Advances in Child Development and Behavior (Vol. 62), “New Approaches to Studying Child Development.”

Computational Approaches to Understanding Interaction and Development

Messinger, D.S.*, Perry, L.K.*, Mitsven, S.G., Tao, Y., Moffitt, J., Fasano, R.M., Custode, S.A., & Jerry, C.M.

November 2022

Note. The first two authors shared equally in the preparation of this chapter.

Acknowledgements. We thank Chaoming Song and Yi Zhang for their insights into detecting social contact and for Figure 10, and thank Chitra Banarjee for Figure 4. We thank the families, Disorders (R01DC018542), the Spencer Foundation, and the Autism Science Foundation. d here was supported by grants from the National Science Foundation (IBSS 1620294), the Institute for Education Sciences (R324A180203), the National Institute on Deafness and Communication

Abstract

Audio-visual recording and location tracking produce enormous quantities of digital data with which researchers can document children's everyday interactions in naturalistic settings and assessment contexts. Machine learning and other computational approaches can produce replicable, automated measurements of these big behavioral data. The economies of scale afforded by repeated automated measurements offer a potent approach to investigating linkages between real-time behavior and developmental change. In our work, automated measurement of audio from child-worn recorders—which quantify the frequency of child and adult speech and index its phonemic complexity—are paired with ultrawide radio tracking of children's location and interpersonal orientation. Applications of objective measurement indicate the influence of adult behavior in both expert ratings of attachment behavior and ratings of autism severity, suggesting the role of dyadic factors in these "child" assessments. In the preschool classroom, location/orientation measures provide data-driven measures of children's social contact, fertile ground for vocal interactions. Both the velocity of children's movement toward one another and their social contact with one another evidence homophily: children with autism spectrum disorder, other developmental disabilities, and typically developing children were more likely to interact with children in the same group even in inclusive preschool classrooms designed to promote interchange between all children. In the vocal domain, the frequency of peer speech and the phonemic complexity of teacher speech predict the frequency and phonemic complexity of children's own speech over multiple timescales. Moreover, children's own speech predicts their assessed language abilities across disability groups, suggesting how everyday interactions facilitate development.

Key words: machine learning, objective, automated measurement, interaction, development, audio, radio frequency identification, language, social, deep learning.

Computational Approaches to Understanding Interaction and Development

New sensing technologies such as first-person recordings that capture video from a child's first-person perspective are becoming increasingly common in developmental research (Gonzalez Villasanti et al., 2020; C. Yu & Smith, 2016; Yurkovic et al., 2021). There is tremendous interest in deploying both audiovisual recorders and newer sensors in naturalistic and semi-naturalistic environments ("in the wild") to understand children's day-to-day interactions. These sensors, however, can produce overwhelming quantities of recorded multimodal data. At the same time, new computational approaches, including machine learning and modeling tools from statistical physics are detecting meaningful patterns in these multimodal data. The efficient computational measurement and rendering of sensing data is sometimes termed behavior imaging (Rehg, 2011). Use of the term emphasizes that understanding real-time developmental processes has the same conceptual importance and requires the same level of computational rigor as neural imaging.

Behavior imaging has the potential to create new insights into children's behavior and development by opening previously intractable fields of inquiry. Examples abound. Visualizations of the temporal sequences of infants and their parents handling and gazing at objects are changing our understanding of joint attention and word learning (C. Yu & Smith, 2016; Yurkovic et al., 2021). Bootstrapped sequences of observed behaviors demonstrate the emergence of emotion regulation (gazing away from the parent during a smile) between three and six months (Yale et al., 2003). Likewise, Ruvolo et al. (2015) used an information theoretic perspective to demonstrate the emergence of infant-mother turn-taking over developmental time. Moreover, computations from statistical physics, as we detail below, allow for the detection of sociality from children's location and orientation without the use of machine learning (Messinger

et al., 2019). A specific behavior imaging approach, machine learning, has been key to understanding children's behavior and interactions as recorded with audio, video, and other sensors.

Machine Learning

Shallow and deep learning

Machine learning refers to algorithms that successively reprocess data into more abstract and compact forms (LeCun et al., 2015). Deep learning has led to a revolution in the flexibility and accuracy of machine learning. In conventional or shallow learning, algorithms are provided with selected features of sensor data (fundamental pitch contours from an audio signal, for example). In deep learning, multiple layered neural networks iteratively abstract information more directly from sensor data (e.g., the audio signal itself). A historical example is illustrative. In 2009, Messinger et al. predicted infant smiling using a one-dimensional support vector machines from the output of manifolds of face image data (shallow learning). By contrast, in 2021, Ertugrul et al. reliably detected infant smiling using a deep neural network directly from a video model of the infant's face.

Unsupervised machine learning

In addition to being shallow or deep, machine learning may be supervised or unsupervised. Unsupervised machine learning algorithms detect high level patterns in raw data, in a fashion metaphorically akin to exploratory factor analysis. Using separate videos of the faces of mothers and infants during their face-to-face interactions, for example, we used an unsupervised branch-and-bound procedure to detect correspondences between the facial expressions of each partner (Chu et al., 2017). Without prior training, this machine learning framework detected commonalities in infant and parent smiling from raw video.

Supervised machine learning

Measuring behavior. In the current chapter (and in developmental research more generally), the primary use of supervised machine learning is the measurement of behavior (Rehg, 2011; Rehg et al., 2014). Supervised algorithms are provided with labeled or annotated sensor data. They receive as input, for example, audio clips that are labeled as vocalizations (and non-vocalizations) or video segments labeled as smiles (and non-smiles). From this input, algorithms learn multilinear or nonlinear correspondences between sensor data and labeled behaviors. Finally, the machine learning algorithm—or sets of interlocking algorithms—are applied to new, unseen data to measure behavior. Throughout the chapter, we refer to the use of supervised machine learning to label sensor data as objective or automated measurement.

Predicting behavior. Machine learning can also be used to predict (to learn) future behavior based on sequences of past behavior. Examining early face-to-face interactions, for example, we found that both infants' smile initiations and terminations were most predictable when the infants themselves (rather than the mother) had most recently smiled or stopped smiling (Messinger et al., 2010). By contrast, mother smile initiations were most predictable in response to an infant smile. At the end of this chapter, we explore the potential of unsupervised machine learning to predict children's vocalizations with peers in social interactions.

The current chapter

Computational modeling of objectively measured social behavior is the focus of this chapter. We begin by considering how computational methods can help us understand development and developmental disabilities. Then we consider objective approaches to understanding attachment and predicting autism symptom levels. The bulk of the chapter is concerned with children's language, social behavior and development in a ubiquitous naturalistic

setting, the preschool classroom. We end the chapter by considering new directions for research in behavior imaging.

The Nested Timescales of Development

We are accustomed to studying behavior at a timescale of minutes and seconds while we assess developmental change over months and years. But developmental change occurs over all these nested timescales. What appear to be small changes in the context of tremendous minute-to-minute and day-to-variation yield striking growth curves over months of observation (Fogel & Thelen, 1987; Smith & Thelen, 2003; Spencer et al., 2011). The frequency and content of pre-linguistic infants' vocalizations during a brief observation, for example, vary as a function of the contingency with which their caregivers respond to specific vocalizations within minutes (Goldstein et al., 2009). Changes in vocalizations also unfold over longer timescales as infants gain new skills through the frequent repetition of vocal interactions in real time (Gilkerson & Richards, 2009; Paul et al., 2011). Measuring the interplay of behavior at multiple timescales requires repeated measurements of second-to-second real-time interaction over weeks and months. The economies of scale afforded by objective measurement are critical to making the study of child development at multiple timescales a tractable problem.

Computational approaches to developmental disabilities

Computational approaches to the study of child development are central to understanding individual differences, including those related to developmental disabilities. Developmental disabilities are life-altering impairments in motor, learning, behavioral, or communicative functioning. Karmiloff-Smith (2018) argues that understanding overlapping developmental cascades at multiple times scales is key to understanding atypical outcomes. The reverse is also true. Careful attention to real-time processes that support developmental change can reveal

similarities between children with and without developmental disabilities.

Autism spectrum disorder (ASD), is a high prevalence developmental disability characterized by social communication deficits and restricted and repetitive behaviors (American Psychiatric Association, 2013), which is a focus of our research. Another research focus, hearing loss (HL), limits access to auditory stimuli and spoken language input, which impacts children's developing language and literacy abilities (Blamey et al., 2001; Johnson & Goswami, 2010; Lund et al., 2015; Moeller et al., 2007; Niparko et al., 2010; E. A. Walker & McGregor, 2013). Both ASD and HL, as well as other developmental disabilities (O-DD) frequently have cascading consequences on children's long term development (Bat-Chava & Deignan, 2001; Hoffman et al., 2014; Ingvalson et al., 2020; Nittrouer et al., 2012).

We apply computational approaches both to the study of children with developmental disabilities and more typically developing (TD) children. An advantage of this strategy is that objective measurement allows researchers to temporally characterize behaviors and symptoms rather than simply noting their presence. Computer vision analyses of head movement applied to an early behavioral symptom of autism, for example, indicated delayed head turning in response to name (Campbell et al., 2019). Thus, computational approaches can quantify behavioral indices of a developmental disorder (latency to head turn) rather than relying on all-or-none ratings of its presence. Below we turn to a brief characterization of some of the computational measurement tools we have employed to characterize children's and adult's vocalizations in our research on attachment, ASD, and interactive behavior in schools (IBIS).

Tools of the trade: LENA

The Language ENvironment Analysis (LENA) system is commonly used by developmentalists to quantify the frequency of child and adult vocalizations occurring during

children's daily interactions (Gilkerson & Richards, 2009; Xu et al., 2009). The LENA system allows for efficient acquisition and automated analysis of day-long audio recordings from child-worn recorders. LENA is lightweight and fits within a pocket on the chest of custom-made vests and t-shirts making it easy to use in a variety of everyday settings, including children's homes, assessment settings, and preschool classrooms. LENA algorithms employ shallow machine learning, pre-trained Gaussian Mixture Models (GMMs) to detect speech, distinguish between speakers, and characterize speech. LENA specifically distinguishes between the speech of the focal child wearing the LENA recorder, and between other children and adults who are proximal to the focal child. In our research, we utilize LENA algorithms to classify focal child speech as speech-related (any phonemic content) or as fixed signals (e.g., crying), and to quantify the rate of these vocalizations.

Comparisons of LENA algorithms to expert human coders evidence moderate reliability with expert coders' classification of speech (Soderstrom & Wittebolle, 2013). Although LENA reliably estimates the relative quantity of focal child and adult speech, recent reports raise concerns about LENA reliability and misspecification of who is speaking (Cristia et al., 2021). To guard against misclassification, our research team outfits all children in a classroom with LENA recorders. Only vocalizations recorded from a child's own recorder are used to quantify child vocalizations. In addition, we assess the reliability of LENA classifications in the contexts we are studying. Using LENA-identified child and adult vocalizations, trained coders blind to LENA's designations listen to and classify individual vocalization segments as either belonging to a child or an adult. We find high levels of agreement (and Cohen's Kappa) between human and LENA categorizations of speech as child speech or adult speech in preschool inclusion classrooms for children with hearing loss, children with ASD, and children with other

developmental delays (O-DD; Fasano et al., 2021; Mitsven et al., 2021; Perry et al., 2021), suggesting the validity of these machine learning classifications. Perhaps more importantly, we and others find consistent associations between LENA's estimates of children's vocalizations and their language development outcomes (Dykstra et al., 2013; Gilkerson et al., 2018; Mitsven et al., 2021; Perry et al., 2018; Wang et al., 2020).

Tools of the Trade: Sphinx Phoneme Detection

While LENA quantifies vocalizations, it is based in part on Sphinx, an evolving software package that provides a window into vocalization quality (Walker et al., 2004; https://cmusphinx.github.io/). One measure of vocalization quality is the number and variety of phonemes vocalizations contain. Quantifying the phonemes present in children's vocalizations can provide insight into their expressive language development. We automatize the detection and estimation of the phonemic make-up of child and adult speech using Sphinx software. Sphinx uses Hidden Markov Models (HMM) to detect and quantify the number of consonants and vowels present within vocalizations (e.g., average count per utterance (ACPU; Lamere et al., 2003). Sphinx HMM estimates phonemes by mapping the speaker's specific production of the vocal signal to a best-fitting latent internal state that characterizes the probability of a given phoneme. Previous and subsequent internal states (generated sequentially in analyzing a vocalization) are employed to predict the current phoneme. In an overly simplified example, an internal state with a high weighting of the consonant /b/ might render a subsequent internal state with a high weighting of /i/ more probable.

