Current Psychiatry Reports (2019) 21:126
https://doi.org/10.1007/s11920-019-1113-1

AUTISM SPECTRUM DISORDERS (ES BRODKIN, SECTION EDITOR)

Pre- and Paralinguistic Vocal Production in ASD: Birth Through School Age

Lisa D. Yankowitz [1,2] · Robert T. Schultz [1,3,4] · Julia Parish-Morris [1,3]

Abstract

Purpose of Review We review what is known about how pre-linguistic vocal differences in autism spectrum disorder (ASD) unfold across development and consider whether vocalization features can serve as useful diagnostic indicators.

Recent Findings Differences in the frequency and acoustic quality of several vocalization types (e.g., babbles and cries) during the first year of life are associated with later ASD diagnosis. Paralinguistic features (e.g., prosody) measured during early and middle childhood can accurately classify current ASD diagnosis using cross-validated machine learning approaches.

Summary Pre-linguistic vocalization differences in infants are promising behavioral markers of later ASD diagnosis. In older children, paralinguistic features hold promise as diagnostic indicators as well as clinical targets. Future research efforts should focus on (1) bridging the gap between basic research and practical implementations of early vocalization-based risk assessment tools, and (2) demonstrating the clinical impact of targeting atypical vocalization features during social skill interventions for older children.

Keywords Autism · Paralinguistics · Prosody · Early diagnosis · Speech production · Acoustic properties

Introduction

Social communication deficits are a core diagnostic feature of autism spectrum disorder (ASD) [1], and paralinguistic aspects of communication (i.e., non-lexical components of the speech signal, such as prosody) may be particularly important. Atypical prosody was part of Kanner's and Asperger's initial descriptions of autism [2, 3], and remains a salient feature of the diagnosis [4, 5••]. Noted in gold standard autism assessment tools, including the Autism Diagnostic Observation Schedule 2 (ADOS-2, [6]) and Autism Diagnostic Interview–Revised (ADI-R, [7]), prosody communicates rich non-lexical information (e.g., emotions, grammar, sarcasm) through pitch, duration, and volume of speech [8]. It can also be defined to include other paralinguistic features (such as voice quality, motor control, pauses, and stress), which have proven useful for describing ASD [9•, 10•].

Given the core relationship between prosody and ASD, along with new tools designed to measure prosody in an automated, quantitative fashion (low-level descriptors of acoustic features over time can be measured with fine-grained accuracy), prosody is a prime candidate for use as an objective marker in screening and diagnostic tools. However, as noted by Fusaroli and colleagues in a recent meta-analysis, promising work in the area of diagnostic prediction based on paralinguistic features has proven difficult to reproduce [5••]. Univariate approaches (i.e., focusing on diagnostic group differences in a single measure, such as pitch or pitch variability) show relatively small effects across studies, with discriminative accuracy of approximately 61–64% [5••]. Multivariate approaches (i.e., combining information across many acoustic features) show reasonably high sensitivity and specificity [11–15] but vary substantially in methodology, making it difficult to compare across studies and draw inferences about clinical relevance.

Topical Collection on *Autism Spectrum Disorders*

✉ Lisa D. Yankowitz
lisayank@sas.upenn.edu

1 Center for Autism Research, Children's Hospital of Philadelphia, 2716 South St, Philadelphia, PA 19104, USA

2 Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, USA

3 Department of Psychiatry, University of Pennsylvania, 3400 Civic Center Blvd, Philadelphia, PA 19105, USA

4 Department of Pediatrics, University of Pennsylvania, 3400 Civic Center Blvd, Philadelphia, PA 19105, USA

Published online: 20 November 2019

Since the review by Fusaroli and colleagues was published in 2017, at least nine new classification studies have emerged. Researchers have also begun to emphasize the use of standardized acoustic feature sets, which is an important first step toward addressing methodological variation and producing studies that are comparable to one another, as well as more reproducible.

Vocal behavior has great potential power to predict later ASD diagnosis, as delays in early speech milestones are often the first cause of concern reported by parents of children with ASD [16]. Non-invasive early markers of ASD risk are something of a holy grail for researchers because early, intensive intervention predicts better outcomes [17, 18], and early intervention cannot be initiated until a need for intervention is identified. Recent work suggests that while it is possible, in some cases, for highly trained clinicians to provide a stable ASD diagnosis at 14 months, the typical age of diagnosis in the USA remains between 4 and 5 years of age [19, 20]. Identifying features of early vocalizations that predict subsequent diagnosis and indicate the need for early intervention could address this critical detection gap. Thus, in this review, we examine recent research on pre-linguistic vocalizations in infants and toddlers who have or who go on to receive an ASD diagnosis, followed by studies of paralinguistic production in pre-school and school-age children on the autism spectrum. Advances in technology and in clinical research make this an opportune time to lower the age of diagnosis through vocal analysis, facilitating earlier intervention and improving outcomes.

Infants and Toddlers

Early vocalizations (e.g., crying, laughing, yelling, squealing, vowels, babbling) are excellent candidates to use as clinically useful biomarkers of ASD. They are easily collected from infants (who produce them freely), are related to a core diagnostic symptom (unusual prosody/atypical communication), and are supported by an emerging literature that documents group differences in infancy and toddlerhood. One common approach to gathering a sample of infants who will go on to receive an ASD diagnosis is by recruiting infants at high risk for ASD, either through genetics—by virtue of having an older sibling with ASD [21]—or who are showing early signs of ASD themselves [22]. Another common approach is to collect and analyze vocalizations produced during retrospective infant home videos of children who currently hold a diagnosis (see [23•] for a review).

In the next three sections, we will review separate lines of evidence indicating various kinds of disruptions in the early vocal behaviors of infants who develop ASD. These sections focus on (1) standardized clinical assessments of expressive language, (2) production rates for different vocalization types, and (3) acoustic features of vocalization. The longitudinal power of early vocalizations to predict later phenotype has only begun to be tested, and the most rigorously cross-validated approaches to prediction have focused primarily on acoustic features. Table 1 provides an overview of these studies.

Standardized Assessments of Expressive Language

For infants later diagnosed with ASD, reliable group-level impairment in expressive language ability (as measured via direct clinical assessment and parent report) is observed by 12 months (see [24] for a meta-analysis of high-risk sibling studies). In longitudinal samples, slower rates of expressive language growth between 6 months and 3 years have been observed in high-risk infant siblings who went on to receive an ASD diagnosis [25, 26]. Difficulties with expressive language are roughly commensurate with receptive language deficits in children aged 1–5 years, according to a recent meta-analysis [27]. Taken together, these findings indicate that reduced language ability emerges early in life for at least a subset of children and adults who develop ASD and is measurable using standardized expressive language assessments.

Early Vocalization Rates

The literature on early vocalizing in ASD is often loosely summarized as indicating reduced overall rates during infancy, but evidence for this claim is mixed. Very few studies report overall lower rates of voluntary vocalizing by infants with ASD across the first few years of life [28, 29], and several studies report no difference between infants that do and do not develop ASD [30, 31, 32••, 33]. Evidence is somewhat stronger that ASD infants and toddlers produce fewer speech-like vocalizations, more non-speech and atypical vocalizations, or a reduced ratio between them [22, 28, 30, 33–43] than control infants, but there are also null [30, 33, 34, 39, 40, 44] and even opposing findings [21, 45]. No clear patterns of study characteristics (age, sex, diagnostic grouping, setting) that might explain the discrepancies emerge from a review of this literature. One possible consequence of reductions in speech-like vocalizing (or lower proportions of speech-like vs. other types of vocalizing) in ASD is that social feedback loops with caregivers might be negatively impacted. For example, an analysis of conversational turns between 8-to-48-month-olds and their caregivers in day-long recordings revealed a social feedback loop for all children: Adults were more likely to respond to speech-like vocalizations than non-speech vocalizations, and toddlers in turn were more likely to produce speech-like vocalizations after an adult response [28]. However, this loop was disrupted in ASD, such that toddlers with ASD produced a lower ratio of speech-like vocalizations and received responses from parents that were less contingent on the type

Table 1 Overview of studies of infant and toddler vocalizations in ASD. *ASD* autism spectrum disorder, *Pre-ASD* infants without a diagnosis at the time of vocalization recording who were later diagnosed, *TD* typically developing, *DD* developmental delayed, *ID* intellectual disability, *LD* language delay, *HR* high risk (diagnostic outcome not reported), *HR+* high risk eventually diagnosed with ASD, *HR−* high risk without an ASD diagnosis, *LR* low risk, *CSBS-DP* Communication and Symbolic Behavior Scales Developmental Profile, *FSP of LTAS* first spectral peak of the long-term average spectrum. Effect sizes are reported as calculated by original authors where possible. When not reported, effect sizes were calculated as Cohen's *d*. When data were non-normal and *z*-scores from non-parametric tests were reported, Cohen's *d* was calculated by first calculating the point-biserial correlation and converting it to Cohen's *d* (Ivarsson et al. 2013)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
Apicella, F., Chericoni, N., Costanzo, V., Baldini, S., Billeci, L., Cohen, D., & Muratori, F. (2013)	0–6 months 6–12 months	Pre-ASD: 10 TD: 9	Home videos	Home videos	Pre-ASD < TD: 6–12 month directed vocalizations ($d = -1.54$) Pre-ASD = TD: 0–6 month directed vocalizations ($d = -.12$)	
Bornstein, M., Costlow, K., Truzzi, A., & Esposito, G. (2016)	5 months	Pre-ASD: 10 TD: 10	Home videos	Home videos	ASD > TD: F0 of cry ($d = 2.15$)	Response time for categorizing ASD cry is slower than TD cry ($d = 2.31$)
Brisson, J., Martel, K., Serres, J., Sirois, S., & Adrien, J.-L. (2014)	0–6 months	Pre-ASD: 13 TD: 13	Home videos	Home videos	No difference in pitch ($d = -.15$) or duration ($d = -.0004$) of speech-like utterances Pre-ASD > TD: simple pitch contours ($d = 1.07$) Pre-ASD < TD: complex pitch contours	Mothers of Pre-ASD used shorter utterances ($d = -.88$), with no difference in pitch ($d = .22$) or pitch contours
Chenausky, K., Nelson, C., & Tager-Flusberg, H. (2017)	12, 18, and 24 months	HR+: 10 HR−: 18 LR: 18	High-risk siblings	30 minute samples from AOSI at 12 months, ADOS at 18 and 24	No difference in rate of non-speech-like vocalizations HR+ < HR−,LR: speech-like vocalizations ($ds = -1.05, -.96$) Age-by-group interaction predicting number of consonants (HR− < LR at 12 months). No main effect of group	
Chenausky, K., & Tager-Flusberg, H. (2017)	18, 24, and 36 months	HR+: 11 HR−: 22 LR: 22	High-risk siblings	First 30 minutes of ADOS	HR+ infants less likely than LR to produce distinct /b/ and /p/ at 36 months, by factor of 0.2 No group difference in word approximation frequency	
Chericoni, N., de Brito Wanderley, D., Costanzo, V., Diniz-Gonçalves, A., Leitgel Gille, M., Parlato, E., … Muratori, F. (2016)	0–6, 6–12, 12–18 months	Pre-ASD: 10 TD: 10	Home videos	Home videos	Pre-ASD < TD: vocalizations (vowels, non-reduplicated consonants and vowels, fusses) at 6–12 months ($d = -1.44$), no difference 0–6 ($d = -.08$) or 12–18 ($d = -.47$) No group difference in rate of babbling. Qualitative difference in directed babbling (less face-gazing in pre-ASD)	
Cohen, D., Cassel, R. S., Saint-Georges, C., Mahdhaoui, A., Laznik,	0–6, 6–12, and 12–18 months	Pre-ASD: 14 TD: 14	Home videos	Home videos	No difference in vocalization responses to parent speech	

