Research Article

Infant–Mother Acoustic–Prosodic Alignment and Developmental Risk

Amanda Seidl,^a Alejandrina Cristia,^b Melanie Soderstrom,^c Eon-Suk Ko,^d
Emily A. Abel,^e Ashleigh Kellerman,^e and A. J. Schwichtenberg^e

Purpose: One promising early marker for autism and other communicative and language disorders is early infant speech production. Here we used daylong recordings of high- and low-risk infant–mother dyads to examine whether acoustic–prosodic alignment as well as two automated measures of infant vocalization are related to developmental risk status indexed via familial risk and developmental progress at 36 months of age.

Method: Automated analyses of the acoustics of daylong real-world interactions were used to examine whether pitch characteristics of one vocalization by the mother or the child predicted those of the vocalization response by the other speaker and whether other features of

infants' speech in daylong recordings were associated with developmental risk status or outcomes.

Results: Low-risk and high-risk dyads did not differ in the level of acoustic–prosodic alignment, which was overall not significant. Further analyses revealed that acoustic–prosodic alignment did not predict infants' later developmental progress, which was, however, associated with two automated measures of infant vocalizations (daily vocalizations and conversational turns).

Conclusions: Although further research is needed, these findings suggest that automated measures of vocalizations drawn from daylong recordings are a possible early identification tool for later developmental progress/concerns.

Supplemental Material: <https://osf.io/cdn3v/>

Early in life, the identification of children at elevated risk for developmental concerns can be difficult. However, early identification is imperative for promoting optimal child outcomes. This study explores the use of daylong acoustic recordings coupled with theory-driven measurements to quantify one type of dyadic alignment, acoustic–prosodic alignment, as well as other measures of early linguistic development to explore their use as early developmental risk markers.

To identify children at elevated risk for developmental concerns, this study builds on a model commonly used in research on children with autism spectrum disorder

(ASD). In this field, young children at risk for ASD and related delays and disorders (e.g., because they have a family history of ASD, tuberous sclerosis, chromosome 15q11.2-13.1 duplication syndrome) are tested in infancy rather than waiting for a confirmatory diagnosis. Targeting young children at risk is an effective strategy because, in families raising a child with ASD, approximately 20% of younger siblings will receive an ASD diagnosis (henceforth referred to as high-risk infant siblings; Ozonoff et al., 2011). In addition to ASD, high-risk infant siblings are also at elevated risk for other developmental concerns, including language and social communicative difficulties (Drumm & Brian, 2013; Miller et al., 2015), cognitive impairments (Brian et al., 2014), and subclinical features of ASD (Constantino, Zhang, Frazier, Abbacchi, & Law, 2010; Messinger et al., 2013; Ozonoff et al., 2014). Starting at 9 months of age, prospective studies have used standardized laboratory-based assessments to identify high-risk infants with developmental concerns (Georgiades et al., 2013; Macari et al., 2012; Northrup & Iverson, 2015). For example, Northrup and Iverson (2015) reported that high-risk infants with language delays at 36 months exhibited less simultaneous speech with their mothers at 9 months of age. Given the range of developmental concerns in early childhood in high-risk infant siblings, recent intervention practices are now targeting social

^aSpeech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN

^bLSCP, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, Paris, France

^cUniversity of Manitoba, Winnipeg, Canada

^dChosun University, Gwangju, South Korea

^eHuman Development and Family Studies, Purdue University, West Lafayette, IN

Correspondence to Amanda Seidl: aseidl@purdue.edu

Editor-in-Chief: Julie Liss

Editor: Stephanie Borrie

Received July 29, 2017

Revision received November 17, 2017

Accepted February 12, 2018

https://doi.org/10.1044/2018_JSLHR-S-17-0287

Disclosure: The authors have declared that no competing interests existed at the time of publication.

communication development (e.g., infant-led initiations; Green et al., 2015, 2017, 2013) in high-risk samples. Most recently, Green and colleagues' (2017) parent-mediated randomized control trial identified increases in infant-led interactions and attention to their mothers for high-risk infants enrolled in the treatment condition. Thus, prospective infant sibling studies provide opportunities to examine early markers of a range of later developmental concerns, including ASD, which may inform current intervention practices. In this study, we use a prospective infant sibling design to gather quantitative measurements of early social communication development including dyadic acoustic-prosodic alignment, which we will explore as a function of risk status (i.e., high, low) as well as a potential marker of developmental progress/concerns.

Our key hypotheses focus on acoustic-prosodic alignment, a form of dyadic synchrony. Prosody includes information present in the acoustic signal above the segmental level (the individual speech sounds), such as pitch, duration, and intensity. The combination of these features generates rhythm and melody of speech, which function to convey information in diverse linguistic and paralinguistic domains such as affective (e.g., happy/sad vocal quality), grammatical (e.g., question/statement), pragmatic (e.g., focus), lexical (e.g., stress on *record* as a noun vs. as a verb). In this paper, we focus on prosody for a few key reasons. First, prosodic differences (e.g., broader pitch range) are common in children with developmental concerns (McCann & Peppe, 2003; Paul, Augustyn, Klin, & Volkmar, 2005). Second, the particular type of prosody explored, acoustic-prosodic alignment, is socially driven. For example, in adult dyads, acoustic alignment is moderated by the quality of social interactions (e.g., positive dyadic interactions are associated with higher acoustic alignment; Lee et al., 2010). This social aspect of acoustic-prosodic alignment may be particularly relevant with high-risk infant siblings (e.g., Baron-Cohen et al., 1996). Specifically, a recent prospective study documents less prosodic alignment between mother-infant dyads when the infant is at high risk for ASD (Quigley, McNally, & Lawson, 2016). Acoustic alignment may therefore distinguish high-risk children from their peers and should be explored as an early risk factor for developmental concerns. Alignment, as we define it, involves the interplay and matching of behaviors between individuals that occur during the course of dyadic interactions. Alignment is often referred to as an intricate dance between dyadic partners, as they each take turns leading and responding to one another, as well as engaging in mutual states (Feldman, 2003). Within this study, acoustic alignment will be explored by examining mother and infant vocalizations; however, it is important to mention that dyadic synchrony/alignment is broadly defined in the literature and can include a variety of behavioral features, such as posture/gestures, gaze patterns, and facial expressions, in addition to physiological measures such as heart rate. Indeed, several studies have assessed early dyadic synchrony and responsiveness and found it to be a promising marker for developmental concerns (e.g., Siller & Sigman, 2008; Yirmiya et al., 2006). Acoustic alignment, in particular, occurs when

conversational partners come to resemble each other acoustically as a function of their interaction, which is obvious at the scale of individual conversational turns.