Sphinx estimates of ACPU exhibit high concordance with human transcriptions of ACPU in child speech samples (Xu et al., 2014) and are statistically indistinguishable from expert transcriptions of speech in predicting children's later expressive language (Woynaroski et al.,

2017). Likewise, Sphinx estimates of the total number of consonants and vowels in child vocalizations predict children's future language abilities and distinguishes children who are at risk for language delays from their typically developing peers (Woynaroski et al., 2017; Xu et al., 2014). Our team has investigated the reliability of Sphinx's estimates of phonemes with expert human coders from LENA recordings of child and adult speech in inclusion classrooms for children with HL. Trained coders recorded the number of unique phonemes present in an equal number of child and adult vocalizations. High overall accuracy of Sphinx's phonemic diversity estimates of child and adult speech (Mitsven et al., 2021) highlight its potential to characterize children's language production and input.

Computational Approaches to Attachment and Autism Assessment

In this section, we look under the hood of two mainstay developmental and child clinical assessment protocols. The reported studies involve unobtrusively capturing relevant child and adult behavior using relatively inexpensive sensors such as Microsoft Kinect cameras, and head-mounted audio-video recorders. A range of computational approaches—some off-the-shelf and some adapted by our team—are used to measure the behavior of all individuals involved in the protocol. Objective measurement of the assessment of attachment behavior in the SSP and the assessment of ASD symptom severity in the ADOS-2 reveals that assessment protocols designed to focus on child behavior may also be influenced by the behavior of adults in the protocols. The results shed light on both child and adult contributions to child assessment paradigms.

Computational Approaches to Infant-Parent Attachment

Infant-parent attachment is a key index of social-emotional functioning and predictor of related outcomes. The gold standard for assessing attachment is the Strange Situation Procedure (SSP). During the SSP, videotaped attachment behaviors are evaluated by expert raters. Raters

observe the infant's interactions with the parent, response to a stranger, their reactions to the parent's departure from the room and, crucially, their behavior during the two reunions with the parent. The assumption of expert SSP rating is that ratings and attachment classification are a reflection of infant and not parent behaviors. It is noteworthy, however, that attachment theory derives in part from ethology. In the ethological tradition, interactive behavior is seen as the product of two or more individuals. Interaction is the braiding of individual behavior into dyadic tresses that cannot be decomposed into separate infant and parent contributions (Bowlby, 1982). However, the interactive basis of infant behavior is typically ignored during the measurement of attachment in the Strange Situation Procedure (Waters, 2002). The assumption is that the infant manifests their attachment-related orientation to the parent, and that the infant's behavior drives expert ratings of that behavior.

To better understand potential dyadic contributions to attachment behavior in the Strange Situation Procedure, we collected objective measures of parent and infant movement via multiple synchronized Microsoft Kinect devices, and processed infant vocalizations using LENA software (Prince, et al., 2021). We used these objective measures of movement and audio data to continuously quantify both infant and parent attachment-relevant behavior during the reunion episodes of the Strange Situation Procedure. We compared objectively measured attachment-related behaviors to expert attachment ratings scales (proximity-seeking, contact-maintenance, resistance, and avoidance) and dimensional summaries (approach/avoidance and resistance/disorganization dimensions) of attachment outcomes.

Objectively assessed dyadic variables (such as the proportion of time the infant was in close contact with the mother), and variables tapping mother behaviors (such as the velocity of mother moving toward and away from the infant) were involved in the prediction of almost all

“infant” attachment ratings. The prediction of attachment *dimensions* was particularly salient because these are alternative dimensional operationalizations of attachment security outcomes (Fraley & Spieker, 2003). Higher dyadic contact duration, higher velocity infant approach to the mother together with higher velocity mother approach to the infant explained more than two thirds of the variance in the central approach/avoidance dimension of attachment. These findings highlight the capacity of objective measures to reveal dyadic (and likely interactive) facets of attachment behavior in the gold standard assessment of infant attachment behavior (Prince, et al., 2021).

Another computational approach to the study of attachment comes from our use of regime switching models to capture “qualitative” changes in movement dynamics during the reunion episodes of the Strange Situation Procedure (S. M. Chow et al., 2018). Attachment theory suggests that infants engage in both comfort-seeking and exploratory behavior (Bowlby, 1982). These appear to be different ‘regimes’ of behavior with different goals and patterns of behavior. To explore these potential behavioral regimes, we applied regime-switching models, a novel modeling approach for representing abrupt changes in a system of differential equation models, to our objective measures of SSP behavior. We distinguished proximity-seeking regimes characterized by infants’ approaching the parent from exploration regimes characterized by infants’ movement away from the parent. Infant vocalizations (often cries) were likely to occasion a change from exploratory to proximity-seeking regimes. Moreover, such changes were more common during the second reunion when the infant attachment system is more activated. The results contribute to validating a key tenet of attachment theory by showing that infants use their vocalizations to help organize attachment seeking behaviors. Thus the application of advanced statistical techniques to multimodal objective behavioral measurements can speak

directly to developmental theory (S. M. Chow et al., 2018).

Objective measurement of autism severity

The characterization of autism spectrum disorder is a hotbed of computational approaches to child behavior (Dawson & Sapiro, 2019; Rehg et al., 2013). Digital behavioral phenotyping employs objective measurements of human behavior with the goal of providing fine-grained identification of diagnostically discriminative differences between individuals (Dawson & Sapiro, 2019). Our group, for example, applied audio signal processing to the vocalizations of infants who were at elevated likelihood of an ASD outcome (because they had an older sibling with ASD). These 'high-risk' baby siblings had higher pitched cries than typically developing infants during the separation phase of the Strange Situation Procedure. Among these infants, those who had ASD outcomes appeared to have the highest pitched cries (Esposito et al., 2014).

We are currently using objective measures of facial expressions and vocalizations to predict clinician assessment of autism symptom severity in preschool age children (Messinger, 2019). Assessment of autism spectrum disorder (ASD) traditionally relies on expert clinician observation and judgment, often of rather restricted samples of child behavior. Thus, we focused on detecting autism-related behaviors in the relatively constrained clinical setting of ASD diagnosis. Diagnostic evaluations for ASD typically include the Autism Diagnostic Observation Schedule (ADOS-2), a semi-structured assessment in which the examiner presents a series of social presses, activities such as activating a remote-controlled bunny, which are designed to reveal autism symptoms (such as a lack of demonstrative pointing). The ADOS-2 yields provide overall symptom severity scores in the domains of Social Affect (SA) and Restricted and Repetitive Behaviors (RRB). The vast majority of our sample met criteria for ASD.

Consequently, we adopted the rather difficult task of using objective characterization of behavior in that clinical setting not to distinguish children with and without ASD, but to predict SA and RRB symptom severity.

Objective measures of child and adult vocalization characteristics as well as automatically detected facial expressions in children were used to predict clinician-rated autism symptom severity. ADOS-2 evaluations were captured using examiner- and parent-worn video-enabled eyeglasses. Faces in each frame of the captured videos were first detected by a pre-trained convolutional neural network or CNN (MTCNN; Zhang et al., 2016), and the age of each detected face was estimated by another CNN, Deep Expectation (Rothe et al., 2018). Using a gaze detection algorithm validated by Chong et al. (2020) we detected children's social gaze (gaze at an adult) from video segments of detected child faces. and we used commercial software, the AFFDEX system (McDuff et al., 2016), to detect child smiles (see Figure 1). Both social gaze toward the parent and the proportion of that gaze that involved smiling (Facial Action Coding System Action Unit 12, lip corner raising) were unique predictors of social affect symptoms. As expected, children with higher levels of social affect symptoms engaged in less social gaze and that gaze was less likely to involve sharing a smile (Ahn, et al., 2021).

Motivated by our earlier research suggesting that vocalization quality may index autism risk (Esposito et al., 2014), we analyzed the audio recordings of the ADOS evaluations (Moffitt et al., 2021). Using Praat speech analysis software (Boersma & Weenik, 2020), we obtained the mean fundamental frequency (F0) of LENA-identified child speech-like and cry vocalizations recorded by the audio recorder in the examiner- and parent-worn glasses. Sphinx estimated the average count of phonemes per vocalization for LENA-detected child and adult vocalizations. We examined the association between these objective measures of vocalizations and children's

calibrated severity scores on the ADOS. A smaller number of child phonemes per vocalization, and child vocalizations with a higher fundamental frequency (pitch), uniquely predicted higher levels of autism symptom severity in the repetitive behavior domain. The association of vocal variables with repetitive behavior symptoms parallels the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) description of this domain.

With respect to dyadic features of vocal behavior, the fundamental frequency of adult vocalizations (higher pitch) functioned in a parallel fashion to the pitch of child vocalizations in the regression equation predicting repetitive behavior symptom severity (Moffitt et al., 2021). The predictive value of adult vocalization pitch (see also Bone et al., 2014) may tap into the responsiveness of adults to more symptomatic children. Specifically, adults may use a more active approach, including the use of more high pitch child-directed speech, to engage more symptomatic children in social interactions.

Overall, these results suggest objective quantification of vocalizations (which are associated with repetitive behavior symptoms) and automated measurements of social gaze and smiling (which are associated with social affect symptoms) characterize complementary diagnostic features of the ASD behavioral phenotype. In clinical settings, objective characterizations of behavior could provide examiners with supplementary information about the severity of ASD symptoms. Moreover, objective measurement approaches can economically be applied to large samples of behavior in multiple contexts to assess the degree to which symptom expression is evident and sustained across settings.

Interactive Behavior in Schools (IBIS): Computational Approaches to the Study of Preschool Classroom Behavior

Studying development *in situ* allows us to observe how behavior and interaction in one

moment gives rise to changes over longer periods of time, facilitating an understanding of the nested timescales of developmental process. As over 60% of 3-5 year-olds attend preschool (Irwin et al., 2021), preschool classrooms represent a key context for children's development. In preschool, children experience a range of emotions, and engage with learning opportunities, peers, and teachers. They develop self-regulatory and perspective-taking abilities in dyadic, small-group, and large group activities. Inclusion—in which children with developmental disabilities are educated in the same classrooms as typically developing children—is a gold standard intervention. Thus, interaction in inclusive classrooms provides a window into the diversity of behavior that is key to understanding the mechanisms driving development.

The labor-intensity of collecting expert observations of behavior has historically been an obstacle to conducting preschool research in both inclusive and non-inclusive classrooms. Moreover, expert observations of interactions can only capture a portion of ongoing classroom social activity. However, recent advances in sensing technologies allow for continuous measurement of classroom interaction, allowing for new insights into classroom behavior and development (Gonzalez Villasanti et al., 2020). Our group is applying these approaches to the study of language and interaction in inclusive preschool classrooms and classrooms for children with developmental delays. We typically conduct objective measurements over the course of the school day (which is often a half-day for preschoolers) at weekly to monthly intervals during the school year. Measurements typically involve all of the children and teachers in the classroom although occasionally a family does not consent to the research protocol or a child does not want to wear their instrumented vest. Children's language abilities are assessed using standardized instruments (typically, the Preschool Language Scales Fifth Edition, PLS-5) at the end of the school year and, increasingly, at the beginning of the year as well. In the following sections, we

review how our work in this domain has expanded our understanding of children's social interactions and language development.

Computing classroom language

LENA technology allows us to simultaneously record the speech of multiple children and teachers over the school day. Our initial work with LENA indicated that children who were exposed to higher rates of peer vocal input and children who engaged in a higher number of conversational turns with their teachers tended to vocalize more themselves (Perry et al., 2018). Further, vocal input from peers and more conversational turns with teachers were associated with greater expressive vocabulary gains (Perry et al., 2018). This research highlighted the role of both peers and teachers in contributing to children's classroom language development, an issue which remains the topic of vigorous research. We have employed two major strategies for understanding classroom language. For the most part, we have examined the frequency of child and teacher vocalizations, their rate of talking. However, we have also investigated the phonemic richness of those vocalizations (the number of speech sounds they contain), an investigation we detail below.

Phonemic production: Teacher to child and child to teacher

Our team supplemented LENA's quantification of the number of teacher and child vocalizations in the classroom with Sphinx estimates of the phonemic composition of those vocalizations to compare the relative contribution of the rate and phonemic diversity of vocalizations to children's language abilities at the end of the school year (Mitsven et al., 2021). Children who were exposed to a higher rate of vocalizations per minute from their teachers produced a higher rate of vocalizations per minute themselves. Similarly, children who were exposed to more phonemically diverse speech from teachers (vocalizations containing a higher

number of unique phonemes), produced more phonemically diverse speech. These associations were observed both in children with hearing loss as well as their typically developing peers, highlighting the importance of teacher speech in influencing the language production of all children in the classroom.