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
M.-C., Apicella, F., … Chetouani, M. (2013)						
English MS, Tenenbaum EJ, Levine TP, Lester BM, Sheinkopf SJ. (2019)	1 month	Pre-ASD: 4 TD: 4	Longitudinal study of infants with prenatal substance exposure	Elicited cries recorded in isolette	Qualitatively, Pre-ASD < TD: number of utterances, longer and more variable utterances and pauses, greater energy, and greater percentage of frames with frication.	Pre-ASD cries rated as more distressed, less typical, and reflecting greater pain
Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2012)	13 months	Pre-ASD: 10 TD: 10	Home videos	Home videos	Pre-ASD > TD: F0 of FSP of LTAS ($d = 2.26$)	Japanese and Italian adults both rated Pre-ASD cries > TD cries for expressed distress and felt distress
Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2013)	13 months	Pre-ASD: 10 TD: 10	Home videos	Home videos	Pre-ASD < TD: duration of pauses in cry ($d = -1.72$) Pre-ASD < TD: number of utterances (vocalizations between pauses) ($d = -1.07$) Pre-ASD > TD: F0 of FSP of LTAS ($d = 2.26$)	Parents of TD: rated Pre-ASD cries > TD for expressed distress (Italian: $\eta_p^2 = 0.09$, Japanese: $\eta_p^2 = 0.11$) and felt distress (Italian: $\eta_p^2 = 0.08$, Japanese: $\eta_p^2 = 0.09$). Both were most predicted by shorter pause length
Esposito, G., Rostagno, M. del C., Venuti, P., Haltigan, J. D., & Messinger, D. S. (2014)	15 months	HR: 13 LR: 14	High-risk siblings	Separation phase of the Strange Situation procedure	HR > TD: F0 of whole cry ($d = 1.41$) and first utterance ($d = .99$) HR > TD: F0-max of first utterance ($d = 1.4$) HR < TD: duration of whole cry ($d = -.97$) HR > TD: duration of first utterance ($d = .89$)	
Esposito, G., Valenzi, S., Islam, T., & Bornstein, M. H. (2015)	13 months	Pre-ASD: 10 TD: 10	Home videos	Home videos	Pre-ASD < TD: duration of pauses in cry ($d = -.93$) Pre-ASD < TD: number of utterances (vocalizations between pauses) ($d = -0.66$) Pre-ASD > TD: in F0 of FSP of LTAS ($d = 1.74$)	Fathers and nonfathers more likely to show increased Galvanic Skin Response to Pre-ASD cries than TD cries; no differences in inter-beat interval or right hand temperature change
Esposito, G., & Venuti, P. (2008)	13 and 20 months	12 total (Pre-ASD, ID, TD)	Home videos	Home videos		Adult women were less accurate at guessing the age and cause of crying when listening to cries of pre-ASD, and more likely to report uneasiness and stress
Esposito, G., & Venuti, P. (2009)	12 months	Pre-ASD: 10 DD: 10 TD: 10	Home videos	Home videos	Pre-ASD > DD,TD: proportional duration of scream ($ds = 1.36, 1.33$) Pre-ASD < DD,TD: in aspiration/expiration phase	Mothers of Pre-ASD < DD and TD responding with tactile or vestibular stimulation ($ds = -.95, -1.10$) Mothers of Pre-ASD > DD and TD verbal responding ($ds = .97, .78$)

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
					(*ds* = − 3.86,− 2.82) and proportion of pause (*ds* = − 1.07,− .78)	
Esposito, G., & Venuti, P. (2010, 60)	5 and 18 months	Pre-ASD: 10 DD: 10 TD: 10	Home videos	Home videos	Pre-ASD > DD, TD: F0 at 18 months (*ds* = 0.58, 1.07). No differences at 5 months	
Esposito, G., & Venuti, P. (2010, 61)	18 months	Pre-ASD: 14 DD: 14 TD: 14	Home videos	Home videos	Pre-ASD > DD,TD: F0 at 18 months (*ds* = 0.60, 0.89)	
Esposito, G., Venuti, P., & Bornstein, M. H. (2011)	18 months	Pre-ASD: 18 DD: 18 TD: 18	Home videos	Home videos		Cries of Pre-ASD judged by adults to be more distressing than TD, and less typical than DD and TD
Gabrielsen, T. P., Farley, M., Speer, L., Villalobos, M., Baker, C. N., & Miller, J. (2015)	15–33 months	ASD: 14 LD: 14 TD: 14	Community screening	2 10-min sections of ADOS	ASD < TD,LD: ratio of typical-to-atypical vocalizations (6:1 in ASD, 31:1 LD, 13:1 TD)	
Garrido, D., Watson, L. R., Carballo, G., Garcia-Retamero, R., & Crais, E. R. (2017)	14 months	23-month-ADOS classification Autism: 34 Spectrum: 25 Non-ASD: 23	Screening	10-min parent-child free play 20-min CSBS-DP	Autism < Spectrum and Non-ASD: speech-like volubility during free play (OR = 1.570), no difference during CSBS Autism < Non-ASD: canonical directed vocalizations (OR = 1.039);Autism > Non-ASD: canonical nondirected (OR = .607) and noncanonical nondirected (OR = 1.2) Autism > Spectrum: atypical vocalizations (OR = .915) No differences in distress/pleasure	
Iverson, J. M., & Wozniak, R. H. (2007)	monthly 5–14 months, follow-up 18 months	HR: 21 LR: 18	High-risk siblings	Monthly home videos and laboratory assessments	More HR than LR significantly delayed in reduplicated babble (29% versus 0%) and first words (29% versus 0%)	
Landa, R. J., Gross, A. L., Stuart, E. A., & Faherty, A. (2013)	14, 18, 24 months	HR+ (early): 28 HR+ (late): 26 LR/HR−: 181	High-risk siblings	CSBS-DP	HR+(early) < LR/HR−: consonant inventory at 14, 18, and 24 months (standardized βs: − .8, − .9, − 1.9) HR+(late) < LR/HR−: consonant inventory at 14 and 24 months (std βs: − 0.5, − 0.1, − 0.9)	
Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007)	14 and 24 months	HR+(early dx): 15 HR+(late dx): 13 HR−BAP: 19 HR−: 51 LR: 17	High-risk siblings	CSBS-DP	HR+(early) < all non-ASD: consonant inventory at 14 and 24 months, word inventory at 24 months HR+(late) < HR− and LR: consonant inventory and word	

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
					inventory at 24 months, no differences at 14 months	
Lee K-S, Shin YJ, Yoo H-J, Lee GJ, Ryu J, Son O, et al. (2018)	13–38 months	ASD: 28 DD: 18	Already diagnosed	10-min parent-child free play	ASD > DD: atypical vocalizations, babbling during non-jointly focused attention DD > ASD: babbling during jointly focused attention	
Northrup, J. B., & Iverson, J. M. (2015)	9 months	HR: 25 LR: 10	High-risk siblings	5-min parent-child free play	No difference in total voluntary vocalization frequency, intrapersonal pause duration, latency to respond duration and variability, or percent simultaneous speech HR infants with later LD showed more simultaneous speech and less-coordinated latency to respond than no LD	
Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., … Warren, S. F. (2010)	ASD: 16–48 months LD: 10–44 months TD: 10–48 months	ASD: 77 LD: 49 TD: 106	Already diagnosed	LENA home recordings	ASD differed from TD in 9 of 12 acoustic parameters of speech-related vocalizations, clustered in measures of rhythm/syllabicity and duration Linear Discriminant Analysis of ASD versus TD using acoustic features showed sensitivity/specificity of 0.86 with leave-one-out cross-validation	
Orlandi, S., Manfredi, C., Bocchi, L., & Scattoni, M. L. (2012)	10 days, 6 weeks, 12 weeks	HR: 2 LR: 17	High-risk siblings	Home recordings	Qualitatively, HR < LR: F0 and percentage of voiced cry intervals at each time-point	
Osterling, J. A., Dawson, G., & Munson, J. A. (2002)	12 months	Pre-ASD: 20 (Pre-ASD+ ID: 9) ID: 14 TD: 20	Home videos	Home recordings (first birthday parties)	No difference in percent of time vocalizing between Pre-ASD and TD, or Pre-ASD+ID and ID	
Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., … Young, G. S. (2010)	6, 12, 18, 24, 36 months	Pre-ASD: 25 TD: 25	High-risk siblings	Visual reception subtest of MSEL	Pre-ASD < TD: directed vocalization frequency at 12, 18, 24, and 36 months, no difference at 6 months	
Ozturk, Y., Bizzego, A., Esposito, G., Furlanello, C., & Venuti, P. (2018)	36–52 months	ASD: 8 TD: 7	Unknown	Structured diagnostic assessment	ASD > TD: F0 of cry (d = 3.06)	ASD cries > TD cries reported stress (ASD-parent ES: 0.54, TD-parent ES: 0.78), arousal (ASD-parent ES: 0.71, TD-parent ES: 0.60), and perceived negative valence (ASD-parent ES: 0.64, TD-parent: 0.61)