This may occur with respect to a variety of acoustic parameters, such as the rate of speech, pitch height and variability, phonemic targets, as well as lexical and syntactic choices (Brennan, 1996; Giles & Coupland, 1991; Levitan & Hirschberg, 2011).

Dyadic synchrony/alignment is also an important proximal process in early development, which may be predictive of general infant developmental outcomes. Previous dyadic synchrony studies report improved social, cognitive, and language outcomes in children who engage in more synchronous behaviors with their dyadic partner (e.g., Feldman, Greenbaum, & Yirmiya, 1999; Feldstein et al., 1993; Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001). Previous ASD-related studies also highlight the developmental importance of dyadic synchrony for informing the development of joint attention and language (for a review, see Mundy, Sullivan, & Mastergeorge, 2009). For example, Apicella and colleagues (2013) retrospectively assessed dyadic synchrony from infant home videos of children diagnosed with ASD. They reported that infants who later had developmental concerns exhibited fewer synchronous motor-vocal patterns at 12 months of age. Similarly, additional studies reported that difficulties with elements of dyadic synchrony between 12 and 18 months of age (but not earlier) preceded clinical diagnoses (Saint-Georges et al., 2011; Wan et al., 2013; Yirmiya et al., 2006). As a whole, these studies provide strong support for the use of dyadic synchrony or alignment as an early behavioral marker of developmental concerns. However, quantifying dyadic alignment as implemented in these studies requires extensive training and coding time, making direct clinical applications challenging. Furthermore, it is difficult to objectively and continuously quantify dyadic synchrony or alignment using these methods. Finally, dyadic synchrony or alignment can be difficult to assess because it is not robustly apparent in all interactions (i.e., it may come and go as interactions and interaction partners change throughout the day).

Our study builds on the concept of dyadic synchrony explored in previous work while addressing several of the limitations highlighted above. Using existing technology (a lightweight recording device able to record all day), coupled with proprietary algorithms to automatically identify who is speaking in these recordings,¹ our study assesses how mother-infant dyadic partners match each other. Specifically, our recordings and subsequent automated analyses require little training to administer and no coding time (as all analyses are automatic) and capture the full day, as opposed to brief interactions, which may or may not display optimal or representative levels of alignment due to time

¹Language Environment Analysis (LENA) and its proprietary algorithms have been instrumental in making research on large-scale analyses of child and caregiver speech feasible and worthwhile. However, we hope that open source collaborative efforts, joint between developmental scientists and speech technologists, will help to move the field forward more quickly.

of day, interaction partners present, or other unrelated variables. The focus on acoustic markers is key in providing an easily recorded and scored measure for clinical use.

Which feature should be selected to explore this acoustic alignment? We have chosen to use prosody and pitch range, specifically, as our key marker of alignment, because in recent work (Ko, Seidl, Cristia, Reimchen, & Soderstrom, 2016; Quigley et al., 2016) it was shown that maternal pitch range significantly predicted typically developing [TD] infants' pitch range within the same conversation, and in a recent meta-analysis (Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017) pitch range was shown to be a strong predictor of diagnosis (ASD, TD), whereas some other prosodic markers were not as good at discriminating groups (e.g., speech rate). However, it is important to note that Quigley and colleagues' (2016) findings were limited to exploring pitch range within a 20-min interaction for usable utterances, not specific to a continuous exchange of vocalizations between mother and infant throughout a given day. Furthermore, the authors noted difficulties in obtaining analyzable samples of infant vocalizations during the dyadic interaction due to the limited frequency and duration of infants producing vocalizations. These limitations highlight that analyzing daylong recordings, as done by Ko and colleagues (2016), may provide a robust index of samples for measuring acoustic alignment throughout the day as well as better capture instances of temporal responsiveness between mothers and their infants.

Daylong automatically analyzed recordings have already been shown to be useful in the identification of children with ASD. For example, measurements gathered from these recordings can discriminate between children (aged 10–48 months) with and without diagnoses with a high rate of accuracy (Oller et al., 2010; Yoder, Oller, Richards, Gray, & Gilkerson, 2013). These findings were generated from a two-phase study of children with ($n = 77$) and without diagnoses ($n = 106$). Child recordings were gathered using hardware from the LENA Foundation and analyzed using software from the LENA Foundation. This software automatically tags timed segments as containing speech by a female adult, a male adult, the target child, another child, and/or background noise. Within these tagged vocalizations, researchers can further assess several features including rhythm/syllabicity, spectral tilt, pitch control, bandwidth, and duration using customized Praat scripts. The automatic approach used above to discriminate groups was primarily data-driven and included several data reduction techniques (e.g., factor analysis) to isolate potentially predictive child speech elements. This powerful approach has not yet been used prospectively to predict developmental progress in young infants at low versus high familial risk, as a function of the linguistic behavior they exhibit in daylong recordings. This is what we do, focusing on pitch range specifically for the reasons stated above.

The current study evaluated acoustic–prosodic alignment between infants at high risk for developmental concerns and their mothers, compared to low-risk infants. Specifically, we assessed the following research questions to determine if acoustic–prosodic alignment informs risk or later development:

Research Question 1: Does acoustic–prosodic alignment differ by risk status? Based on previous research, we expect low-risk dyads to exhibit acoustic–prosodic alignment and high-risk dyads to exhibit lower levels of alignment.

Research Question 2: Do our acoustic–prosodic alignment measures or other automated speech indices predict later developmental progress/concerns?

Method

Participants

Participants were part of a larger infant prospective development study that followed two groups of infants. The first group included infant siblings of children with ASD (high-risk group), and the second group included siblings of children with no known diagnosis at the time of enrollment (low-risk group). Families were eligible to enroll when their infants were 6, 9, 12, 15, or 18 months of age, if they met the following inclusion criteria: (a) they were being raised in a monolingual English-speaking home, (b) they had a home visit scheduled between June 2014 and June 2015, and (3) we had a LENA recording system available at the time of their home visit. The date range for inclusion reflects the funding available for this project rather than any empirical issues. A relatively low percentage (46%) of families agreed to enroll in this supplemental acoustic–prosodic alignment study, which they were told examined infant and mother speech using the LENA device. This low percentage of enrollment likely reflects the relatively heavy burden of the larger study (i.e., home and laboratory visits were every 3–6 months with a week of recording for each, though for the purposes of this study we only used one daylong recording from each child). For this study, we gathered 51 recordings from 44 families, when children were 9–30 months of age. Most of these recordings occurred between 12 and 24 months; therefore, we decided to exclude the relatively few recordings we gathered at 9 ($n = 4$) and 30 months ($n = 3$). We also excluded recordings that were shorter than 6 hr ($n = 4$) and recordings made at an older age when we already had one recording from the family at a younger age (e.g., a recording at 18 months, when we had one already at 12 months; $n = 4$).