Crucially, we found that the phonemic diversity of children's own speech was a stronger predictor of children's end-of-year language abilities than the rate of their speech. Additionally, the phonemic diversity of children's own speech mediated the relationship between the phonemic diversity of their language input from teachers and their end-of-year language abilities (see Figure 2). These findings highlight the role of teachers in supporting children's spoken language skills. They underscore the importance of high-quality language production in predicting language abilities both for children with hearing loss and typically developing children. Finally, the mediation analyses indicate how children's language environments can support their own linguistic productions which, in turn, scaffold their developing language competencies (Mitsven et al., 2021).

The temporal structure of classroom language.

Speech sounds are functionally flexible and can express a variety of meanings while cries express specific emotional content. But both types of vocalizations are embedded in time (Messinger, Custode, et al., 2019). Automated measurement allows us to ask about the utility of two measures of the temporal structure of vocalizations, burstiness and temporal dependency. The burstiness coefficient assesses temporal patterning over relatively long timescales. Burstiness pertains to the time between the onsets of behavioral events such as vocalizations. Of note, the time between the onset of behavioral events involves both the event and the following non-event (e.g., the time between vocal onsets includes the duration of the vocalization and the

duration of the silence that follows). By contrast, the temporal dependency measure ignores non-events. Temporal dependency can be expressed as a correlation that quantifies the degree to which the duration of one event predicts the duration of the next. Thus, temporal dependency quantifies relatively short-term dependencies in the duration of behavioral events such as vocalizations.

We examined burstiness and temporal dependency in infants from 10 to 35 months of age over a two year period (Messinger, Custode, et al., 2019). Burstiness and temporal dependency were not associated at the level of the child, suggesting the relative independence of these indices of temporal structure. Cry vocalizations exhibited higher levels of temporal dependence than speech-like vocalizations and levels of temporal dependence between the two vocalization types were not associated. That is, the duration of vocal expressions of infant negative affect predicted the duration of the next such expression at higher levels than was the case for presumably more affectively neutral speech-like vocalizations. By contrast, levels of cry and speech-like vocalization burstiness were associated, but only the burstiness of infant speech-like vocalizations was associated with the rate of infant-teacher conversational turns. This suggests that vocal interactions may play a role in structuring the onsets of infants' functionally flexible speech-like production over time. Thus, the application of two complementary computational indices of temporal structuring contributed to a better understanding of the function of different types of infant vocalizations in naturalistic groups. Next we turn from individual child vocalizations to dyadic vocalization patterns.

Social interactions and speech

Previous research indicates that children with stronger language abilities tend to engage in a higher number of positive interactions with peers (see Chow, 2018 for a review). This

research, however, focuses on children's overall interactions with all peers, rather than focusing on individual children and their specific peers. That is, the role of both dyad members in co-constructing language experiences and interaction has been relatively neglected. To address this issue, our team asked how children's classroom language production was associated with the social interactions of specific dyads. Using LENA algorithms, we examined associations between the frequency of child vocalizations and teacher input and expert coded observations of children's positive and negative interactions with peers and teachers (Custode et al., 2020). We explored the contributions of both the focal child's relative level of talkativeness and each of their partners' relative level of talkativeness to understand how dyadic language levels were associated with the tenor of interaction. Dyads made up of two highly talkative children had the most positive interactions and those made up of two relatively untalkative children had the least positive interactions. Notably mixed dyads (composed of a highly talkative and a relatively untalkative child) exhibited intermediate levels of positive interactions, suggesting the likelihood of a dyad having a positive interaction is related to their combined language skills. This pattern of results, stemming from the use of automated measurements of language in day-long classroom recordings, highlights the dyadic nature of peer social interaction.

Who are you talking to? Integrating objective measures of social contact

Previous approaches to studying the impact of peer language on child development often involve either labor intensive coding of peer interactions in the classroom, from brief observations of one child. An alternate approach is averaging the language skills of all the children in a classroom as a proxy for the language resources that an individual child experiences. Missing from both of these approaches is a comprehensive real-time understanding of the many simultaneous language-mediated interactions that occur in a classroom and how

those interactions change over time. In the research below, we capture children's interactions by integrating first-person audio recording of language with motion tracking. We begin by examining motion tracking as a stand-alone approach to understanding children's social lives.

Real-time movement to understand sociality

Continuous real-time measurement of children's location and movement offers powerful insights into children's social interaction. Our team utilizes Ubisense's ultrawide radio frequency identification (RFID) system to objectively measure social movement and interaction in preschool classrooms (Ruiz et al., 2017; see Figure 3). Ultrawide RFID tags can be used to track children's and teachers' location and orientation at 2-4 Hz to an accuracy of 15-30 cm by means of triangulation and time of arrival differences (Irvin et al., 2018; Phebey, 2010). An alternative to ultrawide RFID location-based tracking is proximity-based tracking via Bluetooth or similar technologies. Using Bluetooth based tracking, for example, Veiga and Ward (2016) found that the average duration of a child's contact with other children was positively associated with teacher ratings of the child's social competence. In such Bluetooth-based, proximity-based tracking, one knows when individuals are close to one another based on a priori parameters embedded in the badges and their technical capabilities (the distances and orientations at which badges are able to transmit to one another). Unlike in our use of location-based tracking, however, one does not know where children are located in classroom space, nor can one use data-driven methods to determine when children are in social contact or how they approach one another.

Computing social approach: Homophily effects

We have used ultrawide RFID location-based tracking to explore social approach in children with and without ASD. ASD is defined in part by disorders of social interaction that

include atypicalities in social approach such as movement toward peers and adults (Hilton et al., 2012; Schoemaker & Kalverboer, 1994). To investigate social approach, we tracked all children in inclusion classrooms containing children with ASD and other developmental disabilities (O-DD), as well as typically developing (TD) children (Banarjee, et al., 2021). We first ruled out the possibility of overall differences in the velocity of general (non-social) movement. There were no overall velocity differences in the classroom movement of children with ASD, their TD peers, and their peers with O-DD. There were, however, differences in social approach. To understand these differences, we introduce homophily—the tendency of children to interact with children who are similar to themselves on dimensions such as gender and disability status. Homophily is a driving force in children's social interactions in and out of the classroom (Laursen, 2017; Martin et al., 2005). However, evidence for homophily in the classroom is typically constrained by human observation and large-scale research is typically dependent on teacher reports of peer play preferences.

We found striking homophily effects in the velocities at which children approached one another (Banarjee, et al.,2021). Children with ASD and O-DD did not differ from TD children in the velocities with which they approached other children. Instead, effects were dyadic. Children in concordant dyads (e.g., ASD-ASD) approached one another at higher velocities than discordant dyads (e.g., TD-ASD). With respect to social approach, then, ASD deficits in velocity were not evident. Instead, dyads concordant on disability status interacted with one another at higher velocities than discordant dyads, indicating early homophilic patterns of social movement in all three groups.

Computing social contact

Continuous measures of real-time movement led us to ask how we could determine when

children were in interaction with one another. At what distances or orientations can we say that two children are in social contact? We turned to statistical physics for insight. In physics, the radial distribution function, g(r), is used to ascertain the differential structure of molecules in liquids and solids such as crystals (Chandler, 1987). To detect social interaction, we applied the radial distribution function to index when children were in proximity at greater than chance levels (see Figure 4).

Specifically, we used ultra-wide RFID measures of location to calculate the *observed* distance of each child from every other child over time. Crucially, we compare these observed distances to the distances expected by *chance*. The distance expected by chance is calculated from a null model of the cumulative probability of a child being in a given area of the classroom over an entire observation. One can think of this null model as a heat map of areas in which children tend to be located. From this static model, we calculate the chance likelihood of pairs of children being at a continuous range of distances from one another (e.g., .1 meters, .2 meters, and so on; see Figure 4). By dividing the observed proximities of a given distance by the null model for that distance, we derive the radial distribution function. When the function exceeds 1, the probability that two children are to be found within the specified distance is higher than expected by chance. Consequently, when the function exceeds 1, we infer that a pair of children are in social contact, that the children are proximal *because* of the other's presence.

Gendered social contact in the classroom

We first applied the radial distribution function during free play in a kindergarten classroom in which the 5-year-olds wore Ubisenese RFID tags on their wrists (Messinger et al., 2019). The radial distribution function indicated that children were in social contact when they were within 1 meter of one another. Rank-order plots indicated that children's levels of social

contact with individual peers (ordered from most contact to least) showed exponential decay (see Figure 5). That is, children spent tens to hundreds of times more with some peers than others, a relatively unexplored feature of classroom social life (Chen et al., 2019, 2020). Social network analyses suggested moderately strong sex segregation. Children tended to engage in more social contact with same-sex than other-sex peers (i.e., boy-boy and girl-girl pairing). Thus continuous tracking of children's location revealed the exponential structure of children's social contact with peers and shed new light on gender homophily (Messinger et al., 2019).

Integrating orientation: Homophily and social contact

Proximity is an essential social contact criterion, but interpersonal orientation is also relevant. It is unlikely, for example, that two children with their backs to one another are in social contact. To incorporate interpersonal orientation, we equipped children with two hip-worn RFID tags (rather than a single tag as in Messinger et al., 2019). Having a left-side and a right-side tag enabled us to calculate children's relative angular orientation to one another (see Figure 6). Visualizations of children's angular orientations appeared to support our surmise that face-to-face contact was a privileged interpersonal orientation for interaction (see Figure 7). Consequently, we added an angular orientation criterion to the data-derived proximity criterion derived from the radial distribution function. In subsequent research, social contact occurred when two criteria were met: Children were within the proximity limit and they were more-or-less oriented to one another.

We first applied the dual social contact criteria (proximity and orientation) in inclusive classrooms for children with ASD, other development delays (O-DD), and typically developing children. There were powerful homophily effects such that children spent more time in social contact with children in the same disability group (ASD-ASD, O-DD/O-DD, and TD-TD) than

with other children. There also appeared to be disability group effects such that, for example, children with ASD and with O-DD spent less time in social contact than TD children, but the introduction of interaction terms rendered the main effects of disability group non-significant. Instead, the interaction terms indicated that among concordant dyads, ASD-ASD and O-DD/O-DD dyads were in social contact less than TD-TD dyads. ASD-specific deficits in social contact might be expected but instead we found that variations in homophily structured social contact. Children with ASD were in social contact with other children with ASD less than TD-TD dyads, and O-DD dyads showed a corresponding pattern. The results suggest the potential of a new generation of objective measures to extend our understanding of dyadic patterning of children's social lives in preschool inclusion classrooms.

Social Contact and Vocal Interaction

Dyadic conversations

To expand our understanding of linguistic interaction and language development, we sought to integrate classroom language measures with objective tracking of child movement (see also Irvin et al., 2021). Although LENA algorithms identify target child vocalizations and distinguish whether a child's partner is another child (a peer) or an adult, they do not distinguish *who*, specifically, the partner is. To understand who is speaking to whom we synchronized Ubisense ultrawide RFID tag measures of social contact (proximity and orientation to peers) with LENA vocalization measures.

As a first step, we found that objective measurements of children's conversations are associated with sociometric indices of peer relationships and observational measures of classroom engagement. We compared preschool teacher and child reports of peer friendship with the average *dyadic* rate of vocalizing in social contact (i.e., objectively measured peer

interaction). Dyads who engaged in higher rates of vocalizing in social contact were more likely to endorse each other as friends and were more likely to be rated as friends by their teachers (Altman et al., 2019). Likewise, using the inCLASS observational coding system, expert ratings of children's engagement with peers and teachers were associated with children's speech to those partners (Fasano, Perry, Custode, et al., 2021). The results suggest that children's vocal interactions with peers index their friendships, and that children's vocalizations to partners are an index of their interpersonal engagement.

We next considered children's vocalizations in social contact with peers to substantively understand the flow of classroom language. Consistent patterns of dyadic conversation over time were evident in analyses. The more vocalizations a child heard from a peer on one observation, the more vocalizations they produced to that peer on the following observation (see Figure 8). That is, A's level of speech to B predicted B's level of speech to A in the next observation. Analyses controlled for B's previous speech to A (an autoregression effect). Observations occurred every 1 to 4 weeks, but the predictive power of peer vocalizations did not vary based on the time between observations. This finding was evident in inclusion classrooms containing children with hearing loss and their typically hearing peers (Perry et al., 2021), and in classrooms containing children with ASD, other developmental disabilities (O-DD) and their typically developing peers (Fasano, Perry, Custode, et al., 2021). Moreover, no differences were apparent in the strength of the reciprocal peer vocalization effect between TD children and children with any category of developmental disability.