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
						No differences in inter-beat interval (ASD-parent ES: 0.03, TD-parent ES: 0.04)
Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014)	9–12 and 15–18 months	Pre-ASD: 23 TD: 14	Home videos	Home videos	TD more likely to be in canonical stage at 9–12 (OR: 17.1) and 15–18 (OR: 5.96) Pre-ASD < TD: canonical babbling ratio (9–12 $d = -1.09$, 15–18 $d = -.62$) Pre-ASD < TD: syllable volubility (9–12 $d = -2.07$, 15–18 $d = -2.77$)	
Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011)	6, 9, and 12 months	6 month HR: 28; LR: 20 9 month HR: 37; LR: 29 12 month HR: 38; LR: 31	High-risk siblings	5-min parent-child free play	HR < LR: speech-like vocalization rate (6-month $d = -.21$, 9-month d:$-.46$, 12-month $d = -1.19$) HR > LR: non-speech vocalization rate (6-month: $d = .12$, 9-month: $d = .59$, 12-month $d = .75$) HR < LR: consonant inventory size (6-month: $d = -.35$, 9-month: $d = -.83$, 12-month $d = -.28$) HR < LR: percent of syllables which are canonical (9 month $d = -.79$)	
Plumb, A. M., & Wetherby, A. M. (2013)	18–24 months	Pre-ASD: 50 DD: 50 TD: 50	Screening	CSBS-DP	Pre-ASD < TD: proportion speech-like ($d = -.94$), pre-ASD < DD marginally ($d = -.53$) Pre-ASD > TD proportion non-speech ($d = .93$) No differences in syllable structure of speech-like Pre-ASD < TD: proportion directed speech-like No difference directed non-speech	
Pokorny, F. B., Schuller, B., Marschik, P. B., Brueckner, R., Nyström, P., Cummins, N., … Falck-Ytter, T. (2017)	10 months	Pre-ASD: 10 TD: 10	High-risk siblings	12-min parent child interaction	No difference in rate of all non-vegetative vocalizations	
Rozga, A., Hutman, T., Young, G. S., Rogers, S. J., Ozonoff, S., Dapretto, M., & Sigman, M. (2011)	6 months	HR+:8 HR−: 41 LR: 35	High-risk siblings	1-min parent child free play	No difference in rate of social vocalizations	
Santos, J. F., Brosh, N., Falk, T. H., Zwaigenbaum, L., Bryson, S. E., Roberts, W., … Brian, J. A. (2013)	18 months	HR+: 23 LR: 20	High-risk siblings	ADOS Module 1	Classification accuracy of 79.1–97.7% using acoustic features and Support Vector	

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
					Machines or Probabilistic Neural Networks	
Schoen, E., Paul, R., & Chawarska, K. (2009) Study 1	18–36 months	ASD: 12 TD: 11	Already diagnosed	CSBS-DP	ASD > TD: speech-like vocalizations excluding words/approximations (d = .98), greater proportion of utterances greater than .5 seconds, proportion of complex pitch contours (29.84% versus 10.07%) No difference in mean duration, high or low pitch, pitch range	
Schoen, E., Paul, R., & Chawarska, K. (2009) Study 2	18–36 months	ASD: 30 (21 pre-meaningful speech) TD: 15 (4 pre-meaningful speech)	Already diagnosed	CSBS-DP	ASD PreSpeech < TD Meaningful Speech: consonant inventory, number of consonant blends, number of closed syllables, syllable structure complexity ASD > TD: frequency of squeals No differences: delight, distress, atypical, communicative non-speech vocalizations	
Schoen, E., Paul, R., & Chawarska, K. (2011)	ASD/TD: 18–36 Language-matched: 11–13 months	ASD: 30 TD (age-matched): 11 TD (language--matched): 23	Already diagnosed	ASD/TD age: CSBS-DP TD language: parent-child interaction	ASD > TD (age-match): non-speech (d = .95) ASD < TD (age-match): consonant inventory size (d = − 2.33), syllable structure level (d = 1.77) No differences in speech-like or total number of vocalizations ASD > TD(age),TD(language): atypical (ds = .94,1.08), squeals (ds = .99,.96)	
Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., & Lester, B. M. (2012)	6 months	pain cries: HR: 7 LR: 5 non-pain cries: HR: 10 LR: 6	High-risk siblings	Home recordings	ASD > TD: pain-related cry F0 (d = 1.65) and F0 range (d = 1.25) No differences non-pain cry	
Sheinkopf, S. J., Mundy, P., Oller, D. K., & Steffens, M. (2000)	ASD mean: 44.67 months DD mean: 36.09	Preverbal ASD: 15 DD: 11	Already diagnosed	Early Social Communication Scales	ASD > TD: Atypical (d = 1.07) No difference in canonical babble ratio, utterance/syllable count	
Shumway, S., & Wetherby, A. M. (2009)	18–24	Pre-ASD: 50 DD: 50 TD: 50	Screening	CSBS-DP	ASD < TD: percentage of communicative acts with vocalization (d = .97)	
Sullivan, K., Sharda, M., Greenson, J., Dawson, G., & Singh, N. C. (2013)	24 months	ASD: 39 DD: 20 TD: 26	Already diagnosed	Free Play of ADOS	ASD < TD: number of all vocalizations	

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
					No group differences in syllabic rhythm, formant transitions, or place of articulation Low verbal IQ ASD > DD,TD: area of place of articulation	
Swanson, M. R., Shen, M. D., Wolff, J. J., Boyd, B., Clements, M., Rehg, J., … IBIS Network. (2017)	9 months	HR: 40 LR: 19	High-risk siblings	LENA home recordings	HR > LR: speech-like vocalizations No difference in number of conversational turns HR < LR: social babbling (6–9 month AOSI, d = .82)	
Talbott, M. R., Nelson, C. A., & Tager-Flusberg, H. (2016)	9 months	HR+: 6 HR−: 24 LR: 30	High-risk siblings	Toy and book reading sections of home video diaries	No differences in rate of syllable vocalization	No differences in mothers' rate of responding or type of response
Unwin, L. M., Bruz, I., Maybery, M. T., Reynolds, V., Ciccone, N., Dissanayake, C., … Whitehouse, A. J. O. (2017)	12 months	HR: 22 LR: 27	High-risk siblings	Home recordings of cry	HR < LR: cry duration (expiration phase, η^2 = 0.08); No difference in F0	No difference in perceived level of distress by researchers
Venuti, P., Caria, A., Esposito, G., De Pisapia, N., Bornstein, M. H., & de Falco, S. (2012)	13 months	Pre-ASD: 8 TD: 8	Home videos	Home videos	Pre-ASD < TD: duration of pauses in cry (d = − 1.60) Pre-ASD > TD: F0 of FSP of LTAS (d = 1.75)	fMRI: Adults show increased activation in verbal/prosodic processing and emotion regions when hearing Pre-ASD > TD cry
Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014)	ASD: 18–48 months TD: 8–48 months	ASD: 77 TD: 106	Already diagnosed	LENA home recordings	ASD < TD: total vocalizations (β = − 0.274), proportion speech-related vocalizations (β = − 0.274), total child-caregiver interaction, ratio of leading to following, and contingency of adult response to speech-like versus non-speech	
Warren, S., Gilkerson, J., A Richards, J., Oller, D. K., Xu, D., Yapanel, U., & Gray, S. (2009)	16–48 months	ASD: 36 TD: 78	Already diagnosed	LENA home recordings	ASD < TD: speech-like vocalizations (29% fewer, d = − .79), conversation duration (d = − .83) ASD > TD: vocalization duration mean (d = .60) and variability, number of monologues (d = .53)	
Weismer, S. E., Lord, C., & Esler, A. (2010)	Autism mean: 30.9 (3.4) PDD-NOS mean: 30.7 (3.8) DD mean: 30.4 (3.9)	Autism: 179 PDD-NOS: 78 DD: 69	Screening and high-risk siblings	ADOS-PL/Module 1	Frequency of Vocalization score from the ADOS significantly predicted concurrent receptive and expressive language in the Autism group (R^2 = .311 and .425), but not the PDD-NOS or DD group	
Werner, E., & Dawson, G. (2005)	12 and 24 months	Pre-ASD(with regression): 15	Home videos	Home videos	12 months: Pre-ASD(no regression) < TD < Pre-ASD(regression):	

Table 1 (continued)

Authors	Ages	Sample size	Design	Sample setting	Main findings	Perception/parent findings
		Pre-ASD(no regression): 21 TD: 20			frequency of complex babble or words (no regression versus TD $d = -.65$, regression versus TD $d = .34$) 24 months: both Pre-ASD < TD: frequency of complex babble or words (no regression: $d = -.86$, regression $d = -.90$) No differences in simple babble	
Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007)	18–24 months	Pre-ASD: 50 DD: 23 TD: 50	Community screening	CSBS-DP	Pre-ASD < TD: consonant inventory ($d = -1.57$), word inventory ($d = -1.36$) No difference Pre-ASD versus DD (ds $= -0.39, -0.24$) Within Pre-ASD, consonant inventory predicts ADOS communication score at 3 years ($r = -.34$)	
Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004)	Pre-ASD mean: 21 months (3.5) DD mean: 18.4 months(3.7) TD mean: 20.3 months (3.2)	Pre-ASD: 18 DD: 18 TD: 18	Community screening	CSBS-DP	Pre-ASD > DD,TD: unusual prosody (50% of ASD children versus 0% TD,DD) Pre-ASD > TD: lack of communicative vocalization with consonants (83% ASD, 72% DD, 17% TD)	
Winder, B. M., Wozniak, R. H., Parladé, M. V., & Iverson, J. M. (2013)	13 and 18 months	HR: 15 LR: 15	High-risk siblings	45-min semi-structured and unstructured parent child interaction in the home	HR < LR: rate of communicative nonword vocalizations ($\eta_p^2 = 0.14$), words ($\eta_p^2 = 0.17$), non-communicative nonword vocalizations at trend level	

of vocalization they produced [28]. This can result in less instructive interaction for the infant, leading to downstream delays in social development and language learning.

Reduction in the social directedness of vocalizations emerges in the second half of the first year of life in ASD and is consistently reported thereafter [22, 30, 38, 40, 46, 47]. Directedness in these studies is typically defined as the conjunction of vocalization and gaze or other communicative behaviors (e.g., gesture, responding) aimed toward a social interaction partner. When infants fail to direct their vocalizations toward another person, their vocalizations are presumed to lack communicative intent. Although the role that social directedness plays in the feedback loop of language learning has not been directly studied, it is reasonable to hypothesize that directedness affects the responses received from caregivers (and thus, has an influence on infant learning opportunities).