Our final sample consisted of 36 recordings gathered between 12 and 24 months. These recordings were grouped into high- and low-risk cohorts with 18 infants/toddlers in the high-risk group (18 recordings), 17 in the low-risk group (17 recordings), and one child whose risk status was considered as NA because the child had an older sibling who had neither a TD nor ASD diagnosis and thus did not fit cleanly into either risk group. Table 1 conveys the socioeconomic status, sex, and ethnicity of the enrolled children.

Procedure

This study was conducted at Purdue University under the approval of the institutional review board, and all

Table 1. Sample characteristics stratified by familial risk status.

Demographic information	Risk group					
	Low (<i>n</i> = 17)		High (<i>n</i> = 18)		Other ^a (<i>n</i> = 1)	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Enrolled infant						
Male	10	59	12	67	1	100
Infant race and ethnicity						
White, non-Hispanic, or Latino	13	76	14	78	1	100
White, Hispanic, or Latino	0	0	2	11	0	0
Black, non-Hispanic, or Latino	0	0	1	~6	0	0
Black, Hispanic, or Latino	0	0	0	0	0	0
Multiracial, non-Hispanic or Latino	2	12	1	~6	0	0
Multiracial, Hispanic, or Latino	1	6	0	0	0	0
Unknown or unreported	1	6	0	0	0	0
Maternal education						
High school/GED	2	12	0	0	0	0
Associates	2	12	2	11	0	0
College degree	4	23	10	56	1	100
Graduate degree	7	41	2	11	0	0
Other	2	12	3	17	0	0
Unreported	0	0	1	5	0	0
Paternal education						
High school/GED	0	0	2	11	0	0
Associates	1	~6	1	5	0	0
College degree	4	24	5	28	1	100
Graduate degree	7	41	4	~22	0	0
Other	4	24	5	28	0	0
Unreported	1	~6	1	5	0	0
Family income						
Below \$20,000	1	6	0	0	0	0
\$20,001–\$40,000	0	0	4	22	1	100
\$40,001–\$60,000	5	29	4	22	0	0
\$60,001–\$80,000	6	35	5	28	0	0
\$80,001–\$100,000	1	6	2	11	0	0
\$100,001–\$125,00	4	24	0	0	0	0
\$125,001–\$150,000	0	0	1	~6	0	0
\$150,001 or higher	0	0	1	~6	0	0
Unreported	0	0	1	~6	0	0
Outcome						
TD	9	53	5	28	0	0
Non-TD	7	41	12	67	1	100

Note. GED = General Equivalency Diploma; TD = typically developing.

^aOther refers to the infant that did not meet risk status criteria for risk group analyses.

participating families completed the informed consent process. At the time of enrollment, families completed a demographic and family health history form. Laboratory visits included the Autism Diagnostic Observation Schedule–Second Edition (Lord et al., 2012) and Mullen Scales of Early Learning (MSEL; Mullen, 1995) measures as part of a larger battery of assessments. At each assessment age, families completed a home visit and a laboratory visit. If one of their child's assessment visits was scheduled between June 2014 and June 2015, they were invited to enroll in the supplemental study on acoustic–prosodic alignment. Families who opted to enroll in the acoustic–prosodic alignment study were given a small child's vest and a LENA recording device at their home visit. LENA is a speech recording system that includes a small digital recorder worn in the vest. The LENA system also includes specialized software used to analyze the data gathered. Importantly, the recording device and software

track both infant (key child or CHA) and caregiver (female adult near or FAN for our purposes here), which allowed us to examine their language separately and investigate the relationship between the two streams of speech.

With the home visiting examiner, mothers were asked to identify a day in the following week wherein she would be with the enrolled infant/toddler most of the day. Mothers were instructed to place the vest on their child in the morning of that day, turn on the LENA recording device, and have their child wear the vest/device for the entirety of the day (until bedtime). Mothers could decline participation at any time and were given the option to exclude unwanted portions or all of their data from analyses. Mothers returned the vest and LENA device at their laboratory visit later in the week or a second home visit was scheduled to retrieve the device. At the laboratory visit, mothers received \$50 in compensation for their participation in the home and laboratory

visits. Families did not receive additional compensation for completing the supplemental LENA assessment.

Measures

Spontaneous Speech Production Quantity and Vocal Maturity Measures

We extracted several measures of production from the daylong recordings provided by the LENA software. For our study, we used LENA to extract percentiles based on child age for the number of child vocalizations, child Automatic Vocalization Assessment (AVA) scores (an estimate of vocal maturity; Richards et al., 2017), and conversational turns from the LENA output. (See the Appendix for analyses concerning the reliability of LENA's speaker tagging for our sample.)

Acoustic–Prosodic Alignment Measures

Using Praat scripts (available to the interested reader at <https://osf.io/94e3j/>), we extracted pitch measures from each of the two vocalizations constituting a conversational turn for each FAN–CHA (female adult + near-child) dyad from the daylong recordings. Default settings were used for pitch sampling rate and overlap, and separate pitch range settings of 150–300 and 200–550 were used for mother and child, respectively. If a segment was identified as an infant vocalization, followed by his or her mother's vocalization, the conversational turn was labeled as “infant–mother.” Alternatively, if the mother vocalized and this was followed by an infant vocalization, the conversational turn was labeled as “mother–infant.” Imagine a conversational block with vocalizations as follows: infant, mother, mother, infant. The first pair will be classified as infant–mother, the second will be discarded because there is no talker alternation, and the third one is classified as mother–infant. For each turn, pitch range was extracted. Following the procedures in Ko et al. (2016), our planned analyses were based on range defined as maximum minus minimum; however, because this measure is sensitive to outliers, we replaced it here with the interquartile range (75% quartile minus 25% quartile). We report here only on the interquartile range because it is a more robust measure. Also, following Ko et al. (2016), we excluded from consideration segments that were mostly voiceless; specifically, paired turns where either of the turns was 90% or more voiceless were excluded. The number of data points recorded per infant reflects the number of vocalizations the infant produced within conversational turns during the daylong recordings. Thus, similar to Quigley and colleagues (2016), more voluble children and mothers accumulated more data points than less voluble ones. Note that our predictions here are that, if alignment is greater for low-risk than high-risk dyads, then the interquartile ranges should be more similar for conversational turns by dyads who are low risk than high risk.