The findings suggest a reciprocal pattern of dyadic vocalization for all the children in the classroom. That is, A's speech to B predicts B' speech to A which, *in turn*, predicts A's speech to B. To ascertain the impact of peer speech in inclusion classrooms for children with hearing

loss, we divided earlier from later observations (e.g., fall from spring) and predicted children's end-of-year language abilities in mediation analyses. Vocalizations from peers had an indirect association with children's end-of-year language abilities as mediated by children's vocalizations to peers (Perry et al., 2021). The results simultaneously suggest the influence of peer speech and indicate the impact of children's own speech, their own language production, in predicting assessed language abilities (Hirsh-Pasek et al., 2015; Ribot et al., 2018).

Previous research on peer effects indicated that average classroom language ability predicts children's language gains (both measured with standardized assessments) over the school year (Justice et al., 2011). The current computational approach to assessing children's vocal interactions with individual peers suggest a mechanism for the previously documented effects. Within-dyad transmission of speech over the school year appears to lead to individual child language gains. Speech *from* peers appears to be a vehicle for increasing children's subsequent vocalizations to their peers, which provides them with an opportunity to exercise and consolidate language competencies. That is, speech to peers may provide young children opportunities to use words and grammatical constructions that scaffold their developing language abilities. These results suggest the importance of dyadic conversations in shaping children's language capacities over developmental time, but they do not speak to social groupings larger than the dyad.

Social Language Networks

The totality of dyadic interactions between peers constitutes a social network. Children with developmental disabilities and language delays are more peripheral to social networks than their typically developing peers (Chen et al., 2019, 2020; Locke et al., 2013). These findings, however, are dependent on child friendship ratings and teacher ratings of children's interactions

and relationships. While valuable, these approaches offer only a (potentially biased) static snapshot of peer relationships (Shin et al., 2014). In our initial research, social networks were based only on social contact (Messinger et al., 2019). In recent work, automated measures of vocalization frequency in social contact were implemented in inclusion classrooms for children with ASD and other developmental disorders (O-DD) over multiple days of observation (Fasano, Perry, Zhang, et al., 2021). In these networks, nodes are connected by edges weighted by the vocal interaction between the two children (the sum of the rate of the vocalizations made by each child to the other; see Figure 9).

Network analyses (Fasano, Perry, Zhang, et al., 2021) indicate that, as a group, children with ASD exhibited lower modularity (cohesiveness) than children with O-DD or TD (who did not differ from one another in modularity). Modularity indexes the mean weight of edges between group members. Likewise, ASD-TD modularity (edges between children in these groups) was lower than O-DD/O-DD and TD-TD modularity. Thus, continuous measures of vocal interaction suggest that even in inclusion classrooms designed to foster between-group interaction, children with ASD can be isolated from other children’s conversations. To understand the effects of this isolation, we examined the association between degree centrality (the total frequency of child vocalizations to and from peers in social contact) and assessed end-of-year language abilities. Even when accounting for group differences, degree centrality was positively associated with assessed language abilities, highlighting the role of real-time peer vocal interactions in scaffolding developing language capacities. The results speak to the power of network analyses of children’s objectively measured conversations over the course of the school day to reveal group structures that reflect children’s access to peer-based language resources in the classroom.

As we hope to have illustrated, continuous objective measurement has the potential to both differentiate children based on their developmental disabilities, and to identify shared developmental pathways among children with and without disabilities. Identifying specific deficits in behavioral interaction (speech in social contact, for example) that distinguish children with a specific developmental disability can elucidate intervention approaches for those children in their naturalistic environments. Likewise, when speech in social contact does occur its influence appears to be similar across groups, suggests the potential for wide use of interventions targeting that type of social interaction.

Future Directions

Understanding children's language

In our preschool research, we have examined classroom language with respect to the rate per minute of child and adult speech, the number and diversity of phonemes in that speech, conversational turn counts between children and adults, and explored the short-term and long-term temporal structuring of that speech. This has produced insights into peer and teacher sources of children's language development and the social structure of classrooms (Fasano, Perry, Zhang, et al., 2021; Mitsven et al., 2021; Perry et al., 2018). However, questions remain about the best way to record and analyze speech, and how to best investigate whether sociality indeed drives language development in early social groupings like preschool classrooms.

Off-the-shelf LENA software is frequently used by developmentalists to record and then classify adult and child speech segments in audio recordings. Although LENA recorders are small in size and frequently used in developmental research, they record audio monophonically on a single track. A promising alternative is to acquire two audio tracks stereophonically, which can facilitate later sound source separation and denoising (Alamdari et al., 2021; Y. Yu et al.,

2016). Stereo audio recording may be particularly important in noisy environments.

Having recorded home, clinic, or classroom speech, how should researchers quantify the vocalizations and vocal interactions they wish to study? While comparisons of LENA classification to expert human coders, including our own work, show acceptable reliability for specific classification tasks, LENA algorithms also exhibit bias, at least in home and home compound settings. LENA algorithms tend to attribute female adult vocalizations to males, and tend to attribute the vocalizations of focal children (children wearing the LENA recorder) to other children in the audio frame (Cristia et al., 2021). Thus, in classrooms at least, pairing LENA recorders with location tracking may be important to clarifying whether the focal child or a peer is speaking (Fasano, Perry, Custode, et al., 2021; Irvin et al., 2021; Perry et al., 2021).

It is also of note that LENA software uses shallow machine learning models to distinguish speechlike vocalizations and to distinguish between speakers (Gilkerson & Richards, 2009; Xu et al., 2009). As noted, shallow models are trained on selected features of the audio signal rather than the signal itself. Deep learning methods, however, exhibit superior performance to conventional shallow machine learning methods in many applications including audio processing (Lahiri et al., 2020; Lecun et al., 2015). ALICE—Automatic LInguistic unit Count Estimator—is a software toolkit that performs diarization, speaker classification, and phoneme, syllable, and word counts based on deep learning (Räsänen et al., 2021). ALICE provides more accurate measures of adult speech from child-centered daylong recordings than LENA, and represents a promising approach to characterizing classroom language.

Our current models indicate how speech between children and their preschool partners is reciprocated over weekly to monthly observation and from the spring to fall semesters of the preschool year (Fasano, Perry, Zhang, et al., 2021; Mitsven et al., 2021; Perry et al., 2021). But

these models do not speak to real-time vocal interaction. We and others have found that the quantity of LENA-identified adult-child conversational turns are associated with children's reported vocabulary development and language outcomes (Gilkerson et al., 2018; Perry et al., 2018; Romeo et al., 2018). A difficulty with these adult-child conversational turns, however, is that they are conceptually tied to (and strongly correlated with) the frequency of children's own vocalizations (Harbison et al., 2018). By contrast, a partner's responsive vocal feedback to a child's vocalizations is distinct from the vocalizations themselves. Moreover, responsive feedback is a driver of child speech in multiple contexts among children with and without ASD (Cabell et al., 2015; Harbison et al., 2018; McDuffie & Yoder, 2010; Warlaumont et al., 2014). Little is known, however, about the role of responsivity in classroom speech (Hindman et al., 2019). The role of teacher and peer verbal responsivity in fostering children's real-time language production and long-term language development is a promising topic for future research.

A full understanding of children's linguistic interaction will require understanding the content of speech, *what* is being said. Commercial speech to text services offer a machine learning approach to understanding what children and teachers are saying to one another (Chiu et al., 2018). Classrooms are noisy environments, however, and it is unclear when such an approach will be viable. Nevertheless, in a relevant pilot study, investigators used head-mounted video cameras to capture children's first-person experiences. Individual participant faces were identified using Amazon Rekognition processing. Crucially, vocalizations were identified by Amazon Transcribe but then manually coded as being spoken by a teacher, focal child, or peer. Integration of video and audio coding nevertheless yielded promising levels of accuracy in identifying expert human-coded interaction.

Understanding social contact

The radial distribution function, g(r)—and visualizations of children's relative orientation—are typically calculated for all children in the classroom. Yet individual differences between children—for example, disability group—may condition these values. That is, different groups of children may differ in the data-driven parameters (distance and orientation) that define their social contact. Likewise, it remains to be seen whether these differences are individual (e.g., children with ASD are in social contact at longer distances) or dyadic (e.g., children with ASD are in social contact at longer distances when interacting with other children with ASD).

In our work, we have primarily determined social contact using a combined proximity-orientation criterion. However, the orientation criterion has relied on a priori decisions about what constitutes social interaction (i.e., approximately face-to-face mutual orientation between partners). We intend to explore a data-driven approach to determining orientation criteria akin to our radial distribution function used to determine proximity criteria (see Figure 7). That is, when children are within the social contact proximity criterion, we can ask whether the probability of an observed relative orientation exceeds that expected by chance. This procedure provides a data-driven index of when the relative orientation of two children is related to the presence of the other.

More fundamentally, however, social contact criteria based on distance and relative orientation are associated. In pilot analyses, we have synthesized the distance and relative social contact criteria in a joint likelihood distribution of distance and relative orientation (see Figure 10). The joint likelihood distribution suggests that distance and relative orientation criteria are associated. The results indicate the productive potential of computational processing of objective tracking of physical position to understand children's social interaction.

Understanding deep learning

Developmental scientists have increasing access to large volumes of behavioral data collected by sensors such as video cameras, microphones, and RFID tags and badges. Applied to this sensor data, deep learning techniques are increasingly able to produce reliable objective measurements of children's behavior during interactions with peers, teachers, and parents in naturalistic, assessment, and other contexts. Two examples from autism assessment are of note. Deep neural networks are able to differentiate child and adult speech more accurately than more conventional approaches (Lahiri et al., 2020). Likewise, accurate, automated gaze detection from first-person video is now possible with models based on convolutional neural networks trained on a large-scale expert-annotated dataset (Chong et al., 2020).

The potential of deep learning models to understand and predict behavior is as important as their role as measurement tools. An instructive example is our adoption of an inverse optimal control approach to understanding infant-mother, face-to-face smiling interactions (Ruvolo et al., 2015). The premise of the approach is that infants' and mothers' probable goals during interaction can be inferred from the historical consequences of their smiling actions over time. Posterior probabilities indicated that mothers preferred states of mutual smiling. By contrast, infants—who often smile briefly, elicit a mother smile, and then stop smiling—preferred states in which mothers smiled but they did not. Thus, sophisticated learning models can provide new insights into interactive behavior.

In the realm of prediction, deep learning models can use prior behavior as input to "learn" subsequent behavior (Abduallah et al., 2020; Butepage et al., 2107; K. Yu et al., 2017) . In pilot work we applied deep learning models to predict the quantity of children's vocalizing in preschool inclusion classrooms. Based on the previous movement and audio data of children and teachers, convolutional neural networks and recurrent neural networks accurately predicted the

quantity of children's future vocalizations. The results are important because they represent data-based, assumption-free, prediction of speech, a developmentally central child behavior.

Yet the potential predictive strength of deep learning models may come at a cost. Deep learning models typically lack transparency and interpretability (Olah et al., 2018). That is, the complex, multilayer, nonlinear, sometimes iterative relationship between the model inputs and predictions resists human understanding. Reduced interpretability makes it more difficult to gain insights into interaction—the factors, for example, that motivate children's classroom vocalizations—from deep learning models. Consequently, explaining deep neural network predictions and building interpretable deep neural networks is a topic of ongoing investigation (Afrabandpey et al., 2020; You et al., 2021). A promising direction of research, then, is ascertaining the limits of interpretable deep learning in predicting and understanding early interaction and development.

To conclude, it is tantalizing to consider similarities between the developing child and "developing" deep learning algorithms, specifically various types of dynamic neural networks. The analogy here is to dynamic neural networks generally, without focusing on networks constructed to specifically simulate early developmental processes in a neurally plausible fashion (Bhat et al., 2021; Buss, 2017; Perone & Simmering, 2017; Sen et al., 2020; Thibodeau et al., 2020). Dynamic neural networks develop in a fashion similar to human infants. They categorize—sometimes in a supervised fashion but more frequently in an unsupervised fashion—vast quantities of sensory data. Like these neural nets, human infants generate remarkable developmental results in a fashion that is somewhat resistant to mechanistic understanding. A key difference, of course, is that human infants are embodied and motivated to act on the world. Our hope, nevertheless, is that inquiries into these two types of developing systems continue to

fertilize one another.