Consistent with the observation that delays in speech and language are often among the first concerns noted by parents of children later diagnosed with ASD [16], recent studies have found early delays in canonical babbling. Canonical babbles are defined as vocalizations that include mature transitions between a consonant and a vowel (e.g., “ba”, “dada”), and are therefore fundamental building blocks for language. In ASD, delayed attainment of the canonical babbling milestone, reductions in the rate of production of canonical babbling, and smaller consonant inventories have all been reported, beginning in the second half of the first year of life [36, 39, 40, 48–52]. However, the effect sizes of these differences are much smaller when compared to developmentally or language delayed samples [22, 41, 42, 52] vs. typically developing control groups.

Canonical babbling differences have been prospectively associated with ASD symptoms [52], and babble characteristics in the first year of life have been shown to predict an individual’s subsequent diagnosis [35, 36]. Interestingly, several reports have suggested that infants who later develop ASD produce fewer *socially directed babbles*, in particular [22, 30, 40, 41]. This suggests that the interaction between the linguistic maturity of vocalizations and their social directedness may be a particularly potent ASD marker, with greater specificity for predicting ASD than babbling rates alone. Babbling has multiple functions in development, and reductions in babbling in infants with ASD—whether socially directed or not—may decrease their opportunities to practice speech, motor, language, and social interaction skills.

The relationship of these various early vocalization features to later language is not entirely clear. However, a recent meta-analysis found a strong correlation between vocalization (broadly defined, and including most measures noted above) and standardized measures of expressive language in children younger than age 9 (overall $r = 0.5$) [53••]. Although correlations with expressive language were highest for consonant-centric measures (i.e., the most linguistically mature vocalizations, such as canonical babbling, $r = 0.62$) and for concurrent measures ($r = 0.77$, compared to longitudinal measurement, $r = 0.33$), all associations were significant. This suggests that early vocalizations play a role in language development and may carry valuable prospective information about long-term outcomes.

Acoustics of Early Vocalizations

Cries of infants later diagnosed with ASD have been particularly well investigated acoustically, as recently reviewed in detail [54••]. Briefly, these cries have been reported to have unusual acoustic properties (including higher fundamental frequency [55–64] and shorter pause duration [57, 59, 63, 65]) compared to non-ASD peers. Interestingly, naïve raters are sensitive to these differences, rating the cries of infants later diagnosed with ASD as more distressed sounding—and distressing to listen to—than the cries of infants who do not develop ASD [56, 57, 62, 66••, 67, 68]. Cries of infants later diagnosed with ASD produce different neural responses in listeners [63] and behavioral responses from parents [65]. In one study, mothers of children with and without ASD listened to recordings of elicited cries from when the child was 1-month old [66••]. The cries of children who were later diagnosed with ASD were rated as more distressing, more atypical, and more likely reflective of pain than typically developing (TD) infants. Although only eight total cry episodes were rated, the ratings were nearly non-overlapping, suggesting that when cries are elicited in a highly standardized manner, they may hold exceptionally early clues about later diagnosis. Although not yet instituted, a promising future direction is to take advantage of cries that are already elicited during early infancy in standardized (primary care) settings—specifically, cries produced during scheduled vaccinations. Research focused on automated acoustic analysis of these cries would be a valuable step toward identifying a pre-linguistic marker of ASD with a clear path to translation.

The literature detailing acoustic properties of other kinds of pre-linguistic vocalizations in ASD is smaller, with few studies evaluating the same sets of acoustic features. For instance, toddlers with ASD have been reported to use a higher proportion of long duration speech-like utterances [15, 45] than non-ASD toddlers. Different patterns of pitch contour have been reported in this age range as well, albeit in contrasting directions [45, 69]. Further research in this area would benefit from reporting group means on standardized acoustic feature sets to facilitate comparison and should clearly report important clinical characteristics (e.g., developmental delay) to clarify the specificity of acoustic markers.

Using Early Vocalizations to Predict Outcomes

A full picture of the myriad acoustic differences in vocalizations produced by infants later diagnosed with ASD has not yet emerged, but acoustic features have shown good classification accuracy in cross-validated studies of diagnosed toddlers, with emerging evidence of longitudinal predictive accuracy [14, 15, 32••, 40]. Notably, one study reported that high prediction accuracy for diagnostic outcome based on recordings made before diagnosis is currently possible (at 10 months) [32••]. This study extracted a standardized set of acoustic features, the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) [70], using open-source software. eGeMAPS consists of 88 acoustic parameters, selected to provide a minimalist set of features for paralinguistic and clinical speech analysis based on theoretical significance, potential to index affective changes in voice production, and empirical performance. This set includes features related to frequency, energy/amplitude, and spectral (balance/shape/dynamics) parameters. It is notable that eGeMAPS was designed to be used as a standardized set that is easy to extract, reducing the need for individual researchers to select acoustic features and harmonizing methods across studies, which was a need identified in a review of ASD voice signature research [5••]. eGeMAPS features were extracted from vocalizations that were manually segmented from a 12-min interaction between 10-month-olds at high familial risk for ASD who were diagnosed with ASD at age 3 (n = 10) and their parents, and from low-risk TD (n = 10) infant-parent dyads. Linear kernel Support Vector Machines (SVM, a supervised machine learning approach that performs classification on two groups based on a model that best separates the groups) and a 1-layer bidirectional long short-term memory neural network (a nonlinear machine learning approach) both achieved 75% accuracy in subject-level classification with an optimized decision threshold [32••]. Information about the most predictive features from these models was not reported.

Descriptive reports of the kinds of acoustic differences contributing to prediction, and when they emerge in development, will be important for advancing our clinical understanding of ASD. Additional prediction studies, potentially combining acoustic features with other qualities of vocalizations (e.g., directedness, atypicality ratings, babbling), may help shape these prediction models into useful clinical tools for early detection of ASD.

Preschool and School-Age

Atypical Paralinguistics

Beginning when children can be readily diagnosed with ASD (around age 4), a large literature supports the existence of group-level differences in paralinguistic production (e.g., prosody, voice quality) [4, 5••]. That paralinguistic differences exist is nearly indisputable; remaining questions include what exactly those differences are, how they can be detected, what they mean for people with autism, and how they can be used to improve clinical care.

Recent studies of school-aged children with ASD have focused on identifying acoustic signatures of atypical prosody, with particular attention to measures of pitch and pitch variability. In one study, 8- and 9-year-olds with and without Asperger's syndrome (AS) completed the turn-end subtest of the PEPS-C [71], in which they were prompted with pictures to produce a one-word utterance with either sentence or question intonation [71, 72]. Utterances produced by AS participants were rated by naïve adult raters as more atypical and were acoustically characterized as longer in duration, with greater pitch range, and higher mean and maximum pitch than TD participants [72]. A separate sample of Russian children aged 5–14 years (25 ASD, 60 TD) revealed higher pitch and pitch variability in spontaneous speech and word repetition in the ASD group [73]. In recordings of the interview section of the ADOS-2, school-age children with ASD (N = 65) produced higher and more variable vocal pitch than TD children (N = 17), but not a non-ASD clinical group (N = 18) [74].

In contrast, a few recent studies have not found evidence of increased pitch variability in ASD—and some have found the opposite. One small study of Swedish children aged 9–12 (11 ASD, 11 TD) found that the ASD group used significantly more words per utterance when telling a story from pictures but did not differ from the TD group on fundamental frequency mean, range, or variation, nor in perceptual ratings [75]. In another study, 6–9-year-old ASD children showed a smaller coefficient of variation in pitch than TD children during picture-naming, suggesting that the ASD group was acoustically more monotonous, and this difference negatively correlated with a measure of social reciprocity [76]. Of note, this difference was not observed in the pre-school age children in the sample, though the pre-school sample was very small (N = 6). In sum, it is unclear whether there are consistent differences in pitch variability in ASD. Although methodological differences between studies could have caused this pattern of discrepant findings, it is more likely that true clinical heterogeneity is the culprit; individuals with ASD have been described as both "monotonous" (i.e., low pitch variability) and "singsongy" (i.e., high pitch variability) at various times in the literature, e.g., [77].

Social Impact of Paralinguistic Differences in ASD

Atypical acoustics impact social interaction success for individuals with ASD [9•, 78]. For example, when stories told by 34 school-age children (half with ASD and unusual prosody as reported by their school-based speech language pathologist, half TD) were played for lay listeners (i.e., college students) [9•], listeners rated the recordings of the ASD children as more disordered and evaluated them as less likable than TD recordings. Listeners also rated five aspects of speech (articulation, clearness, fluency, accenting, and monotony). Articulation, clearness, and fluency were collapsed into a measure of "intelligibility," which accounted for 63% of the variance in ratings of "disordered" versus typical. Intelligibility accounted for 42% of the variance in likability ratings, and monotony accounted for 21%. Accenting was not a significant predictor of either disordered ratings or likability. Taken together, these findings indicate that lay listeners are sensitive to speech differences in ASD in a way that impacts likability ratings, and which appears to be driven more by acoustic factors related to motor control (e.g., voice quality and intelligibility) than by linguistic prosody per se.

Atypical prosody in autistic speakers has been shown to contribute to poor social interactions—including misunderstandings by individuals with whom they are conversing. For example, one naturalistic study of conversation analyzed "trouble-source turns" (i.e., conversational turns that cause another speaker to make some kind of correction or repair in the conversation) in recordings of group therapy sessions from seven boys with ASD aged 11–13 years [79•]. Of these problematic turns, atypical prosody was observed through creaky voice in 35.3%, quiet voice in 31.4%, pitch excursions (dramatic pitch changes) in 23.5%, stretched syllables in 17.6%, and jerky speech rhythm in 15.7%. High rates of atypical prosodic features in trouble-source turns suggest that prosody contributes directly to sub-optimal conversational interactions.

