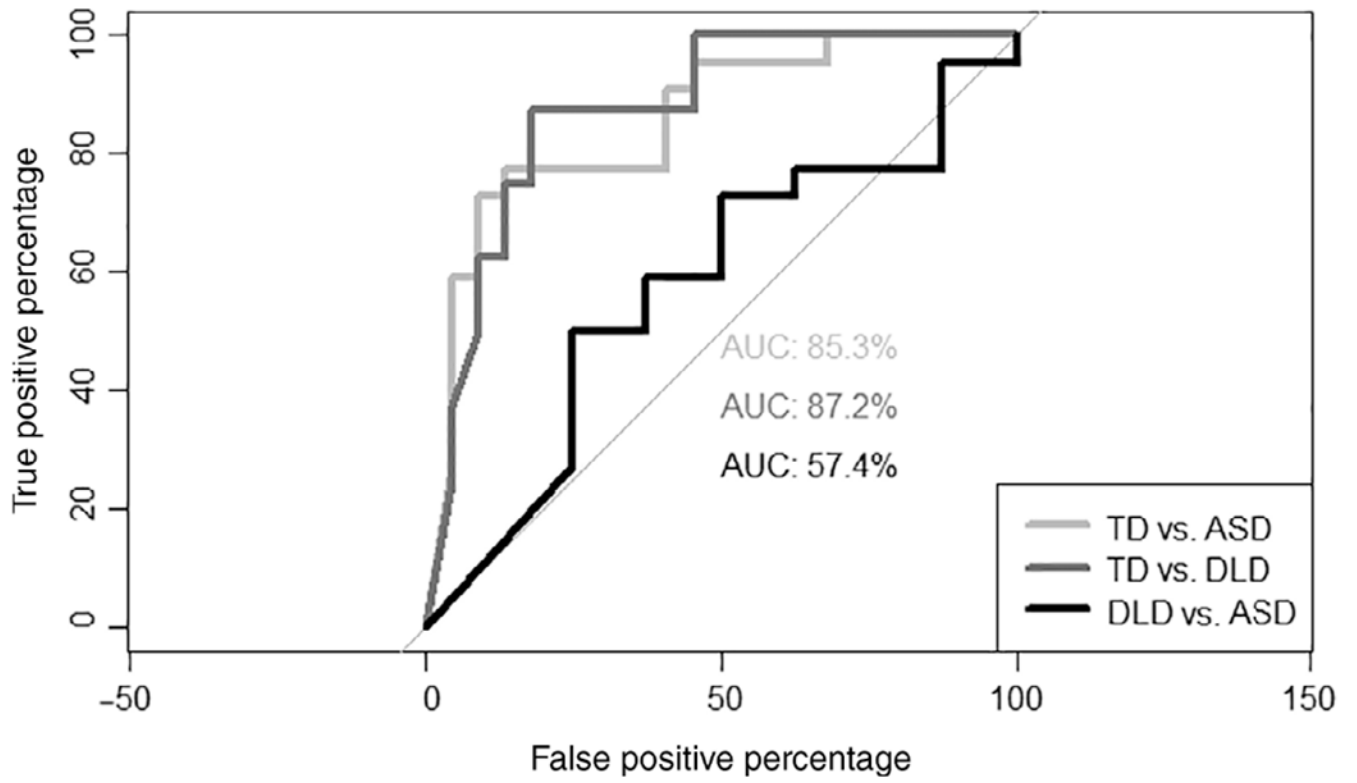
**Figure 4.** Receiver operating characteristic curves and area under the curve for number of nonsyllabic vocalizations by contrast.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5.** Receiver operating characteristic curves and area under the curve for ratio of syllabic vocalizations by contrast.

**Table 1.**

Demographic Information for Participants

	ASD	TD	DLD
Male/female	5/17	5/17	5/3
Age ( <i>SD</i> )	26.19 (4.07)	25.66 (2.39)	23.90 (3.65)
	13.6	13.6	12.5
African American			
Asian	4.5	4.5	0
Caucasian	45.5	68.2	25
Multiracial.	4.5	4.5	37.5
Race %			
Other/unknown	31.8	9	25
Hispanic	4.5	0	0
Non-Hispanic	72.7	90.9	87.5
Ethnicity			
Unknown	22.7	9.1	12.5
Low risk	4	22	8
Medium risk	8	0	0
High risk	10	0	0
M-CHAT-R/F			
ADOS-2 ( <i>SD</i> )	18.81 (4.20)		
	Total score (toddler module)		

Rates of Nonsyllabic Vocalizations, Canonical Syllables, Syllabic Vocalizations, Word Types, Word Tokens, and Syllabic Vocalization Ratios by Group

**Table 2.**

	ASD <i>n</i> = 22		TD <i>n</i> = 22		DLD <i>n</i> = 8	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
# Vocalizations	41.6 (39.3)	0–134	37.2 (43.0)	2–168	10.4 (8.5)	0–24
# Nonsyllabic vocalizations	20.9 (22.1)	0–90	2.8 (2.9)**	0–9	5.1 (5.2)*	0–16
# Canonical syllables	9.8 (15.4)	0–50	4.9 (7.8)	0–32	2.3 (2.9)	0–8
# Syllabic vocalizations	20.8 (34.0)	0–118	34.4 (41.7)	0–163	5.3 (6.4)	0–19
# Word types	3.2 (8.8)	0–38	8.7 (8.2)	0–31	1.0 (1.6)	0–4
# Word tokens	7.7 (20.6)	0–71	20.9 (27.0)	0–114	1.9 (3.8)	0–11
Syllabic vocalization ratio	0.32 (0.36)	0–0.98	0.84 (0.25)**	0–1.0	0.41 (0.35)	0–0.9

Note. Significance reflects Bonferroni corrected post hoc comparison between ASD and TD or DLD.

\*  $P < 0.05$ .

\*\*  $P < 0.01$ .