References

Abduallah, M., Quian, K., Elhoseiny, M., & Claudel, C. (2020). Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14424–14432.

Afrabandpey, H., Peltola, T., Piironen, J., Vehtari, A., & Kaski, S. (2020). A decision-theoretic approach for model interpretability in Bayesian framework. *Machine Learning*, *109*(9), 1855–1867.

Ahn, Y. A., Moffitt, J., Custode, S., Beaumont, A., Cardona, S., Parlade, M., Durocher, J., Hale, M., Alessandri, M., Perry, L., & Messinger, D. (2021, May). Associations between objectively measured facial expressions during the ADOS-2 and the calibrated severity scores in 3-year-olds with suspected ASD [Poster presentation]. The International Society for Autism Research (INSAR) 2021 Annual Meeting. (Virtual Conference).

Ainsworth, M. D. (1985). Patterns of attachment. *Clinical Psychologist*, *38*(2), 27–29.

Alamdari, N., Azarang, A., & Kehtarnavaz, N. (2021). Improving deep speech denoising by Noisy2Noisy signal mapping. *Applied Acoustics*, *17*, 107631.

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

Banarjee, C., Tao, Y., Fasano, R., Vitale, L., Song, C., Shyu, M.L., Perry, L.K., Messinger, D.S. (2021). Quantifying Social Approach in Preschoolers with ASD in Inclusion Classrooms. Poster presented at the Meeting of the International Society for Autism Research (INSAR).

Bat-Chava, Y., & Deignan, E. (2001). Peer Relationships of Children With Cochlear Implants. *Journal of Deaf Studies and Deaf Education*, *6*(3), 186–199.

https://doi.org/10.1093/deafed/6.3.186

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2021). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*, *Advance on*. https://doi.org/https://doi.org/10.1037/rev0000313

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., Psarros, C., Rattigan, K., & Tooher, R. (2001). Relationships among Speech Perception, Production, Language, Hearing Loss, and Age in Children with Impaired Hearing. *Journal of Speech, Language, and Hearing Research*, *44*(2), 264–285. https://doi.org/10.1044/1092-4388(2001/022)

Boersma, P., & Weenik, D. (2020). *Praat: Doing phonetics by computer [Computer program]*.

Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. (2014). The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1162–1177. https://doi.org/10.1044/2014_JSLHR-S-13-0062

Bowlby, J. (1982). *Attachment* (2nd ed.). Basic Books.

Buss, A. T. (2017). Computational Models of Executive Function Development. In *Executive Function* (pp. 124–144).

Butepage, J., Black, M. J., Kragic, D., & Kjellstrom, H. (2107). Deep representation learning for human motion prediction and classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6158–6166.

Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J., & Forston, L. D. (2015). Teacher-child conversations in preschool classrooms: Contributions to children's vocabulary

development. *Early Childhood Research Quarterly*, *30*(PA), 80–92. https://doi.org/10.1016/j.ecresq.2014.09.004

Campbell, K., Carpenter, K. L. H., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., Chang, Z., Qiu, Q., Vermeer, S., Adler, E., Tepper, M., Egger, H. L., Baker, J. P., Sapiro, G., & Dawson, G. (2019). Computer vision analysis captures atypical attention in toddlers with autism. *Autism*, *23*(3), 619–628. https://doi.org/10.1177/1362361318766247

Chandler, D. (1987). Introduction to Modern Statistical Mechanics. Oxford University Press.

Chen, J., Justice, L. M., Rhoad-Drogalis, A., Lin, T.-J., & Sawyer, B. (2020). Social Networks of Children with Developmental Language Disorder in Inclusive Preschool Programs. *Child Development*, *91*(2), 471–487. https://doi.org/https://www.google.com/url?q=https://doi.org/10.1111/cdev.13183&sa=D&source=docs&ust=1637553057859000&usg=AOvVaw3uN9vxaLvruxWJia1NbCJN

Chen, J., Lin, T. J., Justice, L., & Sawyer, B. (2019). The Social Networks of Children With and Without Disabilities in Early Childhood Special Education Classrooms. *Journal of Autism and Developmental Disorders*, *49*(7), 2779–2794. https://doi.org/10.1007/s10803-017-3272-4

Chiu, C.-C., Sainath, T. N., Wu, Y., Prabhavalkar, R., Nguyen, P., Chen, Z., & Kannan, A. (2018). State-of-the-art speech recognition with sequence-to-sequence models. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 4774–4778.

Chong, E., Clark-Whitney, E., Southerland, A., Stubbs, E., Miller, C., Ajodan, E. L., Silverman, M. R., Lord, C., Rozga, A., Jones, R. M., & Rehg, J. M. (2020). Detection of eye contact with deep neural networks is as accurate as human experts. *Nature Communications, 11*(1),

6386. https://doi.org/10.1038/s41467-020-19712-x

Chow, J. C. (2018). Comorbid language and beahvior problems: Development, frameworks, and interventions. *School Psychology Quarterly*, *33*(3), 356–360. https://doi.org/https://www.google.com/url?q=https://doi.org/10.1037/spq0000270&sa=D&source=docs&ust=1637550027541000&usg=AOvVaw0hNA72avp3SxIT9WrEdeE-

Chow, S. M., Ou, L., Ciptadi, A., Prince, E. B., You, D., Hunter, M. D., Rehg, J. M., Rozga, A., & Messinger, D. S. (2018). Representing Sudden Shifts in Intensive Dyadic Interaction Data Using Differential Equation Models with Regime Switching. In *Psychometrika* (Vol. 83, Issue 2). https://doi.org/10.1007/s11336-018-9605-1

Chu, W. S., De la Torre, F., Cohn, J. F., & Messinger, D. S. (2017). A Branch-and-Bound Framework for Unsupervised Common Event Discovery. *International Journal of Computer Vision*, *123*(3), 372–391. https://doi.org/10.1007/s11263-017-0989-7

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2021). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behavior Research Methods*, *53*(2), 467–486. https://doi.org/10.3758/s13428-020-01393-5

Custode, S., Bailey, J., Sun, L., Katz, L. F., Messinger, D. S., & Perry, L. K. (2020). Peer Interactions as Opportunities for Language in an Early Intervention Preschool Classroom. *Poster Presented at the 22nd Biennial International Congress of Infant Studies.*

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

Dykstra, J. R., Sabatos-Devito, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L.

(2013). Using the Language Environment Analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, *17*(5), 582–594. https://doi.org/10.1177/1362361312446206

Ertugrul, I. O., Ahn, Y. A., Bilalpur, M., Messinger, D. S., Speltz, M. L., & Cohn, J. F. (2021). *Infant AFAR: Automated Facial Action Recognition in Infants [Manuscript submitted for publication]*.

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

Fasano, R. M., Perry, L. K., Custode, S. A., Adams, A., Bulotsky Shearer, R. J., & Messinger, D. S. (2021). Associations between Social Language Use and Behavioral and Social Engagement in Inclusive ASD Classrooms. *Poster Presented at the 39th Biennial Society for Research in Child Development, Virtual Meeting*.

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

Fogel, A., & Thelen, E. (1987). Development of Early Expressive and Communicative Action: Reinterpreting the Evidence From a Dynamic Systems Perspective. *Developmental Psychology*, *23*(6), 747–761. https://doi.org/10.1037/0012-1649.23.6.747

Fraley, R. C., & Spieker, S. J. (2003). Are infant attachment patterns continuously or categorically distributed? A taxometric analysis of strange situation behavior.

Developmental Psychology, *39*(3), 387–404.

Gilkerson, J., & Richards, J. A. (2009). The Power of Talk (2nd Edition). *LENA Technical Report (LTR-01-2)*, 1–30.

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language Experience in the Second Year of Life Predicts Language Outcomes in Late Childhood. *Pediatrics*, *142*(4). https://doi.org/10.1542/peds.2017-4276.

Goldstein, M. H., Shwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing: Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child Development*, *80*(3), 636–644. https://doi.org/10.1111/j.1467-8624.2009.01287.x.The

Gonzalez Villasanti, H., Justice, L. M., Chaparro-Moreno, L. J., Lin, T. J., & Purtell, K. (2020). Automatized analysis of children's exposure to child-directed speech in reschool settings: Validation and application. *PLoS ONE*, *15*(11 November), 1–18. https://doi.org/10.1371/journal.pone.0242511

Harbison, A. L., Woynaroski, T. G., Tapp, J., Wade, J. W., Warlaumont, A. S., & Yoder, P. J. (2018). A New Measure of Child Vocal Reciprocity in Children with Autism Spectrum Disorder. *Autism Research*, *11*(6), 903–915. https://doi.org/10.1002/aur.1942.

Hilton, C. L., Zhang, Y., Whilte, M. R., Klohr, C. L., & Constantino, J. (2012). Motor impairment in sibling pairs concordant and discordant for autism spectrum disorders. In *Autism* (Vol. 16, Issue 4, pp. 430–441). https://doi.org/10.1177/1362361311423018

Hindman, A. H., Wasik, B. A., & Bradley, D. E. (2019). How Classroom Conversations Unfold: Exploring Teacher–Child Exchanges During Shared Book Reading. *Early Education and Development*, *30*(4), 478–495. https://doi.org/10.1080/10409289.2018.1556009

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust,

P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science*, *26*(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Hoffman, M. F., Quittner, A. L., & Cejas, I. (2014). Comparisons of social competence in young children with and without hearing loss: A dynamic systems framework. *Journal of Deaf Studies and Deaf Education*, *20*(2), 115–124. https://doi.org/10.1093/deafed/enu040

Ingvalson, E. M., Grieco-Calub, T. M., Perry, L. K., & VanDam, M. (2020). Rethinking Emergent Literacy in Children With Hearing Loss. *Frontiers in Psychology*, *11*(January). https://doi.org/10.3389/fpsyg.2020.00039

Irvin, D. W., Crutchfield, S. A., Greenwood, C. R., Kearns, W. D., & Buzhardt, J. (2018). An automated approach to measuring child movement and location in the early childhood classroom. In *Behavior Research Methods* (Vol. 50, Issue 3, pp. 890–901). https://doi.org/10.3758/s13428-017-0912-8

Irvin, D. W., Luo, Y., Huffman, J. M., Grasley-Boy, N., Rous, B., & Hansen, J. H. L. (2021). Capturing talk and proximity in the classroom: Advances in measuring features of young children's friendships. *Early Childhood Research Quarterly*, *57*, 102–109. https://doi.org/10.1016/j.ecresq.2021.05.003

Irwin, V., Zhang, J., Wang, X., Hein, S., Wang, K., Roberts, A., York, C., Barmer, A., Bullock Mann, F., Dilig, R., & Parker, S. (2021). *Report on the Condition of Education 2021*. https://nces.ed.gov/pubsearch/pubsinfo.%0Aasp?pubid=2021144.

Johnson, C., & Goswami, U. (2010). Phonological Awareness, Voabulary, and Reading in Deaf Children. *Journal of Speech Language and Hearing Research*, *53*(April), 237–262. https://doi.org/10.1044/1092-4388(2009/08-0139)

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer Effects in Preschool Classrooms: Is Children's Language Growth Associated With Their Classmates' Skills? In *Child Development* (Vol. 82, Issue 6, pp. 1768–1777). https://doi.org/10.1111/j.1467-8624.2011.01665.x

Karmiloff-Smith, A. (2018). Development itself is the key to understanding developmental disorders. *Thinking Developmentally from Constructivism to Neuroconstructivism: Selected Works of Annette Karmiloff-Smith*, *2*(10), 97–117. https://doi.org/10.4324/9781315516691-5

Lahiri, R., Kumar, M., Bishop, S., & Narayanan, S. (2020). Learning Domain Invariant Representations for Child-Adult Classification from Speech. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2020-May*, 6749–6753. https://doi.org/10.1109/ICASSP40776.2020.9054276

Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., Warmuth, M., & Wolf, P. (2003). *The CMU SPHINX-4 speech recognition system. 1*, 2–5.