Research Gap: Preschool Children

A paucity of studies into paralinguistic features produced by pre-school age children represents a noticeable gap in the recent literature. Because many speech features change dramatically across development [80], it is imperative to conduct developmentally sensitive studies that include every age range and report age-specific effects. The preschool age is especially important in ASD, as this is the time frame during which the majority of children first receive a diagnosis [19]. As mentioned above, one study of picture naming detected no acoustic differences in 4- to 6-year-olds with ASD despite finding differences in pitch variation in school-aged children [76]. The failure to detect differences in preschoolers may have been due to insufficient power, especially considering that an older study detected a wider pitch range in this age group in ASD [13]. Additionally, at least two recent studies have indicated the presence of paralinguistic differences in preschoolers with ASD. In the first, 21% of children with ASD and no intellectual disability (N = 83, 4–6 years of age) showed clinically relevant phonological speech problems on a phonological speech-production assessment [81]. In the second, utterances were extracted from a standardized social interaction and parent-child interactions [82] and the Prosody Voice-Screening Profile (PVSP [83]) was applied (N = 7 children with ASD, N = 7 TD children, 24–68 months of age). No differences were detected in rate, loudness, or pitch, but significant differences were observed in stress (e.g., atypical stress of multi-syllable words [82]). Of note, many nonsignificant group differences reported in this age range stand in contrast to consistently detected differences in school-age children. Taken together, this work suggests that there could be developmentally-specific patterns of acoustic differences in ASD, but more research is needed in this age range to understand the exact nature of those differences.

Diagnostic Classification Using Paralinguistic Features

The goal of most prediction studies in school-aged children is to predict diagnosis and/or symptom severity from concurrently measured vocalization features. This maps onto three potential clinical uses: clinical screening, diagnostic decision support (helping clinicians determine whether a presenting child meets criteria for ASD at that moment in time, including differential diagnosis), and phenotyping/individualized treatment planning (determining which intervention—or how much intervention—might be most helpful to improve ASD symptoms).

Several recent studies have demonstrated good accuracy in predicting diagnostic labels from sets of acoustic speech features (see Table 2 for studies not already reviewed in [5••]). Prediction of utterance-level and subject-level diagnostic category from a set of acoustic speech features was reasonable (67.6% and 66.7%) in a sample of six Japanese 10- to 13-year-olds [84]. In 3- to 10-year-old children (30 with ASD, 54 TD), 74.9% accuracy was achieved using 24 dimensions related to pitch when dividing utterances into thirds, which was slightly higher than prediction using the whole utterance [85], indicating the need for attention to the kinds of utterances used for prediction and the time-scale of analysis.

As in infant research, several recent studies of school-aged children have used standardized acoustic sets, including eGeMAPS and the ComParE feature set (a set of 6373 features which won the Interspeech 2013 Computational Paralinguistics Challenge for classifying ASD diagnosis) [86]. Notably, much of the development in this area comes from engineering, computational, and linguistics conferences, highlighting the importance of collaborative relationships between clinicians and technical researchers. For example, using a standard acoustic set on a sample of 803 utterances from 14 autistic children (4 to 10 years old), support vector machine learning correctly classified severity according to the Social Responsiveness Scale, 2nd edition (SRS-2) (mild, moderate, severe) with 73.7% accuracy [87, 88]. This is particularly

Table 2 Recent studies of paralinguistic diagnostic classification in children. This table is designed to align with Table 5 in Fusaroli et al. [5••] and includes only studies that were not included in that table. In combination, these tables provide an updated overview of prediction studies. All classifiers predicted subject-level diagnosis, unless otherwise specified. *UAR* unweighted average recall, which represents the mean value of recognition accuracy for each class. The F-measure combines precision and recall. Accuracy is the percentage of correct predictions. Precision is the proportion of positive predictions that are correct. Recall is the proportion of true positives that were predicted correctly

Authors	Sample size and matching criteria	Age	Level of function of the ASD group	Task	Features	Feature selection (FS), validation (V), classifier (C), performance (P)
Baird, A., Amiriparian, S., Cummins, N., Alcorn, A.M., Batliner, A., Pugachevskiy, S., et al. (2017)	14 ASD No TD	4–10 year	Not reported	Spontaneous production (interaction with therapeutic robot)	Interspeech 2009 Emotion Challenge (IS09) Interspeech 2010 Paralinguistic Challenge (IS10) Interspeech 2013 Computational Paralinguistics Challenge (ComParE) 4096 spectrogram features (deep spectrum)	FS: No feature selection for IS09, IS10, or ComParE; deep spectrum features selected through convolutional neural networks and competitive swarm optimization(CSO) V: Test/train C: Classifying mild, moderate, or severe based on SRS-2. Support Vector Machines for IS09-emotion, IS10-paraling, and COMPARE; hybrid end-to-evolution (e2ev) for deep spectrum P: Unweighted Average Recall (UAR) (chance = 33%) IS09: 64.8% IS10: 73.7% ComParE: 62.2% e2ev: 61.9%
Cho, S., Liberman, M., Ryant, N., Cola, M., Schultz, R.T., Parish-Morris, J. (2019)	35 ASD, 35 TD Matched on Full Scale, Verbal, and Nonverbal IQ estimates, and sex ratio	ASD mean 11.42 (2.51) year TD mean 10.57 (2.82) year	Full Scale IQ mean: 105.51	Social interaction	ComParE features from participant and confederate (mean, median, standard deviation, and interquartile range - 352 features) and linguistic features (272 features)	FS: Features significantly correlated with diagnostic status, and Principle Component Analysis V: Leave-one-out cross-validation C: Gradient Boosting Model P: For ASD, Accuracy = .66, Precision = 0.82, Recall = 0.66, F1-score = 0.73
Deng, J., Cummins, N., Schmitt, M., Qian, K., Ringeval, F., Schuller B. (2017)	[a]11 Autistic disorder (AD), 10 pervasive developmental disorder not otherwise specified (PDD-NOS), 13 Specific Language Impairment, 68 TD. Patients matched on age, sex, academic grades, lexical abilities. TD children	6–18 year	Performance IQ > 70	Spontaneous production (narrative production)	eGeMAPS, ComParE	FS: Generalized Adversarial Networks (GAN), no selection used as comparison V: Test/train C: Unselected features + Support Vector Machines (linear, radial basis function, multi-layer perceptron) P: UAR (Chance = 25%) linear SVM eGeMAPS: 39.91%, ComParE: 42.83%

Table 2 (continued)

Authors	Sample size and matching criteria	Age	Level of function of the ASD group	Task	Features	Feature selection (FS), validation (V), classifier (C), performance (P)
	matched with patients on age and sex					C: GAN + linear Support Vector Machine P: UAR (Chance = 25%) eGeMAPS: 44.06%, ComParE: 46.93%
Nakai, Y., Takiguchi, T., Matsui, G., Yamaoka, N., Takada, S. (2017)	30 ASD, 51 TD (Japanese)	3–10 year	IQ: 68.2 (16.9), > 30 expressive words	Spontaneous production (lexical elicitation)	24 pitch-related statistics (static pitch and change over time)	FS: None V: 10-fold cross validation C: utterance-level prediction using Support Vector Machines P: Accuracy = 0.66, F-measure = .59 C: subject-level prediction using Support Vector Machines P: Accuracy = 0.76, F-measure = .73
Pokorny, F. B., Schuller, B., Marschik, P. B., Brueckner, R., Nyström, P., Cummins, N., … Falck-Ytter, T. (2017)	10 pre-ASD, 10 TD	10 m	Not reported	Spontaneous production (parent-child interaction)	Extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS)	FS: None V: 3-fold cross validation C: utterance-level prediction using linear Support Vector Machines and 1-layer bidirectional neural network; subject-level classification of 3-year-old diagnosis based on utterance classification P: subject-level Accuracy = 75% for both classifiers
Ringeval, F., Marchi, E., Grossard, C., Xavier, J., Chetouani, M., Cohen, D., et al. (2016)	[a]12 Autistic disorder (AD), 10 pervasive developmental disorder not otherwise specified (PDD-NOS), 13 Specific Language Impairment, 70 TD Patients matched on age, sex, academic grades, lexical abilities. TD children matched with patients on age and sex	6–18 year	Performance IQ > 70	Spontaneous production (narrative production)	eGeMAPS, IS09, IS10, Interspeech 2011 Challenge (IS11), ComParE	FS: None V: Leave One Speaker Out C: Support Vector Machines (linear, polynomial, and Gaussian kernels, with best kernel reported) P: Typical versus patient (UAR, chance = 50%) eGeMAPS: 83.2% IS09: 84.0% IS10: 87.6% IS11: 89.4% ComParE: 86.3% Diagnosis (UAR, chance = 25%) eGeMAPS: 48.2% IS09: 51.2% IS10: 53.7% IS11: 56.4% ComParE: 56.2%

Table 2 (continued)

Authors	Sample size and matching criteria	Age	Level of function of the ASD group	Task	Features	Feature selection (FS), validation (V), classifier (C), performance (P)
Schmitt, M., Marchi, E., Ringeval, F., Schuller, B. (2016)	English: 8 ASD, 10 TD [a]French: 13 ASD, 16 TD Hebrew: 7 ASD, 10 TD Swedish: 9 ASD, 11 TD	5–18 year	Not reported	Spontaneous production	eGeMAPS, ComParE	FS: None V: Test/train C: Diagnosis (within and between languages) Support Vector Machine with linear kernel P: Within language (UAR chance = 50%): English: eGeMAPS 81.8%, ComParE 86.9% French: eGeMAPS 82.9%, ComParE 87.1% Hebrew: eGeMAPS 62.2%, ComParE 71.6% Swedish: eGeMAPS 73.2%, ComParE 78.6% Classifier trained on other language(s) (UAR ranges, chance = 50%) English: eGeMAPS 43.7–60.6%, ComParE 38.9–56.9% French: eGeMAPS 40.2–57.5%, ComParE 44.4–55.4% Hebrew: eGeMAPS 50.6–64.9%, ComParE 54.7–67.0% Swedish: eGeMAPS 46.0–63.5%, ComParE 42.5–56.6%
Tanaka, H., Sakti, S., Neubig, G., Toda, T., Nakamura, S. (2014)	4 ASD, 2 TD No matching	10–13 year	IQ above 70	Spontaneous production (narrative production)	Standard deviation and coefficient of variation of pitch and intensity, rate (words per voiced second), voice quality (amplitude of the third formant, difference between the first and second harmonic, difference between first harmonic and third formant)	FS: None V: 10-fold cross-validationC: Native Bayes (utterance-level labels) P: Accuracy = 57.6% C: Support Vector Machine (utterance-level labels) P: Accuracy = 67.6%

[a] The sample was taken from the Child Pathological and Emotional Speech database (CPESD)

notable given that the task of classifying severity levels based on narrow bands within ASD is much more challenging than separating ASD from TD. In a different sample of 35 French school-age children with autism, pervasive developmental disorder-not otherwise specified, or specific language impairment, and 70 TD children, several standardized acoustic sets produced high accuracy in discriminating picture book narrations from TD group vs. diagnosed groups (83.2–89.4% across different feature sets) [89]. Prediction accuracy was lower when predicting specific diagnosis (56.4% in the best case), though the inclusion of both autism and pervasive developmental disorder-not otherwise specified makes this task more challenging than necessary, as these labels both fall under the category of ASD in DSM-5 [90]. By applying generative adversarial networks (a machine learning approach in which two neural networks compete to produce new data similar to the provided data) to the same dataset, researchers achieved an unweighted average recall for the four groups of 44.06% using eGeMAPs features, and 46.93% for COMPARE features (both significantly above chance, 25%) [91]. In one attempt to test the cross-linguistic prediction capability of such machine learning, SVM classifiers were trained and tested on a sample of children speaking either English, French, Swedish, or Hebrew (aged 6 to 11 years, ASD $N = 37$, TDC $N = 47$) [91]. Models trained and tested within a language using eGeMAPs or ComParE features did reasonably well (UARs ranging from 62.2 to 87.1% in the test set, where chance is 50%) [92•]. However, models trained on one or more languages and tested on another language did not fare well, with many models performing at or worse than chance, with even the best models achieving less than 70% accuracy [92•]. This work suggests that while paralinguistic features that are predictive of ASD diagnosis exist in every language, the specific features that predict membership may need to be selected using non-ASD speakers of the same language as a baseline reference.