Developmental Progress Measures

Each child was also tested using the MSEL at 18, 24, and 30 months to index each child's general developmental

progress as well as to examine how performance on the MSEL correlated with our acoustic–prosodic alignment measures. The MSEL is a standardized and normed developmental assessment given to children aged birth through 68 months (Mullen, 1995). It provides subscale scores of Receptive Language, Expressive Language, Visual Reception, Gross Motor, and Fine Motor Skills. Subscale scores on the MSEL are expressed as *T* scores, with a mean of 50 and a standard deviation of 10. The MSEL is commonly used in studies of high-risk infant siblings (Miller et al., 2015; Ozonoff et al., 2015) and is psychometrically sound for use in children with developmental concerns, including ASD (Akshoomoff, 2006). For this study, MSEL scores were used to inform clinical best estimates of developmental progress/concerns at the child's outcome visit (at 24, 30, or 36 months; $n = 1$ at 24 months, $n = 16$ at 30 months, $n = 17$ at 36 months; outcomes were unavailable for two children).² The clinical best estimate was determined independent of the LENA data, and highly trained clinicians classified each child into one of six clinically defined outcome categories based on clinical judgment: ASD, broader autism phenotype, attention deficit hyperactivity disorder concerns, global developmental delay, speech-language problems, or typical development. Given the small number of children in each of five clinical categories within this study, all were combined into one non-TD group. Therefore, child developmental progress/concerns included two groups: non-TD and TD.

Family Demographic and Health History Form

Infant sex, race, ethnicity, and family demographic characteristics were reported at the time of enrollment. As depicted in Table 1, values for high-risk and low-risk dyads were approximately matched. Family health history, including any ASD diagnoses, was also confirmed at enrollment. Initial diagnostic information was provided via parent report, and diagnostic confirmation was achieved by obtaining a copy of their community diagnosis, by an independent research assessment, or by scores above the ASD cutoff on the Social Communication Questionnaire.

Analyses

Given the high dimensionality of the data, we identified two key research questions, a primary question and an exploratory question. Our primary question was: Are acoustic–prosodic alignment or other language sample features associated with developmental risk (as indexed by risk-group status and an atypical developmental outcome)? Although a plan for data inclusion and analytic decisions for the primary question were preregistered to allow for a clear separation of confirmatory and exploratory analyses,

²Best clinical estimates for outcome status are unavailable for these children because of incomplete study participation. To receive outcome classification, children needed to have completed at least three developmental laboratory assessments, with their final visit occurring at either 24, 30, or 36 months of age.

the latter revealed that conclusions drawn from preregistered analyses did not generalize. Therefore, we concentrate here on what we believe is the most robust analysis.³ As will be seen later, this does not entail that the *p* values reported are *p*-hacked, because results of inferential analyses were not a criterion that biased our choice. See <https://osf.io/hscx8/> for the preregistration and <https://osf.io/cdn3v/wiki/home/> for the project history.

Results

Is Acoustic–Prosodic Alignment Associated With Developmental Risk?

To address this question, we entered the paired observations of pitch range for infant and mother described above into a mixed model (using the R package lme4) with the following shape: $\text{lmer}(\text{SecondIQR_z} \sim \text{FirstIQR_z} * \text{SecondSpeaker} * \text{Group} + \text{Age_cont} * \text{SecondSpeaker} + (\text{FirstIQR} | \text{Dyad_ID}) + (1 | \text{Block_ID}))$. Thus, the dependent variable is the pitch interquartile range of the second of two segments in a turn. The random structure includes a random slope and intercept for each dyad and random intercepts for the conversational block (nested in the child's individual ID because block names are not repeated across children). We have data from recordings gathered throughout a whole day, and paired segments are not randomly distributed. Instead, it typically happens that mother and child interact briefly, leading to a number of paired utterances, and then stop interacting for a while to start a new conversational block later. Declaring a random factor for block allows us to capture the fact that sometimes several data points are drawn from the same conversational bout and sometimes they are drawn from different bouts, because in LENA each conversation (defined as a bout of speech bounded by 5 s of nonspeech on either side) is given a unique identifier to account for the fact that sometimes multiple turns are drawn from the same block.

As for fixed effects, we fit the model to predict SecondIQR_z from the following fixed effects: FirstIQR_z is the pitch interquartile range of the first segment; SecondSpeaker (mother, child) is the speaker of the second segment; Group is low risk or high risk (risk status); and Age_cont is child age as a continuous variable (rounded to months) at the time of recording. The “_z” in FirstIQR_z and SecondIQR_z indicates that both pitch interquartile ranges (IQRs) have been *z*-scored within each speaker type, which was necessary to avoid a violation of homogeneity of variances due to infants and mothers having distinct but partially

overlapping distributions. The beta for range represents how much this variable explains of the dependent variable or, put otherwise, directly captures alignment effects. As for interactions, the interaction $\text{FirstIQR} * \text{SecondSpeaker}$ captures to what extent mothers align more or less than their children, $\text{FirstIQR} * \text{Group}$ captures to what extent risk groups differ in terms of their alignment level, etc. In other words, the crucial test of our prediction emerges from an interaction involving FirstIQR and Group , and possibly also including SecondSpeaker , as follows: If risk status modulates acoustic–prosodic alignment, then the relationship between FirstIQR (the pitch range of the first of two segments in a turn) and SecondIQR (the pitch range of the conversational partner's response) should be weaker for the high-risk group than the low-risk group. This may be particularly evident in the mother–child turns, where the child is the one expected to align with his or her mother. As noted above, this model specification was arrived at after exploration triggered by feedback from reviewers and thus departs from the one specified in the preregistration. Nonetheless, it is not the case that the conclusions below have been selected because they present a more exaggerated picture of the difference between low- and high-risk infants.

As a main test of our hypothesis, we compared a model with the two- and three-way interaction against another that did not have it, namely $\text{lmer}(\text{SecondIQR_z} \sim \text{FirstIQR_z} * \text{SecondSpeaker} + \text{SecondSpeaker} * \text{Group} + \text{Age_cont} * \text{SecondSpeaker} + (\text{FirstIQR_z} | \text{Dyad_ID}) + (1 | \text{Block_ID}), \text{data} = \text{pairedTurns})$. The more complex model was not statistically significantly better than the simpler model: $\chi^2(2) = 4.82, p = .088$.