Laursen, B. (2017, 2017/12/01). Making and Keeping Friends: The Importance of Being Similar. Child Development Perspectives, 11(4), 282-289. https://doi.org/10.1111/cdep.12246

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Locke, J., Kasari, C., Rotheram-Fuller, E., Kretzmann, M., & Jacobs, J. (2013). Social Network Changes Over the School Year Among Elementary School-Aged Children with and Without an Autism Spectrum Disorder. *School Mental Health*, *5*(1), 38–47. https://doi.org/10.1007/s12310-012-9092-y

Lund, E., Werfel, K. L., & Schuele, C. M. (2015). Phonological awareness and vocabulary

performance of monolingual and bilingual preschool children with hearing loss. *Child Language Teaching and Therapy*, *31*(1), 85–100. https://doi.org/10.1177/026565904531261

Martin, C. L., Fabes, R. A., Hanish, L. D., & Hollenstein, T. (2005). Social dynamics in the preschool. Developmental Review, 25(3-4), 299-327. https://doi.org/10.1016/j.dr.2005.10.001

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. E. (2016). AFFDEX SDK: A cross- platform real-time multi-face expression recognition toolkit. *2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 3723–3726. https://doi.org/https://doi.org/10.1145/2851581.2890247

McDuffie, A., & Yoder, P. (2010). Types of Parent Verbal Responsiveness That Predict Language in Young Children With Autism Spectrum Disorder. *Journal of Speech, Language, and Hearing Research*, *53*(4), 1026–1039. https://doi.org/10.1044/1092-4388(2009/09-0023)

Messinger, D. S. (2019). Data Drive Development - Objective Phenotyping of Autism Spectrum Disorder. *Keynote Presented at the Translational Multimedia Data Mining (TMDM) Workshop at the IEEE International Conference on Data Mining*.

Messinger, D. S., Custode, S., A., Perry, L. K., Mitsven, S. G., Ullery, M., Katz, L. F., Vitale, L., Valtierra, A. M., Prince, E. B., & Johnson, N. (March, 2019). Classroom cries exhibit more temporal structure than non-cries at one year. Talk presented at the Society for Research in Child Development meeting, Baltimore, MD.

Messinger, D. S., Mahoor, M. H., Chow, S. M., & Cohn, J. F. (2009). Automated measurement of facial expression in infant - Mother interaction: A pilot study. *Infancy*, *14*(3), 285–305. https://doi.org/10.1080/15250000902839963

Messinger, D. S., Prince, E. B., Zheng, M., Martin, K., Mitsven, S. G., Huang, S., Stölzel, T., Johnson, N., Rudolph, U., Perry, L. K., Laursen, B., & Song, C. (2019). Continuous measurement of dynamic classroom social interactions. *International Journal of Behavioral Development*, *43*(3), 263–270. https://doi.org/10.1177/0165025418820708

Messinger, D. S., Ruvolo, P., Ekas, N. V., & Fogel, A. (2010). Applying machine learning to infant interaction: The development is in the details. *Neural Networks*, *23*(8–9), 1004–1016. https://doi.org/10.1016/j.neunet.2010.08.008

Mitsven, S. G., Perry, L. K., Tao, Y., Elbaum, B. E., Johnson, N. F., & Messinger, D. S. (2021). Objectively measured teacher and preschooler vocalizations: Phonemic diversity is associated with language abilities. *Developmental Science*, *July 2020*, 1–16. https://doi.org/10.1111/desc.13177

Moeller, M. P., Tomblin, J. B., Yoshinaga-Itano, C., Connor, C. M. D., & Jerger, S. (2007). Current state of knowledge: Language and literacy of children with hearing impairment. *Ear and Hearing*, *28*(6), 740–753. https://doi.org/10.1097/AUD.0b013e318157f07f

Moffitt, J. M., Ahn, Y. A., Custode, S., Tao, Y., Mathew, E., Parlade, M., Hale, M., Durocher, J., Alessandri, M., Perry, L. K., & Messinger, D. S. (2021). *Objective measurement of vocalizations in the assessment of autism spectrum disorder symptoms in preschool age children [Manuscript submitted for publication]*.

Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N. Y., Quittner, A. L., & Fink, N. E. (2010). Spoken language development in children following cochlear implantation. *JAMA - Journal of the American Medical Association*, *303*(15), 1498–1506. https://doi.org/10.1001/jama.2010.451

Nittrouer, S., Caldwell, A., & Holloman, C. (2012). Measuring what matters: Effectively

predicting language and literacy in children with cochlear implants. In *International Journal of Pediatric Otorhinolaryngology* (Vol. 76, Issue 8, pp. 1148–1158). https://doi.org/10.1016/j.ijporl.2012.04.024

Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., & Mordvinstev, A. (2018). *The Building Blocks of Interpretabiligy*. Distill. https://doi.org/10.23915/distill.00010

Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the Mouths of Babes: Vocal Production in Infant Siblings of Children with ASD. *Journal of Child Psychology and Psychiatry*, *53*(5), 588–598. https://doi.org/10.1111/j.1469-7610.2010.02332.x.Out

Perone, S., & Simmering, V. R. (2017). Chapter Two - Applications of Dynamic Systems Theory of Cognition and Development: New Frontiers. *Advances in Child Development Behavior*, *52*(43–80). https://doi.org/https://doi.org/10.1016/bs.acdb.2016.10.002

Perry, L. K., Mitsven, S. G., Custode, S., Vitale, L., Laursen, B., Song, C., & Messinger, D. S. (2021). *Reciprocal patterns of peer speech in preschoolers with and without hearing loss [Manuscript submitted for publication]*.

Perry, L. K., Prince, E. B., Valtierra, A. M., Rivero-Fernandez, C., Ullery, M. A., Katz, L. F., Laursen, B., & Messinger, D. S. (2018). *A year in words: The dynamics and consequences of language experiences in an intervention classroom*. https://doi.org/10.1371/journal.pone.0199893

Phebey, T. (2010). The Ubisense assembly control solution for BMW. *Proceedings of RFID Journal Europe Live*.

Prince, E. B., Ciptadi, A., Tao, Y., Rozga, A., Martin, K. B., Rehg, J., & Messinger, D. S. (2021). Continuous measurement of attachment behavior: A multimodal view of the strange

situation procedure. Infant Behavior and Development, 63, 101565. https://doi.org/https://doi.org/10.1016/j.infbeh.2021.101565

Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., & Casillas, M. (2021). ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, *53*(2), 818–835. https://doi.org/10.3758/s13428-020-01460-x

Rehg, J. M. (2011). Behavior imaging : Using computer vision to study autism. *MVA2011 IAPR Conference on Machine Vision Applications*. https://doi.org/10.1145/2522848.2532203

Rehg, J. M., Abowd, G. D., Rozga, A., Romero, M., Clements, M. A., Sclaroff, S., Essa, I., Ousley, O. Y., Li, Y., Kim, C., Rao, H., Kim, J. C., Presti, L. Lo, Zhang, J., Lantsman, D., Bidwell, J., & Ye, Z. (2013). Decoding children's social behavior. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3414–3421. https://doi.org/10.1109/CVPR.2013.438

Rehg, J. M., Rozga, A., Abowd, G. D., & Goodwin, M. S. (2014). Behavioral Imaging and Autism. *IEEE Pervasive Computing*, *13*(2), 84–87. https://doi.org/10.1109/MPRV.2014.23

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language Use Contributes to Expressive Language Growth: Evidence from Bilingual Children. *Child Development*, *89*(3), 929–940. https://doi.org/10.1111/cdev.12770

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710. https://doi.org/https://doi.org/10.1177/0956797617742725

Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep Expectation of Real and Apparent Age

from a Single Image Without Facial Landmarks. In *International Journal of Computer Vision* (Vol. 126, Issues 2–4, pp. 144–157). https://doi.org/10.1007/s11263-016-0940-3

Ruiz, A., Jimenez, R., & Granja, F. S. (2017). Comparing ubisense, bespoon, and deawave uwb location systems: Indoor performance analysis. *IEEE Transactions on Instrumentation and Measurement*, *66*(8), 2106–2117.

Ruvolo, P., Messinger, D., & Movellan, J. (2015). Infants time their smiles to make their moms smile. *PLoS ONE*, *10*(9), 1–10. https://doi.org/10.1371/journal.pone.0136492

Schoemaker, M. M., & Kalverboer, A. F. (1994). Social and affective problems of children who are clumsy: How early do they begin? *Adapted Physical Activity Quarterly*, *11*(2), 130–140. https://doi.org/10.1123/apaq.11.2.130

Sen, A., Cox, C. R., Borkenhagen, M. C., Seidenberg, M. S., & Zhu, X. (2020). *Learning to Read through Machine Teaching*. 1–11. http://arxiv.org/abs/2006.16470

Shin, N., Kim, M., Goetz, S., & Vaughn, B. E. (2014). Dyadic Analyses of Preschool-aged Children's Friendships: Convergence and Differences between Friendship Classifications from Peer Sociometric Data and Teacher's Reports. *Social Development*, *23*(1), 178–195. https://doi.org/10.1111/sode.12043

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. https://doi.org/10.1016/S1364-6613(03)00156-6

Soderstrom, M., & Wittebolle, K. (2013). When Do Caregivers Talk? The Influences of Activity and Time of Day in Caregiver Speech and Child Vocalizations in Two Childcare Environments. *PLoS ONE*, *8*(11), e80646. https://doi.org/10.1371/journal.pone.0080646

Spencer, J. P., Perone, S., & Buss, A. T. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. In *Child Development Perspectives*

(Vol. 5, Issue 4, pp. 260–266). https://doi.org/10.1111/j.1750-8606.2011.00194.x

Thibodeau, P. H., Blonder, A., & Flusberg, S. J. (2020). A connectionist account of the relational shift and context sensitivity in the development of generalisation. *Connection Science*, *32*(4), 384–397. https://doi.org/https://doi.org/10.1080/09540091.2020.1728519

Veiga, J. J. D., & Ward, T. E. (2016). Data collection requirements for mobile connected health: An end user development approach. *Mobile! 2016 - Proceedings of the 1st International Workshop on Mobile Development, Co-Located with SPLASH 2016*, 23–30. https://doi.org/10.1145/3001854.3001856

Walker, E. A., & McGregor, K. K. (2013). Word learning processes in children with cochlear implants. *Journal of Speech Language and Hearing Research*, *56*(2), 375–387. https://doi.org/10.1044/1092-4388(2012/11-0343)

Walker, W., Lamere, P., Kwok, P., Bhiksha, R., Singh, R., Gouvea, E., Wolf, P., & Woelfel, J. (2004). *Sphinx-4: A Flexible Open Source Framework for Speech Recognition*.

Wang, Y., Williams, R., Dilley, L., & Houston, D. M. (2020). A meta-analysis of the predictability of LENA^{TM} automated measures for child language development. *Developmental Review*, *57*(May), 100921. https://doi.org/10.1016/j.dr.2020.100921

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A Social Feedback Loop for Speech Development and Its Reduction in Autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

Waters, E. (2002). Comments on Strange Situation Classification. In. Retrieved 1/5/12 from http://www.johnbowlby.com.

Woynaroski, T., Oller, D. K., Keceli-Kaysili, B., Xu, D., Richards, J. A., Gilkerson, J., Gray, S., & Yoder, P. (2017). The stability and validity of automated vocal analysis in preverbal

preschoolers with autism spectrum disorder. *Autism Research*, *10*(3), 508–519. https://doi.org/10.1002/aur.1667

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

Xu, D., Richards, J. A., Gilkerson, J., Yapanel, U., Gray, S., & Hansen, J. (2009). *Automatic Childhood Autism Detection by Vocalization Decomposition with Phone-like Units General Terms*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.8098&rep=rep1&type=pdf

Yale, M. E., Messinger, D. S., Cobo-Lewis, A. B., & Delgado, C. F. (2003). The Temporal Coordination of Early Infant Communication. *Developmental Psychology*, *39*(5), 815–824. https://doi.org/10.1037/0012-1649.39.5.815

You, Y., Sun, J., Guo, Y., Tan, Y., & Jiang, J. (2021). Interpretability adn accuracy trade-off in the modeling of belief rule-based systems. *Knowledge-Based Systems*, *10479*.

Yu, C., & Smith, L. B. (2016). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, K., Tao, Z., & Fu, Y. (2017). Deep sequential context networks for action prediction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1473–1481.

Yu, Y., Wang, W., & Han, P. (2016). Localization based stereo speech source separation using probabilistic time-frequency masking and deep neural networks. *EURASIP Journal on Audio, Speech, and Music Processing*, *1*, 1–18.

Yurkovic, J. R., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., Kennedy, D. P., & Yu, C. (2021). Using head-mounted eye tracking to examine visual and manual exploration during naturalistic toy play in children with and without autism spectrum disorder. *Scientific Reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-021-81102-0

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, *23*(10), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342

Figure Legends

Figure 1. ADOS-2 Assessment. A child's face and expression are captured by examiner-worn video-enabled eyeglasses and detected by a machine learning program during administration of the ADOS-2.