Two studies directly addressed the question of whether automated, machine-learning classification is better at determining diagnosis than human judgment. One study of 35 children with ASD and 35 TD children used ComParE features and linguistic features from both the child and a confederate to achieve 66% accuracy (precision 82%, recall 66%) in classifying ASD (Gradient Boosting Model) [93]. This was similar to the accuracy achieved by undergraduate students listening to the conversation—67% (precision (73%, recall 89%). In a second study, using the same recordings of picture-naming described in [76], a set of 24 features related to fundamental frequency were used to classify diagnosis using SVM [94•]. Ten speech therapists rated these sets of words as being from a child with ASD or TD. With only these data available, the machine-learning classifier outperformed therapists using the F-measure (an index of both precision and sensitivity). While this comparison is unfair to speech therapists in that they would traditionally have significantly more information available to them (e.g., full utterances), it nonetheless serves as proof of principle that automated detection of acoustic features holds promise as a method for objectively augmenting the work of expert therapists. Further research in this area should examine how automated detection can support the diagnostic accuracy of a clinician working with a more reasonably limited set of data, such as the data available to a clinician in an under-resourced clinic.

Conclusions

A brief review of the recent literature on pre- and paralinguistic vocalizations in ASD indicates two primary ways that these features are meaningful in ASD: as developmentally important clinical targets and as potential behavioral biomarkers. Clinically, atypical acoustic productions impact children's experiences across the first decade of life, beginning with parental responses to infant cries and speech-like vocalizations, and later in the form of negative ratings by observers and difficulties during conversation. Taken together, the impact of atypical vocalizing likely contributes to a sub-optimal developmental social feedback loop, including altered parental responsiveness early in life and negative social evaluation as children get older, all of which compound to deprive children of valuable social and language learning experiences (Fig. 1). Thus, paralinguistic speech differences are important intervention targets, as ameliorating these differences (while acquiring words and grammar) could have cascading effects on long-term social and linguistic development.

The second value of these speech features is as an objective bio-behavioral marker, which could be used to indicate ASD risk in infants and as an adjunctive clinical decision support tool for young children. In infants, promising existing research is sufficient to justify further explorations in this area. Multivariate studies suggest that early vocalization acoustics can predict diagnostic outcome, as measured later in life. The full realization of this potential has deep implications for clinical care, as earlier flagging for treatment will improve intervention effectiveness and outcomes.

In older children, translational research that bridges the basic and clinical domains is critically needed. A number of reports using different samples, settings, languages, and feature sets collectively indicate that acoustic features can correctly classify diagnostic labels with reasonable accuracy. Echoing conclusions drawn by Fusaroli and colleagues [5••], we argue that it is currently difficult to draw conclusions from the multivariate acoustic feature prediction literature—because we need to know significantly more about which acoustic features predict best for whom and in what setting. However, it is also important to note that a full theoretical

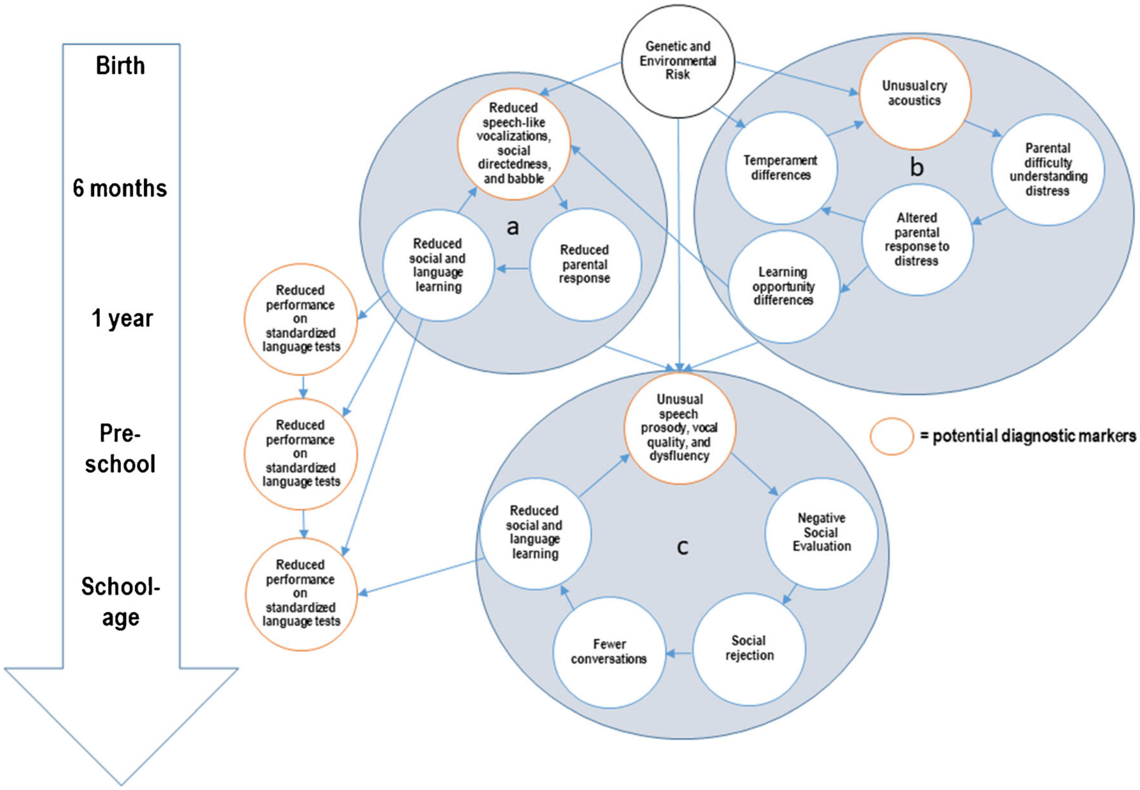

Fig. 1 The developmental feedback loops of pre- and paralinguistic features. Orange circles represent features that may be useful in diagnostic decision support. Loop "a" is an infant/toddler social feedback loop for speech development [28]. This model posits that reduced babbling, speech-like vocalizing, and directed vocalizations elicit reduced parental response, which reduces language-learning opportunities. Loop "b" represents a cry feedback loop, in which atypical cry serves as a "self-generated environmental factor" which results in parental difficulty understanding cry, eliciting different parental response [54]. This could alter learning opportunities and temperamental development. Both of these loops may lead developmentally to Loop "c," a negative social feedback loop. Atypical prosody and voice features result in negative social evaluation and sub-optimal social interaction, leading to fewer social interactions and social/language learning opportunities

understanding of why a tool works is not necessary for the tool to be clinically useful.

To achieve the goal of diagnostic decision support based on paralinguistics, several intermediary steps are required. First, there is a need for methodological advances that will automatically segment child vocalizations from complex audio streams in a variety of contexts. Most of the successful classification work presented in this review relies on audio segments that have been painstakingly segmented by human coders. Although some automatic speech detectors for children exist (e.g., the Language Environment Analysis System (LENA) algorithm [95]), they are imperfect [96], and technological improvements in this domain could improve both the accuracy and scalability of diagnostic classifiers. It is essential for engineers and clinicians to form partnerships to accomplish this goal and other methodological advancements. Second, research on the kinds of samples that can realistically be gathered in a clinical setting is needed. The discriminability of acoustic features of speech is sensitive to collection strategies (e.g., spontaneous versus elicited), and so translation of classification algorithms will depend on training the algorithms on appropriate data. Of note, although recordings of children completing the ADOS-2 are likely convenient for researchers who already have such samples collected, this may not be the most valuable for clinicians who wish to rely on automated speech detection specifically to eliminate the need to administer the ADOS-2. For this purpose, research on cries administered by vaccine administration or unstructured speech samples that can be easily collected by primary care providers are needed. Third, studies should investigate how vocalization analysis might be useful for clinicians. Although studies of vocalizations and their predictive power focus exclusively on speech features to rigorously demonstrate that they have predictive value, a real-world application would incorporate other knowledge, such as responses to screener questionnaires or brief history questionnaires. Research investigating the incremental value of paralinguistic classifiers (i.e., asking how much they improve classification *beyond* a standard screener) is necessary to understand what role they might realistically play in clinical practice. Fourth, studies of preschool aged children are particularly needed. Most children in the US are diagnosed during the preschool years [19], but many children diagnosed in community settings never receive an ADOS-2 or other time-consuming gold-standard measures as part of their diagnostic evaluation. Clinicians in settings with fewer resources might benefit from decision support generated by an automated paralinguistic risk detection tool. In summary, pre- and paralinguistic features hold promise as clinically useful markers of ASD in infants and young children, and with further research effort, may form the foundation for automated tools that seamlessly integrate with clinical care.

