scientific reports

Check for updates

OPEN

Intra-topic latency as an automated behavioral marker of treatment response in autism spectrum disorder

Elizabeth P. McKernan[1], Manoj Kumar[2], Adriana Di Martino[3], Lisa Shulman[4], Alexander Kolevzon[5], Catherine Lord[6], Shrikanth Narayanan[2] & So Hyun Kim[1]✉

Data science advances in behavioral signal processing and machine learning hold the promise to automatically quantify clinically meaningful behaviors that can be applied to a large amount of data. The objective of this study was to identify an automated behavioral marker of treatment response in social communication in children with autism spectrum disorder (ASD). First, using an automated computational method, we successfully derived the amount of time it took for a child with ASD and an adult social partner (*N* pairs = 210) to respond to each other while they were engaged in conversation bits ("latency") using recordings of brief, natural social interactions. Then, we measured changes in latency at pre- and post-interventions. Children with ASD who were receiving interventions showed significantly larger reduction in latency compared to those who were not receiving interventions. There was also a significant group difference in the changes in latency for adult social partners. Results suggest that the automated measure of latency derived from natural social interactions is a scalable and objective method to quantify treatment response in children with ASD.

Automatically-derived, objective audio–video measures of behavior can complement and enhance traditional manual behavioral coding, providing an alternative way to identify treatment response[1,2]. Signal processing and machine learning techniques have been applied to study social interactions of children with autism spectrum disorder (ASD). These efforts have yielded a number of findings relevant to treatment outcome, including the identification of acoustic-prosodic properties that can be measured objectively and are related to autism symptom severity[3,4].

The goal of the present study was to assess response to treatment as measured by automatically extracted conversational features from a brief, natural social interaction (i.e., Brief Observation of Social Communication Change [BOSCC][5]). To date, outcome measures for ASD have largely consisted of study-specific, manually coded behavioral observations or parent-, therapist- or self-reports[6]; many of these measures do not directly target core autism symptoms such as social communication impairments even though understanding whether core autism symptoms are effectively impacted by treatments is the key to evaluate their efficacy[7–9]. It has also been shown that relying solely on caregiver-, therapist-, and/or self-reports may introduce measurement error[10,11]. This is due to a variety of factors including bias and unblinding. Thus, unbiased automated coding can complement traditional behavioral measures by humans. The detection of treatment response using an automated method also has high scalability to process a large amount of data, allowing us to overcome the constraints in time and efforts.

Previous outcome measures for interventions focused on improving communicative or conversational skills have included task analyses[12], Likert scales[13], and probe data[14], none of which incorporated computational methods of data interpretation using automatic signal and information processing techniques. The automatically derived conversation behaviors have the potential to transform the field by providing an objective and scalable measure of treatment response that augments human judgment and enhances traditional manual coding. Previous work examined automated detection, modeling, and analyses of behaviors of children with ASD interacting

[1]Department of Psychiatry, Weill Cornell Medicine, New York, NY, USA. [2]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA. [3]Child Mind Institute, New York, NY, USA. [4]Department of Pediatrics, Albert Einstein College of Medicine, New York, NY, USA. [5]Seaver Autism Center for Research and Treatment, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ✉email: sok2015@med.cornell.edu

	1	2	3	4	5	6	7	8
1. VIQ	1							
2. NVIQ	0.84**	1						
3. Expressive language level	− 0.09	− 0.09	1					
4. CSS SA	− 0.07	− 0.01	0.05	1				
5. CSS RRB	− 0.03	0.02	− 0.01	0.25**	1			
6. Age	− 0.02	− 0.01	0.36**	0.12*	− 0.01	1		
7. Child latency intra-topics	− 0.03	− 0.04	− 0.05	0.24**	− 0.04	− 0.02	1	
8. Examiner latency intra-topics	− 0.01	− 0.03	− 0.02	0.08	0.04	0.04	0.36**	1

Table 1. Bivariate correlations between clinical measures and intra-topic latency. N = 333–350. *$p < 0.05$, **$p < 0.01$.

with a social partner in the context of diagnostic evaluations. Results suggested that children with greater autism symptom severity spoke less, took longer to respond (longer latency), as well as used personal pronouns and affect language less often[4]. Expanding on these findings, for the first time in the field, we used automatically extracted conversational features, specifically latency, or the time it takes for social partners to respond to each other, to measure treatment response in children with ASD. Automated coding of latency provides a highly specific temporal resolution which augments human coding, given that these features are impossible for human coders to measure with high precision.

Latency. Previous work focused on latency demonstrated that children with ASD often take longer to respond within a conversation. For instance, children with ASD were found to take 23.7% longer to respond than typically developing (TD) children in back-and-forth conversation, and 34.4% longer to respond to questions than TD children[15]. Another study[16] showed that response latency positively correlated with the overall Autism Diagnostic Observation Schedule (ADOS[17]) severity score and the social communication ADOS severity score, pointing to the utility of using "intra-topic" latency (the time it takes to respond to a social partner within a conversation topic). Therefore, in our study, we hypothesized that latency would serve as an effective treatment outcome measure for children with ASD. In addition, recent work suggested that adult social partners while interacting with a child with ASD adjust their own behaviors; thus, their speech properties are more predictive of the child's social communication impairments than the child's own atypical speech patterns[18]. Based on this, we hypothesized that dynamic interactions between dyads would lead to changes in latency in both children with ASD and their social partners over the course of treatment.