Full results of the model fit to the whole data set and separating low- and high-risk status participants are shown in Table 2. In a nutshell, these models show that there is a main effect of Group , with lower ranges for low-risk dyads, and interactions between Group and SecondSpeaker (with mothers having significantly greater ranges than their infants in the low-risk group, but the opposite being true for the high-risk dyads), and Age and SecondSpeaker , with mothers' pitch ranges reducing as their infants age. There was also a main effect of SecondSpeaker , with the beta indicating that mothers overall had smaller pitch ranges than infants, but such effects are hard to visualize in the presence of interactions, as is the case here: We point out also that the main variable, which would indicate alignment, FirstIQR_z in this model was not significant. This means that we do not replicate the main effect of alignment found in Ko et al. (2016) in either group, a question to which we return in the discussion. Figure 1 shows the relationship between first and second IQR, with regression lines fit separately to mothers and infants. As clear from these data, the slopes are low incline and remarkably similar for low- and high-risk dyads.

Analyses Addressing Developmental Progress/Concerns

Of the 36 children who provided recordings for the study, we were able to assess developmental progress/outcomes for 34. Of these 34, three children received an ASD

³We carried out analyses to assess what pitch parameter led to more accurate pitch measurements by extracting 50 paired observations from each child and having a human coder tag FAN and CHN vocalizations in Praat. We found that measurements from this human coding correlated best with measurements extracted from the LENA segmentation when a type-driven range per speaker type was used, such that all FANs got the same pitch floor and ceiling values and all CHNs got another set of values. We are preparing a separate fuller report on this as a methodological contribution to the literature.

Table 2. Results of fixed effects of the linear mixed models.

Measure	Estimate	SE	df	t	Pr(> t)
Overall					
Observations: 11,133, groups: block_ID: 3833; Dyad_ID: 36					
(Intercept)	0.032	0.033	38.956	0.964	.341
firstlqr_z	0.000	0.023	52.831	-0.018	.986
secondSpeakerFAN	-0.053	0.027	9818.741	-1.978	.048
groupLR	-0.095	0.045	39.924	-2.111	.041
age_cont	0.005	0.006	31.665	0.875	.388
firstlqr_z:secondSpeakerFAN	-0.031	0.028	10739.576	-1.103	.270
firstlqr_z:groupLR	-0.005	0.031	56.061	-0.158	.875
secondSpeakerFAN:groupLR	0.111	0.038	9881.095	2.958	.003
secondSpeakerFAN:age_cont	-0.013	0.004	9766.989	-2.975	.003
firstlqr_z:age_cont	-0.001	0.003	13.843	-0.472	.644
firstlqr_z:scndSpkrFAN:groupLR	0.069	0.038	9982.167	1.815	.069
High risk					
Observations: 5,372, groups: block_ID: 1775; Dyad_ID: 19					
(Intercept)	0.037	0.038	15.609	0.994	.335
firstlqr_z	-0.004	0.023	25.274	-0.162	.872
secondSpeakerFAN	-0.047	0.027	4752.691	-1.785	.074
age_cont	0.018	0.008	13.789	2.111	.054
firstlqr_z:secondSpeakerFAN	-0.031	0.028	5270.567	-1.137	.256
secondSpeakerFAN:age_cont	-0.023	0.005	4685.715	-4.195	< .001
firstlqr_z:age_cont	-0.003	0.004	6.879	-0.654	.534
Low risk					
Observations: 5,761, groups: block_ID: 2058; Dyad_ID: 17					
(Intercept)	-0.076	0.028	20.103	-2.721	.013
firstlqr_z	-0.001	0.022	31.492	-0.059	.953
secondSpeakerFAN	0.068	0.027	5159.473	2.571	.010*
age_cont	-0.013	0.008	20.219	-1.735	.098
firstlqr_z:secondSpkrFAN	0.042	0.026	5673.082	1.624	.104
secondSpkrFAN:age_cont	0.004	0.007	5127.022	0.512	.609
firstlqr_z:age_cont	0.001	0.005	11.598	0.201	.844

Note. There are two categorical variables, both binomial: secondSpeaker (which codes whether it is the child or the mother who is speaking the range functioning as dependent variable) and group (which codes whether the child is high risk versus low risk). For both of these, treatment coding was used, with the treatment level coded between the parentheses.

Figure 1. SecondIQR as a function of FirstIQR for each group. Each point represents f0 ranges in a pair of vocalizations (e.g., the f0 range measured in a vocalization uttered by the mother in the x axis, the f0 range in the child's next vocalization as predicted by the main mixed model—see main text for further details on the models). The regression lines represent linear fits between these two variables.

diagnosis (all three from the high-risk group), another 17 were classified as non-TD (nine from the high-risk group and eight from the low-risk group), and the remaining 14 were classified as TD (five from the high-risk group and nine from the low-risk group). Here we explore whether our measure of acoustic–prosodic alignment is related to child risk status, developmental progress/concerns. Because only three children received an ASD diagnosis at outcome, we collapsed the data from these three children within our larger non-TD group ($n = 17$) for a total of 20 children with a non-TD outcome. We explored predictors of this larger categorization, rather than ASD specifically, because our limited sample size did not allow for an exploration of specificity (please see the linked materials on [osf.io](https://osf.io/kz95s/) for greater detail on outcomes for each test; <https://osf.io/kz95s/>, pp. 16–25). In all figures below, however, data from these three children with ASD outcomes are overlaid with an “x” as opposed to a dot to visually distinguish their data from the larger non-TD group. Other exploratory correlations are provided in the supplemental materials together with the data and R scripts used for analyses (see <https://osf.io/cdn3v/>).

Table 3 displays p values for tests of association between outcome status (non-TD $n = 20$, TD $n = 14$) and LENA-generated measures (AVA, conversational turns, infant vocalizations), as well as with a simple acoustic–prosodic alignment measure (the Spearman correlation coefficient in pitch interquartile range calculated over all paired conversational turns for each dyad, collapsing across infant–mother and mother–infant turns). We used Spearman correlations when assessing association between two continuous measures, Wilcoxon tests when dealing with one continuous and one categorical measure, and Phi tests for association between two categorical measures.