Funding Information This work supported by the Autism Science Foundation ASF #19-006 (grantee: Yankowitz), and NIDCD R03DC017944, "Infant Vocalizations as Early Markers of Autism Spectrum Disorder" (PI: Parish-Morris).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington: American Psychiatric Association; 2013.
2. Kanner L. Autistic disturbances of affective contact. Neuro Child. 1943;2:217–50.
3. Asperger H. Die "Autistischen Psychopathen" im Kindesalter. Arch Für Psychiatr Nervenkrankh. 1944;117:76–136.
4. McCann J, Peppé S. Prosody in autism spectrum disorders: a critical review. Int J Lang Commun Disord. 2003;38:325–50.
5.•• Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. Autism Res. 2017;10:384–407. **This systematic review and meta-analysis demonstrates significant differences in pitch and pitch range, and calls for systematic study of multivariate prediction based on high accuracy in studies conducted so far.**
6. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. J Autism Dev Disord. 2000;30:205–23.
7. Lord C, Rutter M, Couteur AL. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994;24:659–85.
8. Cutler A, Dahan D, van Donselaar W. Prosody in the Comprehension of Spoken Language: A Literature Review. Lang Speech. 1997;40:141–201.
9.• Redford MA, Kapatsinski V, Cornell-Fabiano J. Lay Listener Classification and Evaluation of Typical and Atypical Children's Speech. Lang Speech. 2018;61:277–302. **This study demonstrates that lay listeners are sensitive to speech features in ASD, which contribute to perceptions of disorder and likeability.**
10.• Patel SP, Kim JH, Larson CR, Losh M. Mechanisms of voice control related to prosody in autism spectrum disorder and first-degree relatives. Autism Res. 2019;12(8):1192–210. **This study suggests that atypical audio-vocal integration may be a genetically-based mechanism of prosodic differences in ASD.**
11. Marchi E, Schuller BW, Baron-Cohen S, Golan O, Bölte S, Arora P, et al. Typicality and emotion in the voice of children with autism spectrum condition: evidence across three languages. INTERSPEECH. 2015.
12. Asgari M, Bayestehtashk A, Shafran I. Robust and Accurate Features for Detecting and Diagnosing Autism Spectrum Disorders. Lyon: Proc Interspeech; 2013.
13. Bonneh YS, Levanon Y, Dean-Pardo O, Lossos L, Adini Y. Abnormal speech spectrum and increased pitch variability in young autistic children. Front Hum Neurosci. 2011;4:237.
14. Santos JF, Brosh N, Falk TH, Zwaigenbaum L, Bryson SE, Roberts W, et al. Very early detection of Autism Spectrum Disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. 2013 IEEE Int Conf Acoust Speech Signal Process. 2013. p. 7567–71.
15. Oller DK, Niyogi P, Gray S, Richards JA, Gilkerson J, Xu D, et al. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. Proc Natl Acad Sci U S A. 2010;107:13354–9.
16. Herlihy L, Knoch K, Vibert B, Fein D. Parents' first concerns about toddlers with autism spectrum disorder: Effect of sibling status. Autism Int J Res Pract. 2015;19:20–8.
17. Granpeesheh D, Dixon DR, Tarbox J, Kaplan AM, Wilke AE. The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders. Res Autism Spectr Disord. 2009;3:1014–22.
18. Zwaigenbaum L, Thurm A, Stone W, Baranek G, Bryson S, Iverson J, et al. Studying the Emergence of Autism Spectrum Disorders in High-risk Infants: Methodological and Practical Issues. J Autism Dev Disord. 2007;37:466–80.
19. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. Morb Mortal Wkly Rep Surveill Summ Wash DC 2002. 2014;63:1–21.
20. Pierce K, Gazestani VH, Bacon E, Barnes CC, Cha D, Nalabolu S, et al. Evaluation of the Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. JAMA Pediatr. 2019;173:578–87.
21. Swanson MR, Shen MD, Wolff JJ, Boyd B, Clements M, Rehg J, et al. Naturalistic Language Recordings Reveal "Hypervocal" Infants at High Familial Risk for Autism. Child Dev. 2018;89:e60–73.
22. Plumb AM, Wetherby AM. Vocalization Development in Toddlers With Autism Spectrum Disorder. J Speech Lang Hear Res. 2013;56:721–34.
23.• Roche L, Zhang D, Bartl-Pokorny KD, Pokorny FB, Schuller BW, Esposito G, et al. Early Vocal Development in Autism Spectrum Disorder, Rett Syndrome, and Fragile X Syndrome: Insights from Studies Using Retrospective Video Analysis. Ther Adv Neurol Disord. 2018;2:49–61. **This article reviews studies of infant vocalizations gathered through retrospective home videos.**

24. Garrido D, Petrova D, Watson LR, Garcia-Retamero R, Carballo G. Language and motor skills in siblings of children with autism spectrum disorder: A meta-analytic review. Autism Res. 2017;10:1737–50.
25. Ozonoff S, Young GS, Belding A, Hill M, Hill A, Hutman T, et al. The broader autism phenotype in infancy: when does it emerge? J Am Acad Child Adolesc Psychiatry. 2014;53(53:398):398–407.e2.
26. Leonard HC, Bedford R, Pickles A, Hill EL. Predicting the rate of language development from early motor skills in at-risk infants who develop autism spectrum disorder. Res Autism Spectr Disord. 2015;13–14 Complete:15–24.
27. Kwok EYL, Brown HM, Smyth RE, Oram CJ. Meta-analysis of receptive and expressive language skills in autism spectrum disorder. Res Autism Spectr Disord. 2015;9:202–22.
28. Warlaumont AS, Richards JA, Gilkerson J, Oller DK. A Social Feedback Loop for Speech Development and Its Reduction in Autism. Psychol Sci. 2014;25:1314–24.
29. Sullivan K, Sharda M, Greenson J, Dawson G, Singh NC. A novel method for assessing the development of speech motor function in toddlers with autism spectrum disorders. Front Integr Neurosci. 2013;7.
30. Lee K-S, Shin YJ, Yoo H-J, Lee GJ, Ryu J, Son O, et al. Vocalization of Emotional and Social Expressions in Korean-Speaking Toddlers with Autism Spectrum Disorder and Those with Developmental Delay. Yonsei Med J. 2018;59:425–30.
31. Northrup JB, Iverson JM. Vocal Coordination During Early Parent-Infant Interactions Predicts Language Outcome in Infant Siblings of Children with Autism Spectrum Disorder. Infancy Off J Int Soc Infant Stud. 2015;20:523–47.
32.•• Pokorny FB, Schuller B, Marschik PB, Brueckner R, Nyström P, Cummins N, et al. Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach. ISCA. 2017:309–13. **This small study demonstrates that machine learning applied to a standardized acoustic feature set from vocalizations collected at 10 months can accurately predict diagnostic outcome at age 3.**
33. Schoen E, Paul R, Chawarska K. Phonology and vocal behavior in toddlers with autism spectrum disorders. Autism Res. 2011;4:177–88.
34. Chenausky K, Nelson C, Tager-Flusberg H. Vocalization Rate and Consonant Production in Toddlers at High and Low Risk for Autism. J Speech Lang Hear Res. 2017;60:865–76.
35. Patten E, Belardi K, Baranek GT, Watson LR, Labban JD, Oller DK. Vocal Patterns in Infants with Autism Spectrum Disorder: Canonical Babbling Status and Vocalization Frequency. J Autism Dev Disord. 2014;44:2413–28.
36. Paul R, Fuerst Y, Ramsay G, Chawarska K, Klin A. Out of the mouths of babes: vocal production in infant siblings of children with ASD. J Child Psychol Psychiatry. 2011;52:588–98.
37. Warren S, Gilkerson JA, Richards J, Oller DK, Xu D, Yapanel U, et al. What Automated Vocal Analysis Reveals About the Vocal Production and Language Learning Environment of Young Children with Autism. J Autism Dev Disord. 2009;40:555–69.
38. Winder BM, Wozniak RH, Parladé MV, Iverson JM. Spontaneous Initiation of Communication in Infants at Low and Heightened Risk for Autism Spectrum Disorders. Dev Psychol. 2013;49:1931–42.
39. Chericoni N, de Brito WD, Costanzo V, Diniz-Gonçalves A, Leitgel Gille M, Parlato E, et al. Pre-linguistic Vocal Trajectories at 6–18 Months of Age As Early Markers of Autism. Front Psychol. 2016;7.
40. Garrido D, Watson LR, Carballo G, Garcia-Retamero R, Crais ER. Infants at-risk for autism spectrum disorder: Patterns of vocalizations at 14 months. Autism Res. 2017;10:1372–83.
41. Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early Indicators of Autism Spectrum Disorders in the Second Year of Life. J Autism Dev Disord. 2004;34:473–93.
42. Sheinkopf SJ, Mundy P, Oller DK, Steffens M. Vocal Atypicalities of Preverbal Autistic Children. J Autism Dev Disord. 2000;30:345–54.
43. Gabrielsen TP, Farley M, Speer L, Villalobos M, Baker CN, Miller J. Identifying Autism in a Brief Observation. Pediatrics. 2015;135:e330–8.
44. Talbott MR, Nelson CA, Tager-Flusberg H. Maternal Vocal Feedback to 9-Month-Old Infant Siblings of Children with ASD. Autism Res Off J Int Soc Autism Res. 2016;9:460–70.
45. Schoen E, Paul R, Chawarska K. Vocal productions in toddlers with autism spectrum disorders. Speech Sound Disord Child San Diego Plur Publ Inc. 2009:181–204.
46. Ozonoff S, Iosif A-M, Baguio F, Cook IC, Hill MM, Hutman T, et al. A Prospective Study of the Emergence of Early Behavioral Signs of Autism. J Am Acad Child Adolesc Psychiatry. 2010;49:256–266.e2.
47. Shumway S, Wetherby AM. Communicative Acts of Children With Autism Spectrum Disorders in the Second Year of Life. J Speech Lang Hear Res. 2009;52:1139–56.
48. Iverson JM, Wozniak RH. Variation in Vocal-Motor Development in Infant Siblings of Children with Autism. J Autism Dev Disord. 2007;37:158–70.
49. Landa RJ, Gross AL, Stuart EA, Faherty A. Developmental Trajectories in Children With and Without Autism Spectrum Disorders: The First 3 Years. Child Dev. 2013;84:429–42.
50. Landa RJ, Holman KC, Garrett-Mayer E. Social and Communication Development in Toddlers With Early and Later Diagnosis of Autism Spectrum Disorders. Arch Gen Psychiatry. 2007;64:853–64.
51. Werner E, Dawson G. Validation of the Phenomenon of Autistic Regression Using Home Videotapes. Arch Gen Psychiatry. 2005;62:889–95.
52. Wetherby AM, Watt N, Morgan L, Shumway S. Social Communication Profiles of Children with Autism Spectrum Disorders Late in the Second Year of Life. J Autism Dev Disord. 2007;37:960–75.
53.•• McDaniel J, D'Ambrose Slaboch K, Yoder P. A meta-analysis of the association between vocalizations and expressive language in children with autism spectrum disorder. Res Dev Disabil. 2018;72:202–13. **This meta-analysis demonstrates that early vocalizations are predictive of expressive language ability in ASD, suggesting they may be both a marker and an intervention target.**
54.•• Esposito G, Hiroi N, Scattoni ML. Cry, Baby, Cry: Expression of Distress As a Biomarker and Modulator in Autism Spectrum Disorder. Int J Neuropsychopharmacol. 2017;20:498–503. **This article reviews a line of evidence from humans and mouse models suggesting atypical cry in ASD, which elicits altered response from caregivers.**
55. Bornstein M, Costlow K, Truzzi A, Esposito G. Categorizing the cries of infants with ASD versus typically developing infants: A study of adult accuracy and reaction time. Res Autism Spectr Disord. 2016;31:66–72.
56. Esposito G, Nakazawa J, Venuti P, Bornstein MH. Perceptions of distress in young children with autism compared to typically developing children: a cultural comparison between Japan and Italy. Res Dev Disabil. 2012;33:1059–67.
57. Esposito G, Nakazawa J, Venuti P, Bornstein MH. Componential Deconstruction of Infant Distress Vocalizations via Tree-Based Models: A Study of Cry in Autism Spectrum Disorder and Typical Development. Res Dev Disabil. 2013;34:2717–24.
58. Esposito G. Rostagno M del C, Venuti P, Haltigan JD, Messinger DS. Brief Report: Atypical Expression of Distress During the Separation Phase of the Strange Situation Procedure in Infant Siblings at High Risk for ASD. J Autism Dev Disord. 2014;44:975–80.