Figure 2. Phonemic Diversity and Assessed Language Ability. (A) Mediation model. The phonemic diversity of teachers' speech was associated with the phonemic diversity of children's speech, which, in turn, positively predicted children's end-of-year expressive language abilities. (B) Scatterplot. Children who produced speech with a higher number of unique phonemes scored higher on a standardized measure of expressive language ability, the PLS-5. (Parallel results with receptive language ability not shown.) Credit: Mitsven, S. G., Perry, L. K., Tao, Y., Elbaum, B. E., Johnson, N. F., & Messinger, D. S. (2021). Objectively measured teacher and preschooler vocalizations: Phonemic diversity is associated with language abilities. *Developmental science*.

Figure 3. Materials used to continuously track children's and teachers' location and orientation in the classroom. Ubisense sensors are arrayed, one in each corner of the classroom, and track the ultrawide RFID signal emitted by each of the two tags worn by children and by teachers.

Figure 4. The radial distribution function. The radial distribution function, g(r), indicates distances at which the probability of child–child and child-teacher pairs being in contact exceeds chance (g(r) = 1). The grey area above between 0.2 and 2 m index illustrates the proximity criterion for social contact across cohorts (classes of children). Figure compliments of Chitra Banarjee, Yale University.

Figure 5. Semi-log plots of the mean frequency of children's social contacts by ordinal rank. Levels of social contact, aggregated for all children, are displayed for each child's most to least contacted peer. Rank 1 indicates the most contacted peer, Rank 2 the second-most, and so on. Boys and girls' mean contact frequency are shown separately. Both males and females consistently show levels of social contact with their most contacted peers that are ten to hundreds of time higher than levels of contact with their least contacted peers. Credit: Messinger, D. S., Prince, E. B., Zheng, M., Martin, K., Mitsven, S. G., Huang, S., ... & Song, C. (2019). Continuous measurement of dynamic classroom social interactions. *International Journal of Behavioral Development*, *43*(3), 263-270.

Figure 6. Classroom visualization. A visualization of objective measurements of child and teacher location in an inclusion classroom. The tip of each triangle indicates the direction each child and teacher is facing (orientation). Orientation is calculated from the 2 Ubisense tags

Figure 7. Mutual orientation. The heat map indicates the angle of each child to other children & teachers (θ_1 to θ_2) from .2–2m. Children tend to be shoulder to shoulder (~=90 ^{o}), & face-to-face (within |45^{o}| of 0^{o}) with partners. The color bar indicates the ratio of observed mutual orientations to chance (=1, yellow line; black line =1.5). Chance relative orientation refers to the cumulative probability of relative orientations between all children over an entire observation.

Figure 8. Peer speech transmission. Visualized associations between peer vocalizations in social contact with focal child at observation t and child vocalizations to those peers at t+1. The

association did not differ by the focal child's hearing status. The vocalization rate per hour is presented on a log10(x+1) scale. Each point represents one child's vocalizations to one peer. Error bands show standard errors of the mean.

Figure 9. Classroom speech network. Each node is one child in a classroom. Node size is proportional to total vocalization rate (input and output) in social contact with all peers. Tie width indexes the sum of vocalizations in social contact between two children (co-talk). For visualization purposes, the network was pruned such that only ties of greater than mean weight (both per observation and overall) are displayed. Credit: Fasano, R. M., Perry, L. K., Zhang, Y., Vitale, L., Wang, J., Song, C., & Messinger, D. S. (2021). A granular perspective on inclusion: Objectively measured interactions of preschoolers with and without autism. Autism Research.

Figure 10. Mutual orientation and proximity. Mutual orientation is indexed by the relative orientation of two children (theta1 minus theta2) where 0 indicates a face-to-face orientation and ±180 indicates shoulder-to-shoulder orientation. *R* indicates the radius, the distance between the two children. The heat map indicates the likelihood of a given orientation and a given distance occurring with respect to chance (observed divided by chance). Reds indicate higher likelihoods probabilities and blues indicate lower likelihoods. The green line encloses all areas where the joint probability exceeds 1, and the black line encloses areas where the joint probability exceeds 2. Face-to-face contact tends to occur over relatively long distances (through 3 meters) but is concentrated in a small band around 1 meter. By contrast shoulder-to-shoulder contact (around plus and minus 180 degrees) tends to occur much more than chance in a relatively restricted

range of distances between approximately .2 and 1.8 meters. Figure compliments of Chaoming Song and Yi Zhang, University of Miami.

Figure:1A Figure:1B

Figure 1. ADOS-2 Assessment.

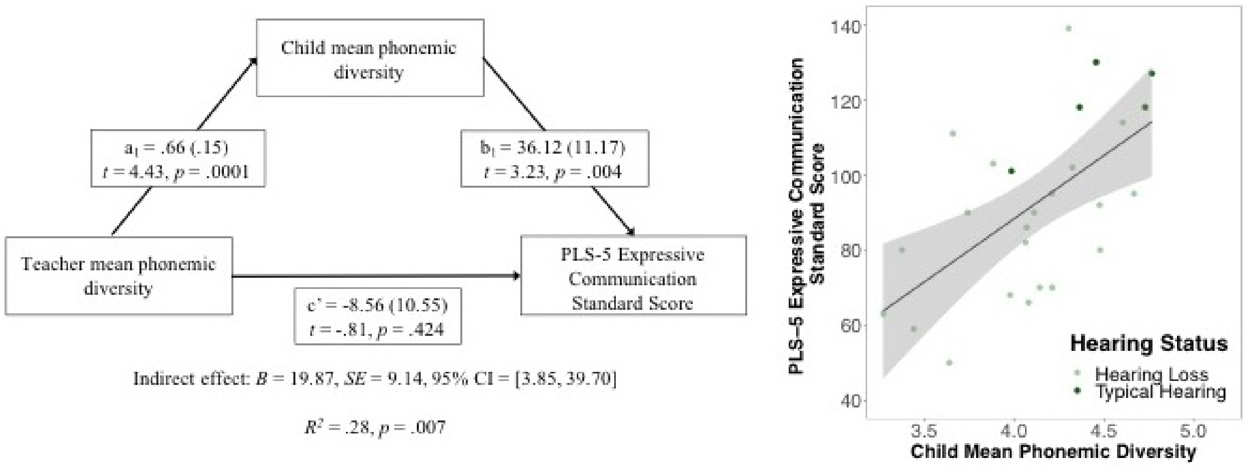

Figure 2. Phonemic Diversity and Assessed Language Ability.

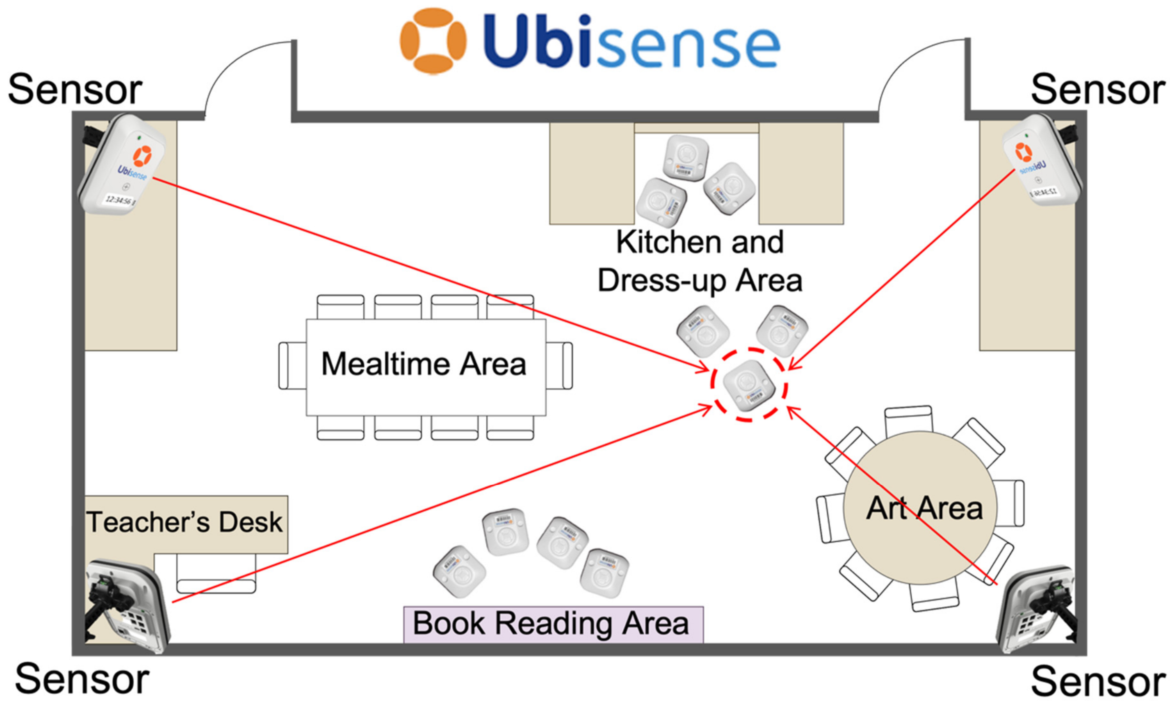

Figure 3. Materials used to continuously track children's and teachers' location and orientation in the classroom.

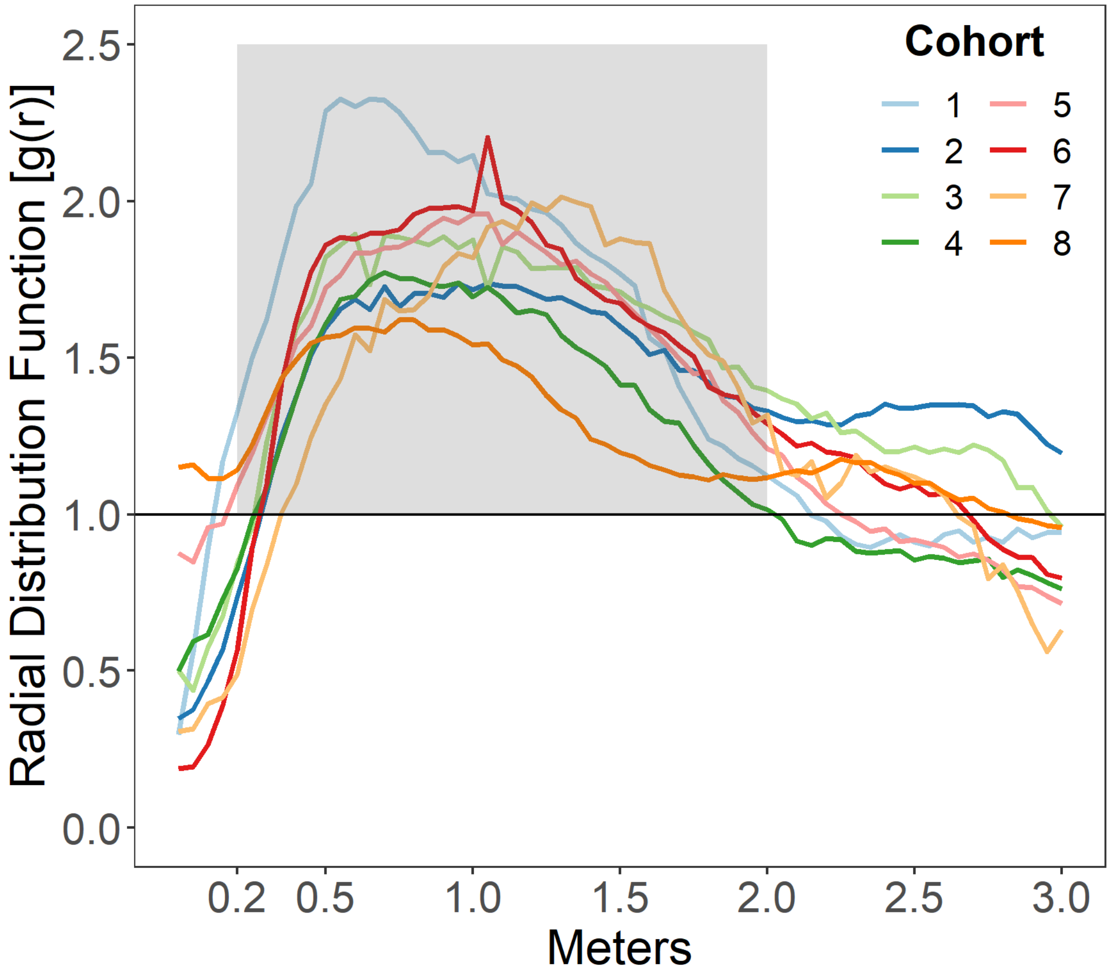

Figure 4. The radial distribution function.