Although this simple alignment measure is not related to developmental progress/concerns, two of the LENA-generated measures were associated with developmental progress/concerns: daily vocalizations and conversational turns. These two variables (daily vocalizations and conversational turns) are highly correlated with each other as well. Children with TD outcomes had higher conversational turns and higher numbers of daily vocalizations in infancy (see Figures 2 and 3). Note that the children with LENA-provided measures who went on to receive ASD diagnoses do not clearly differ from the pattern of the remainder of the non-TD group.⁴

In the next set of models, we conducted regression analyses to explore which variables were robust predictors of developmental progress/concerns (see Table 4). Specifically, we fit a binomial model to predict outcome status (TD vs. non-TD). The basic model, attempting this prediction solely from risk status, revealed that being low risk was not associated with a significantly higher likelihood of

TD classification. The daily vocalizations model revealed that this variable was a significant predictor of outcome above and beyond risk status. A chi-square test on an analysis of deviance confirmed that the addition of daily vocalizations was a significant improvement over the basic model (reducing the residual deviance from 42.604 to 34.879, $p = .005$). A similar pattern emerged when comparing the basic model against one declaring conversational turns (where the residual deviance was reduced to 34.436, $p = .004$). In contrast, our alignment measure did not uniquely contribute to the basic model (above and beyond risk status (residual deviance 39.14, $p = .41$).

Discussion

This study was designed to assess acoustic–prosodic alignment as a correlate of developmental risk. First, we explored whether risk status was related to degree of acoustic–prosodic alignment with the conversational partner. Contrary to our hypothesis that high-risk infants would show less prosodic alignment with caregivers, these analyses revealed that high-risk and low-risk infant–mother dyads showed similar acoustic–prosodic alignment. In addition, we hypothesized that infant–mother alignment would be less affected than mother–infant alignment for the high-risk group, because we expected that high-risk children would have more difficulties matching their mothers’ pitch range. However, our analyses did not support this prediction either. However, it must be noted that the level of alignment was overall very low and not statistically significant in the analyses presented here.

Because we departed from Ko et al.’s (2016) methods in several ways and did not find an overall alignment effect here, readers may wonder whether their analyses would be more appropriate and show a greater difference between groups. Indeed, although we carried out other analyses (see supplemental materials, <https://osf.io/7vps9/>), we chose to focus here on an analysis that departs from theirs in numerous ways. Most of these choices do not affect whether alignment is detected or not, namely having a more complex random structure (random alignment slopes per dyad), a covariate not considered in their analysis (age), and—to a certain extent—which pitch parameter is used (pitch range and IQR lead to similar results, whereas average tends to lead to lower alignment estimations). One factor turned out to be crucial, namely whether the program used to extract the f_0 measures tracked f_0 within a specific range for mothers (150–350 Hz) versus children (200–550 Hz) or whether a single more extensive range was used with both (150–650 Hz). It is difficult, in our view, to establish which of these strategies objectively leads to more accurate alignment measurements. We favored the former approach here because a coding study revealed that correlations between f_0 measurements drawn from human coding and those from the LENA segments were better correlated when the range was set for each speaker type than when the same range was used with both. In other words, at least in what relates to measuring f_0 , a talker type range leads to more replicable results.

⁴Note that one child who received a diagnosis of ASD is not included in these analyses or in Figures 2 and 3 because he or she did not contribute a recording that was sufficiently long for the LENA to accurately compute daily vocalizations or turn count.

Table 3. *p* values table.

Measure	Risk group	Daily vocalizations	AVA percentile	Conversational turns	Synchrony measure
Outcome	.173	.016	.635	.013	.769
Risk status		.481	.458	.871	.389
Daily vocalizations			.363	.000	.107
AVA percentile				.555	.378
Conversational turns					.100

Note. Bold values indicate significant effects of interest. *p* values for tests of association between risk group and LENA-provided measures and a simple synchrony measure. The latter is the Spearman correlation coefficient calculated over all paired conversational segments within each dyad separately.

Nonetheless, what is important to bear in mind is that when we do follow the Ko et al. (2016) pipeline decisions, we replicate their findings (overall alignment, driven mainly by the mother aligning to the child more than the child to the mother) in the present data set but still do not find a difference between the groups. Thus, we are confident in our statement that alignment levels do not differ across the two groups. It is unclear why this is the case. It is possible that, although pitch has been argued to be a marker for ASD, production differences in pitch do not manifest in infancy. However, we did find both a main effect of group and an interaction between speaker and group (which were replicated in most alternative analyses). These analyses suggest that low-risk infants tend to have smaller pitch ranges than their mothers, whereas high-risk children, if anything, have slightly larger ranges than their mothers. Thus, we do find differences in pitch use across high- and low-risk samples in these daylong recordings, but these differences do not carry over to the phenomenon of alignment. Nonetheless, these findings may be fruitful to explore in future work and to extend to larger data sets.

Next, we carried out exploratory analyses concerned with predicting developmental progress/concerns. We assessed our measures' ability to determine overall developmental

progress (as indexed by a TD vs. non-TD outcome classification). Our results showed that, although our simple acoustic-prosodic alignment measure was not related to developmental progress/concerns, two LENA-generated measures, daily vocalizations and conversational turns, were. Both measures were predictors of developmental progress/concerns above and beyond risk status, although neither was a statistically stronger predictor than the other. This finding extends previous work (Oller et al., 2010) in which automated recordings were related to concurrent diagnoses (i.e., ASD) by showing that such measures are potentially useful prospective markers of developmental outcome status.

Why was our measure of alignment not related to developmental progress/concerns? The lack of relationship between acoustic-prosodic alignment and developmental progress/concerns is inconsistent with many dyadic alignment studies. For example, previous prospective infant sibling studies documented better developmental outcomes with higher rates of maternal responsiveness, maternal utterances, and matched dyadic behaviors. Most notably, our discrepant findings may reflect our assessment of general developmental progress/concerns (non-TD vs. TD). Our discrepant findings may also indicate that our data are an

Figure 2. Individual percentiles for conversational turns as a function of the child's outcome status. To facilitate inspection, observations have been jittered along the x-axis.

Figure 3. Individual percentiles for daily vocalization count as a function of the child's outcome status. To facilitate inspection, observations have been jittered along the x-axis.

Table 4. Estimates for logistic regression models.