59. Esposito G, Valenzi S, Islam T, Bornstein MH. Three physiological responses in fathers and non-fathers' to vocalizations of typically developing infants and infants with Autism Spectrum Disorder. Res Dev Disabil. 2015;0:43–50.
60. Esposito G, Venuti P. Developmental changes in the fundamental frequency (f0) of infants' cries: a study of children with Autism Spectrum Disorder. Early Child Dev Care. 2010;180:1093–102.
61. Esposito G, Venuti P. Understanding early communication signals in autism: a study of the perception of infants' cry. J Intellect Disabil Res. 2010;54:216–23.
62. Ozturk Y, Bizzego A, Esposito G, Furlanello C, Venuti P. Physiological and self-report responses of parents of children with autism spectrum disorder to children crying. Res Dev Disabil. 2018;73:31–9.
63. Venuti P, Caria A, Esposito G, De Pisapia N, Bornstein MH, de Falco S. Differential brain responses to cries of infants with autistic disorder and typical development: An fMRI study. Res Dev Disabil. 2012;33:2255–64.
64. Sheinkopf SJ, Iverson JM, Rinaldi ML, Lester BM. Atypical Cry Acoustics in 6-Month-Old Infants at Risk for Autism Spectrum Disorder. Autism Res. 2012;5:331–9.
65. Esposito G, Venuti P. Comparative Analysis of Crying in Children with Autism, Developmental Delays, and Typical Development. Focus Autism Dev Disabil. 2009;24:240–7.
66.•• English MS, Tenenbaum EJ, Levine TP, Lester BM, Sheinkopf SJ. Perception of Cry Characteristics in 1-Month-Old Infants Later Diagnosed with Autism Spectrum Disorder. J Autism Dev Disord. 2019;49:834–44. **This small study suggests that cry may be an extremely early-emerging bio-behavioral marker of ASD. Cries of 1-month-old infants later diagnosed with ASD were rated as more distressing, atypical, and indicative of pain than cries of TD infants.**
67. Esposito G, Venuti P. How is crying perceived in children with Autistic Spectrum Disorder. Res Autism Spectr Disord. 2008;2:371–84.
68. Esposito G, Venuti P, Bornstein MH. Assessment of distress in young children: A comparison of autistic disorder, developmental delay, and typical development. Res Autism Spectr Disord. 2011;4:1510–6.
69. Brisson J, Martel K, Serres J, Sirois S, Adrien J-L. Acoustic analysis of oral productions of infants later diagnosed with autism and their mother. Infant Ment Health J. 2014;35:285–95.
70. Eyben F, Scherer KR, Schuller BW, Sundberg J, André E, Busso C, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. IEEE Trans Affect Comput. 2016;7:190–202.
71. Peppé S, McCann J. Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. Clin Linguist Phon. 2003;17:345–54.
72. Filipe MG, Frota S, Castro SL, Vicente SG. Atypical Prosody in Asperger Syndrome: Perceptual and Acoustic Measurements. J Autism Dev Disord. 2014;44:1972–81.
73. Lyakso E, Frolova O, Grigorev A. A Comparison of Acoustic Features of Speech of Typically Developing Children and Children with Autism Spectrum Disorders. In: Ronzhin A, Potapova R, Németh G, editors. Speech Comput. Springer International Publishing; 2016. p. 43–50.
74. Parish-Morris J, Liberman M, Ryant N, Cieri C, Bateman L, Ferguson E, et al. Exploring Autism Spectrum Disorders Using HLT. In: Proc Third Workshop Comput Linguist Clin Psychol. San Diego: Association for Computational Linguistics; 2016. p. 74–84.
75. Dahlgren S, Sandberg AD, Strömbergsson S, Wenhov L, Råstam M, Nettelbladt U. Prosodic traits in speech produced by children with autism spectrum disorders – Perceptual and acoustic measurements. Autism Dev Lang Impair. 2018;3:2396941518764527.
76. Nakai Y, Takashima R, Takiguchi T, Takada S. Speech intonation in children with autism spectrum disorder. Brain and Development. 2014;36:516–22.
77. DePape A-MR, Chen A, Hall GB, Trainor LJ. Use of Prosody and Information Structure in High Functioning Adults with Autism in Relation to Language Ability. Front Psychol. 2012;3.
78. Lyakso E, Frolova O, Grigorev A. Perception and Acoustic Features of Speech of Children with Autism Spectrum Disorders. In: Karpov A, Potapova R, Mporas I, editors. Speech Comput. Springer International Publishing; 2017. p. 602–12.
79.• Wiklund M, et al. J Pragmat. 2016;94:76–97. **This naturalistic conversational analysis implicates paralinguistic speech features (prosody and voice quality) in conversational difficulty in ASD.**
80. Stathopoulos E, Huber J, Sussman J. Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4-93 Years of Age. J Speech Lang Hear Res JSLHR. 2011;54:1011–21.
81. Kjellmer L, Fernell E, Gillberg C, Norrelgen F. Speech and language profiles in 4- to 6-year-old children with early diagnosis of autism spectrum disorder without intellectual disability. Neuropsychiatr Dis Treat. 2018;14:2415–27.
82. McAlpine A, Plexico LW, Plumb AM, Cleary J. Prosody in Young Verbal Children With Autism Spectrum Disorder. Contemp Issues Commun Sci Disord Rockv. 2014;41:120–32.
83. Shriberg L, Kwiatkowski J, Rasmussen CR, Lof GL, Miller J. The Prosody-Voice Screening Profile (PVSP): Psychometric Data and Reference Information For Children Phonology Project Technical Report No. 1. 1997.
84. Tanaka H, Sakti S, Neubig G, Toda T, Nakamura S. Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorders in Children's Narrative. In: Proc Workshop Comput Linguist Clin Psychol Linguist Signal Clin Real. Baltimore: Association for Computational Linguistics; 2014. p. 88–96.
85. Kakihara Y, Takiguchi T, Ariki Y, Nakai Y, Takada S. Investigation of Classification Using Pitch Features for Children with Autism Spectrum Disorders and Typically Developing Children. Am J Signal Process. 2015;5:1–5.
86. Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, et al. The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. 2013.
87. Baird A, Amiriparian S, Cummins N, Alcorn AM, Batliner A, Pugachevskiy S, et al. Automatic Classification of Autistic Child Vocalisations: A Novel Database and Results. Interspeech 2017. ISCA; 2017. p. 849–53.
88. Constantino J, Gruber C. The Social Responsiveness Scale Manual, Second Edition (SRS-2). Los Angeles: Western Psychological Services; 2012.
89. Ringeval F, Marchi E, Grossard C, Xavier J, Chetouani M, Cohen D, et al. Automatic Analysis of Typical and Atypical Encoding of Spontaneous Emotion in the Voice of Children. Proc INTERSPEECH 2016 17th Annu Conf Int Speech Commun Assoc ISCA. San Francisco, CA, United States; 2016. p. 1210–4.
90. American Psychiatric Association, American Psychiatric Association, editor. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Washington: American Psychiatric Association; 2013.
91. Deng J, Cummins N, Schmitt M, Qian K, Ringeval F, Schuller B. Speech-based Diagnosis of Autism Spectrum Condition by Generative Adversarial Network Representations. Londres: 7th Int Digit Health Conf; 2017. p. 53–7.
92.• Schmitt M, Marchi E, Ringeval F, Schuller B. Towards Cross-lingual Automatic Diagnosis of Autism Spectrum Condition in Children's Voices. Speech Commun 12 ITG Symp. 2016. p. 1–5. **This machine learning classification study finds that across several languages, acoustic features can accurately classify ASD**

diagnosis when the model is trained on only speakers of the same language. However, model performance dramatically decreases when trained on speakers of a different language.

93. Cho S, Liberman M, Ryant N, Cola M, Schultz RT, Parish-Morris J. Automatic detection of Autism Spectrum Disorder in children using acoustic and text features from brief natural conversations. Proc Interspeech. Graz, Austria; 2019.

94.• Nakai Y, Takiguchi T, Matsui G, Yamaoka N, Takada S. Detecting Abnormal Word Utterances in Children With Autism Spectrum Disorders: Machine-Learning-Based Voice Analysis Versus Speech Therapists. Percept Mot Skills. 2017;124:961–73. **This study represents a first step toward evaluating how automatic speech-based diagnostic classifiers perform compared to professionals. Machine learning performed slightly better than speech language pathologists, who were provided with an extremely limited dataset (single word utterances).**

95. Xu D, Yapanel U. Gray S. LENA Found: Reliability of the LENA Language Environment Analysis System in young children's natural home environment; 2009.

96. Jones RM, Plesa Skwerer D, Pawar R, Hamo A, Carberry C, Ajodan EL, et al. How effective is LENA in detecting speech vocalizations and language produced by children and adolescents with ASD in different contexts? Autism Res Off J Int Soc Autism Res. 2019;12:628–35.