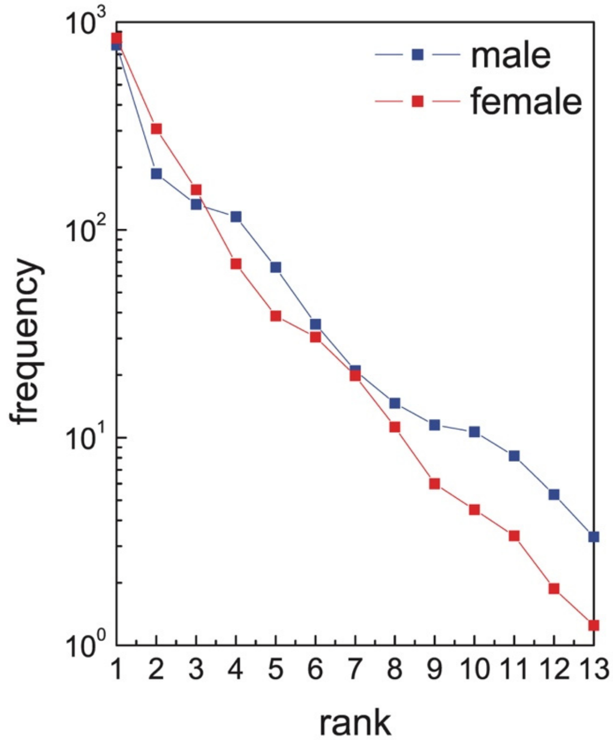

Figure 5. Classroom social network illustration.

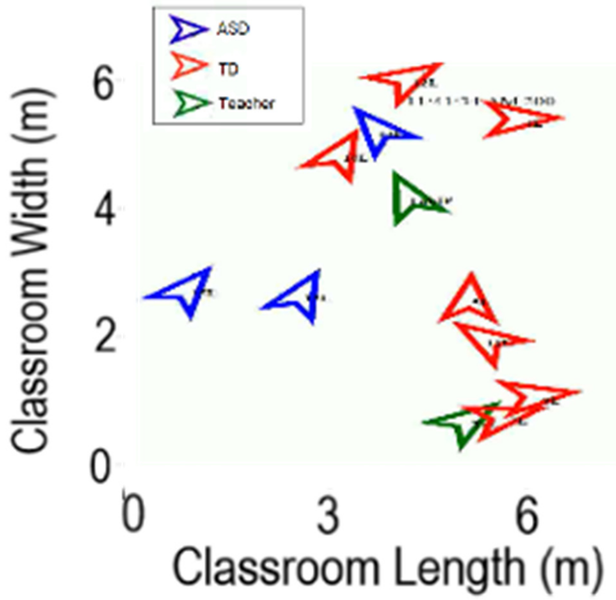

Figure 6. Classroom visualization.

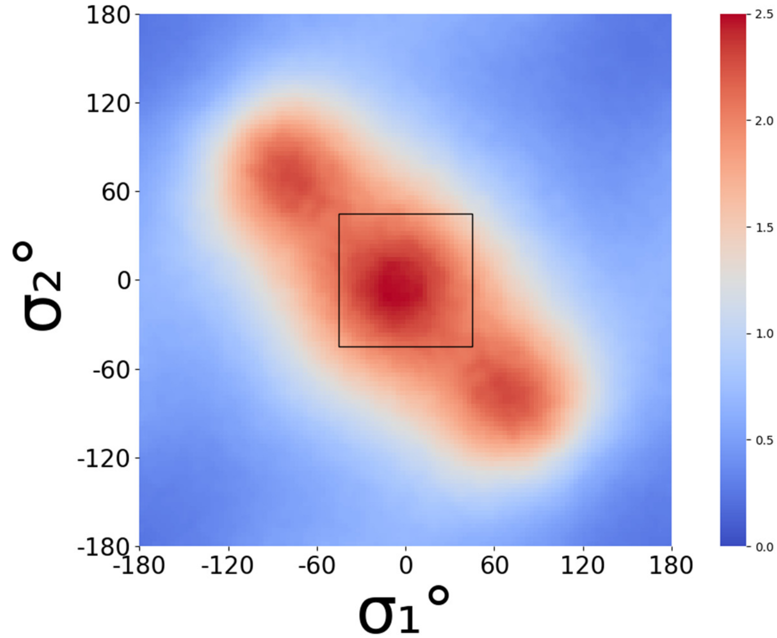

Figure 7. Mutual Orientation.

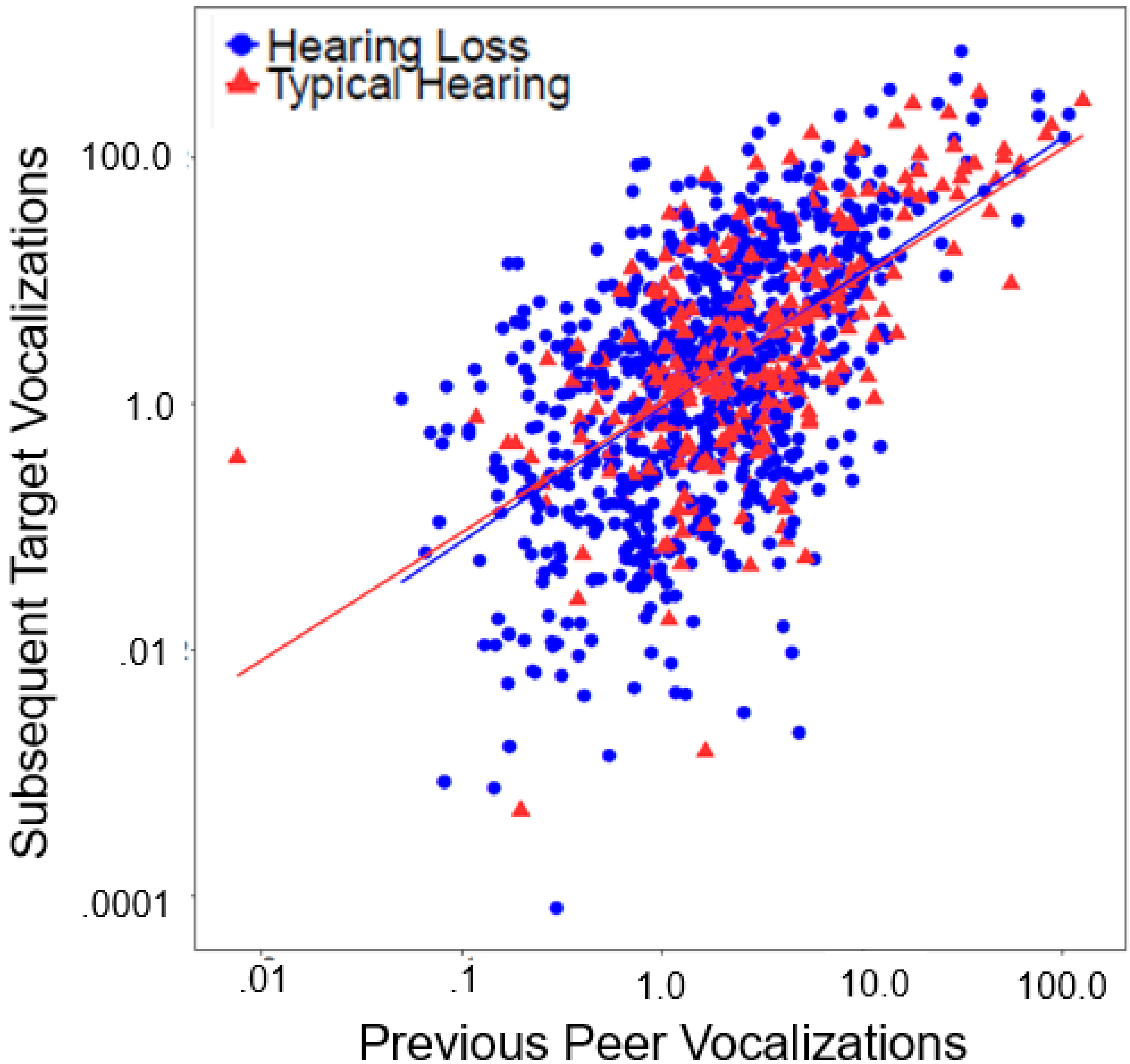

Figure 8. Peer speech transmission

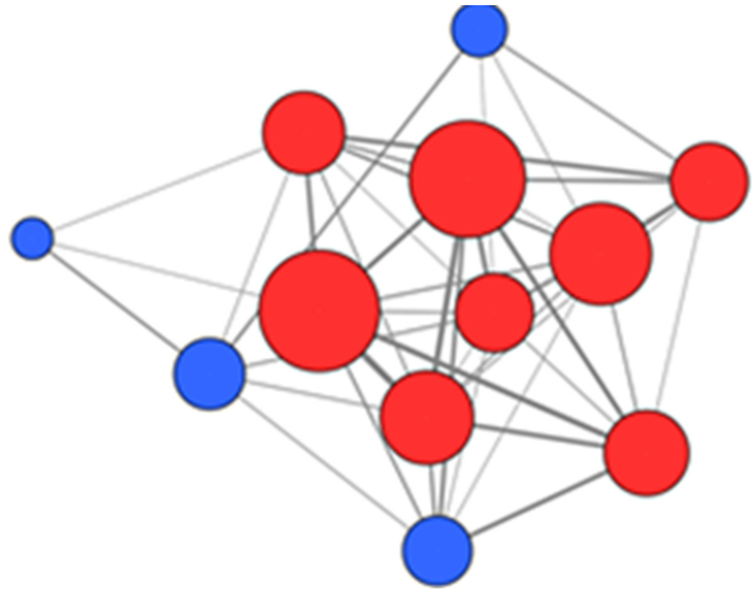

Figure 9: Classroom Speech Network.

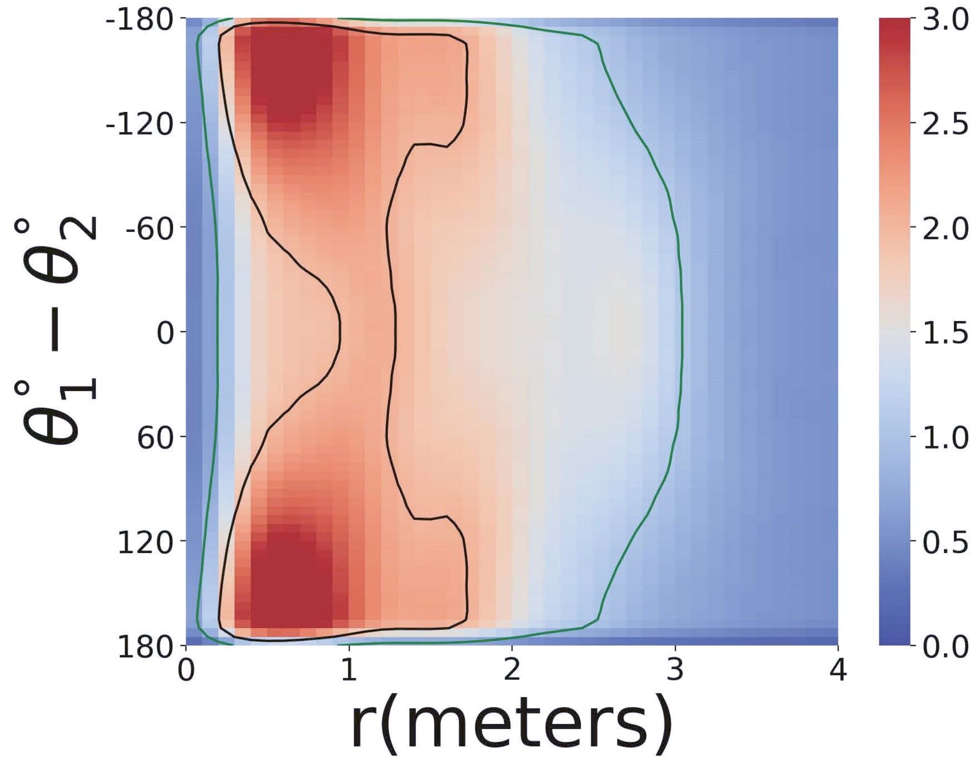

Figure10. Mutual orientation and distance.