Measure	Model 1 (basic)	Model 2 (daily vocalizations)	Model 3 (conversational turns)	Model 4 (synchrony measure)
Risk status (low risk)	$\beta = 0.81, SE = 0.73, p = .268$	$\beta = 1.31, SE = 0.88, p = .14$	$\beta = 1.10, SE = 0.86, p = .20$	$\beta = 0.955, SE = 0.78, p = .22$
Daily vocalizations		$\beta = 0.046, SE = 0.02, p = .018$		
Conversational turns			$\beta = 0.044, SE = 0.02, p = .014$	
Synchrony measure				$\beta = -3.60, SE = 4.43, p = .42$
Degrees of freedom	30	29	29	29
Residual deviance	42.6	34.8	34.4	39.14

Note. Estimates for logistic regression models predicting outcome from risk status alone or additionally one of the estimates extracted from daylong recordings (daily vocalizations, conversational turns, and a simple synchrony measure).

outlier in the larger field or that acoustic-prosodic alignment is distinct from other types of social/behavioral dyadic alignment in infant siblings (i.e., different mechanisms may be functioning in acoustic-prosodic alignment). However, an alternative interpretation is that our null result is merely due to diversity within the non-TD sample.

In conclusion, we did not find acoustic-prosodic alignment differences between low-risk and high-risk infants and their mothers using a method that is based on naturalistic daylong recordings but did find group differences in child speech from LENA, both in terms of acoustics (pitch range) and quantity (vocalization and turn counts) of speech. Our results, however, are limited by a small sample, which could mean that some of our null results could be due to lack of power. Furthermore, we cannot extrapolate as to how these group differences relate to long-term developmental outcomes because this study only followed children until 36 months of age. Despite these limitations, we argue that a noninvasive assessment of alignment in child and caregiver is a promising avenue of research. Furthermore, this method is easy to implement and quantify and is suitable for infants of a variety of ages, and exploratory results suggest that several measures gleaned from the LENA could be promising predictors for non-TD outcomes.

Acknowledgments

A. Cristia acknowledges the institutional support of ANR-10-IDEX-0001-02 PSL*, ANR-10-LABX-0087 IEC and the financial support of ANR-14-CE30-0003 MechELex. Also, A. Seidl would like to thank Purdue University for the Study in a Second Discipline Grant, which allowed for the completion of this work. We are grateful to all the families who opened their homes, shared their lives, and allowed us to watch their little ones grow—thank you. We also acknowledge and thank the senior mentors who supported the design and success of the larger study (for which this study was a supplement). Thank you, Sally Ozonoff and Thomas F. Anders.

Contributions: AS and AJS designed the primary research question, AJS oversaw the data collection, AC and AS performed the data analyses and interpretation, AS wrote the first draft, all authors provided input in critical revisions and approved the final version.

Compliance With Ethical Standards: This study was funded by the National Institutes of Health R00MH092431 (PI: Schwichtenberg).

Ethical Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

References

- Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology, 12*, 269–277.
- Apicella, F., Chericoni, N., Costanzo, V., Baldini, S., Billeci, L., Cohen, D., & Murarori, F. (2013). Reciprocity in interaction: A window on the first year of life in autism. *Autism Research and Treatment, 2013*, 705895. <https://doi.org/10.1155/2013/705895>
- Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., ... Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry, 168*, 158–163.
- Brennan, S. (1996). Lexical entrainment in spontaneous dialog. *Proceedings, ISSD-96, International Symposium on Spoken Dialogue, 1996*, 41–44.
- Brian, J., Roncadin, C., Duku, E., Bryson, S. E., Smith, I. M., Roberts, W., ... Zwaigenbaum, L. (2014). Emerging cognitive profiles in high-risk infants with and without autism spectrum disorder. *Research in Autism Spectrum Disorders, 8*, 1557–1566. <https://doi.org/10.1016/j.rasd.2014.07.021>
- Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M., & Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry, 167*, 1349–1356. <https://doi.org/10.1176/appi.ajp.2010.09101470>
- Drumm, E., & Brian, J. (2013). The developing language abilities and increased risks of “unaffected” siblings of children with autism spectrum disorder. *Neuropsychiatry, 3*, 513–524.
- Feldman, R. (2003). Infant-mother and infant-father synchrony: The coregulation of positive arousal. *Infant Mental Health Journal, 24*, 1–23.
- Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother-infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology, 35*(5), 223–231.

- Feldstein, S., Jaffe, J., Beebe, B., Crown, C. L., Jasnow, M. J., Fox, H., & Gordon, S. (1993). Coordinated interpersonal timing in adult–infant vocal interactions: A cross-site replication. *Infant Behavior and Development*, 16, 455–470.
- Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, 10(3), 384–407. <https://doi.org/10.1002/aur.1678>
- Georgiades, S., Szatmari, P., Zwaigenbaum, L., Bryson, S., Brian, J., Roberts, W., . . . Garon, N. (2013). A prospective study of autistic-like traits in unaffected siblings of probands with autism spectrum disorder. *Journal of the American Medical Association Psychiatry*, 70, 42–48. <https://doi.org/10.1001/2013.jamapsychiatry.1>
- Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Milton Keynes, England: Open University Press.
- Green, J., Charman, T., Pickles, A., Wan, M. W., Elsabbagh, M., Slonims, V., . . . Johnson, M. H. (2015). Parent-mediated intervention versus no intervention for infants at high risk of autism: A parallel, single-blind, randomised trial. *Psychiatry*, 2, 133–140. [https://doi.org/10.1016/S2215-0366\(14\)00091-1](https://doi.org/10.1016/S2215-0366(14)00091-1)
- Green, J., Pickles, A., Pasco, G., Bedford, R., Wan, M. W., Elsabbagh, M., . . . Johnson, M. H. (2017). Randomised trial of a parent-mediated intervention for infants at high risk for autism: Longitudinal outcomes to age 3 years. *The Journal of Child Psychology and Psychiatry*, 58(12), 1330–1340. <https://doi.org/10.1111/jcpp.12728>
- Green, J., Wan, M. W., Guiraud, J., Holsgrove, S., McNally, J., Slonims, V., . . . Johnson, M. H. (2013). Intervention for infants at risk of developing autism: A case series. *Journal of Autism and Developmental Disorders*, 43, 2502–2514. <https://doi.org/10.1007/s10803-013-1797-8>
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. D. (2001). Rhythms of dialogue in infancy. *Monographs of the Society for Research in Child Development*, 66, 1–131.
- Ko, E.-S., Seidl, M., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, 43(2), 284–309.
- Lee, C., Black, M., Katsamanis, A., Lammert, A., Baucom, B., Christensen, A., . . . Narayanan, S. S. (2010). Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. *Eleventh Annual Conference of the International Speech Communication Association*, 2010, 793–796.
- Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule–Second Edition (ADOS-2) Manual (Part I): Modules 1–4*. Torrance, CA: Western Psychological Services.
- Levitan, R., & Hirschberg, J. (2011). Measuring acoustic–prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech*. Columbia University Academic Commons. <https://doi.org/10.7916/D8V12D8F>
- Macari, S. L., Campbell, D., Gengoux, G. W., Saulnier, C. A., Klin, A. J., & Chawarska, K. (2012). Predicting developmental status from 12 to 24 months in infants at risk for autism spectrum disorder: A preliminary report. *Journal of Autism and Developmental Disorders*, 42, 2636–2647. <https://doi.org/10.1007/s10803-012-1521-0>
- McCann, J., & Peppe, S. (2003). Prosody in autism spectrum disorder: A critical review. *International Journal of Language & Communication Disorders*, 38(4), 325–350.
- Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., . . . Sigman, M. (2013). Beyond autism: A Baby Siblings Research Consortium study of high-risk children at three years of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52, 300–308. <https://doi.org/10.1016/j.jaac.2012.12.011>
- Miller, M., Young, G. S., Hutman, T., Johnson, S., Schwichtenberg, A. J., & Ozonoff, S. (2015). Early pragmatic language difficulties in siblings of children with autism: Implications for DSM-5 social communication disorder? *The Journal of Child Psychology and Psychiatry*, 56(7), 774–781.
- Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service, Inc.
- Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, 2, 2–21. <https://doi.org/10.1002/aur.61>
- Northrup, J., & Iverson, J. (2015). Vocal coordination during early parent–infant interactions predicts language outcome in infant siblings of children with autism spectrum disorder. *Infancy*, 20(5), 523–547.
- Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., . . . Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, 107, 13354–13359.
- Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., . . . Iosif, A. M. (2014). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child & Adolescent Psychiatry*, 53, 398–407. <https://doi.org/10.1016/j.jaac.2013.12.020>
- Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., . . . Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. *Pediatrics*, 128, e488–e495.
- Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., . . . Iosif, A. M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A Baby Siblings Research Consortium study. *The Journal of Child Psychology and Psychiatry*, 56(9), 988–998. <https://doi.org/10.1111/jcpp.12421>
- Paul, R., Augustyn, A., Klin, A., & Volkmar, F. (2005). Production and perception of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(2), 205–220.
- Quigley, J., McNally, S., & Lawson, S. (2016). Prosodic patterns in interaction of low-risk and at-risk-of-autism spectrum disorders infants and their mothers at 12 and 18 months. *Language, Learning, and Development*, 12(3), 1–16.
- Richards, J. A., Xu, D., Gilkerson, J., Yapanel, U., Gray, S., & Paul, T. (2017). Automated assessment of child vocalization development using LENA. *Journal of Speech, Language, and Hearing Research*, 60(7), 2047–2063.
- Saint-Georges, C., Mahdhaoui, A., Chetouani, M., Cassel, R. S., Laznik, M. C., Apicella, F., . . . Cohen, D. (2011). Do parents recognize autistic deviant behavior long before diagnosis? Taking into account interaction using computational methods. *PLoS One*, 6(7), e22393. <https://doi.org/10.1371/journal.pone.0022393>
- Siller, M., & Sigman, M. (2008). Modeling longitudinal change in the language abilities of children with autism: Parent behaviors and child characteristics as predictors of change. *Developmental Psychology*, 44(6), 1691–1704.
- VanDam, M., & Silbert, N. H. (2016). Fidelity of automatic speech processing for adult and child talker classifications. *PLoS One*, 11(8), e0160588. <https://doi.org/10.1371/journal.pone.0160588>

-
- Wan, W., Green, J., Elsabbagh, M., Johnson, M., Charman, T., & Plummer, F. (2013). Quality of interaction between at-risk infants and caregiver at 12–15 months is associated with 3-year autism outcome. *The Journal of Child Psychology and Psychiatry*, 54(7), 763–771.
- Yirmiya, N., Gamliel, I., Pilowsky, T., Feldman, R., Baron-Cohen, S., & Sigman, M. (2006). The development of siblings of children with ASD at 4 and 14 months: Social engagement, communication, and cognition. *The Journal of Child Psychology and Psychiatry*, 47(5), 511–523.
- Yoder, P. J., Oller, D. K., Richards, J. A., Gray, S., & Gilkerson, J. (2013). Stability and validity of an automated measure of vocal development from day-long samples in children with and without autism spectrum disorder. *Autism Research*, 6(2), 103–107.

Appendix

Reliability of Speaker Attribution in LENA

It is important to mention that, as with any automatic speech processing system, LENA's performance in identifying talkers (i.e., diarization) is not perfect. Recent studies, however, suggest that, compared with manual annotation, LENA tags agree with human tags for adult talkers 82% of the time and for child talkers 73%–76% of the time (VanDam & Silbert, 2016). This is comparable to the 84% agreement between manual codes and LENA codes found in Ko et al. (2016). Given that our key question is in exploring differences between high- and low-risk groups for convergence and that the accuracy of the tags is not expected to interact with our key variables, we were not overly concerned with LENA's less-than-perfect speaker tagging. However, given that the aforementioned estimates concerning accuracy come from recordings of typically developing and low-risk children, it was important to assess the accuracy of LENA's speaker tagging for the high-risk group separately. Thus, for reliability purposes, we manually coded 10 blocks in which a female adult (FAN) interacted with the key child (CHN) from each of 10 high-risk dyad recordings. Ten blocks were randomly selected by a program from each dyad, which resulted in a total of 1,384 segments that LENA had tagged as either CHN or FAN. Reliability was performed by volunteer research assistants with undergraduate-level linguistics training using a tailored program, which allowed them to tag segments as from an adult, a child, junk, or mixed. Each research assistant first listened to the whole conversational block. Next, she listened to each utterance that had been segmented by the LENA system and judged the segment as either adult, child, mixed, or junk, while being completely blind to the LENA tags for this. We then compared the LENA-generated codes with our manual tagging. Results revealed that LENA accurately tagged FAN and CHN as such 72% of the time, but that it reversed these tags (e.g., tagged a child as an adult) 15% of the time. Furthermore, LENA tagged junk (e.g., silence, breathing, nonspeech) as either FAN or CHN 11% of the time and mixed segments (where a single utterance was actually composed approximately half adult and half child vocalization) as either FAN or CHN approximately 2% of the time. In summary, error rates in LENA's diarization were similar to those found in other data sets (despite slightly more rigorous coding categories in our data set) and should not adversely impact the performance of our analyses.