Results

Association between latency and baseline clinical features. Bivariate correlations were conducted to determine whether there was a significant association between latency and clinical measures of autism severity (ADOS-2 Calibrated Severity Score–Social Affect [CSS SA]) and cognitive functioning (nonverbal IQ [NVIQ], verbal IQ [VIQ]). Child and examiner latency variables were significantly positively correlated with each other. Child intra-topic latency was significantly positively correlated with the ADOS-2 CSS SA. On the other hand, neither of the latency variables were significantly correlated with VIQ nor NVIQ. Bivariate correlations are presented in Table 1.

Changes in latency over the course of interventions. There was a significant interaction between time and treatment condition for both child intra-topic latency as well as examiner intra-topic latency (see Table 2). As depicted in Fig. 1a, the intra-topic latency of children who received treatment significantly decreased from pre- to post-assessment ($t[172] = 2.05$, $p = 0.042$, $d = 0.16$; $M = 2.94$, $SD = 1.77$ at Pre-intervention, $M = 2.69$, $SD = 1.33$ at Post-intervention), whereas intra-topic latency significantly increased over time for children in the "treatment as usual" (TAU) condition ($t([175] = -3.46$, $p = 0.001$, $d = -0.28$; $M = 2.63$, $SD = 1.21$ at Pre-intervention, $M = 3.05$, $SD = 1.79$ at Post-intervention). As shown in Fig. 1b, examiners' intra-topic latency was stable from pre- to post-assessment when interacting with children who received treatments ($t[172] = 1.82$, $p = 0.071$, $d = 0.07$; $M = 2.54$, $SD = 2.49$ at Pre-intervention, $M = 2.38$, $SD = 2.38$ at Post-intervention) but significantly increased over time when interacting with children in the TAU group ($t[175] = -2.07$, $p = 0.04$, $d = -0.17$; $M = 2.29$, $SD = 1.36$ at Pre-intervention, $M = 2.55$, $SD = 1.64$ at Post-intervention). Across groups, the effects of baseline expressive language level and autism symptom severity on latency were also significant such that children with higher expressive language levels and those with lower symptom severity had shorter intra-topic latency.

When treatment duration, site, and the interaction between treatment condition and site were controlled within the mixed models, treatment duration ($F[1, 215] = 0.008$, $p = 0.929$ for child latency; $F[1, 215] = 0.118$, $p = 0.732$ for examiner latency) and the interaction between treatment condition and site ($F[1, 215] = 0.23$, $p = 0.631$ for child latency; $F[1, 215] = 0.431$, $p = 0.512$ for examiner latency) were not significant, but there was a significant effect of site for child latency ($F[3, 215] = 2.96$, $p = 0.033$ for child latency; $F[3, 215] = 1.39$, $p = 0.245$ for examiner latency); however, all of the significant results on the interaction effect between the time and

Outcome variable	Predictors	*B*	*df*	*F*	Significance
Child intra-topic latency	Overall model		[17, 313]	2.53	0.001**
	Time	0.37	[1, 313]	0.001	0.969
	Treatment condition	0.52	[1, 313]	0.732	0.393
	Time * treatment condition	− 0.73	[1, 313]	5.32	0.022*
	Expressive language	4.74	[4, 313]	4.93	0.001**
	Level = 5	− 0.24			
	Level = 6	0.08			
	Level = 7	− 0.35			
	NVIQ	< 0.001	[1, 313]	0.210	0.647
	CSS SA	− 1.28	[7, 313]	2.12	0.041*
	CSS SA = 4	− 1.19			
	CSS SA = 5	− 1.17			
	CSS SA = 6	− 0.94			
	CSS SA = 7	− 0.64			
	CSS SA = 8	− 0.52			
	CSS SA = 9	− 0.38			
	Age	− 0.002	[1, 313]	0.703	0.402
	Gender	− 0.07	[1, 313]	0.079	0.779
Examiner intra-topic latency	Overall model		[17, 313]	1.05	0.408
	Time	0.29	[1, 313]	0.087	0.768
	Treatment condition	0.42	[1, 313]	0.080	0.777
	Time * treatment condition	− 0.67	[1, 313]	4.40	0.037*
	Expressive language	2.52	[4, 313]	1.02	0.396
	Level = 5	0.43			
	Level = 6	− 0.15			
	Level = 7	− 0.49			
	NVIQ	− 0.001	[1, 313]	0.320	0.572
	CSS SA	− 1.19	[7, 313]	1.14	0.339
	CSS SA = 4	− 0.71			
	CSS SA = 5	0.51			
	CSS SA = 6	− 1.07			
	CSS SA = 7	− 0.56			
	CSS SA = 8	− 0.80			
	CSS SA = 9	− 0.64			
	Age	< 0.001	[1, 313]	0.002	0.962
	Gender	− 0.28	[1, 313]	0.427	0.514

Table 2. Generalized linear mixed models of time and treatment effects for intra-topic latency. *NVIQ* nonverbal IQ, *CSS SA* calibrated severity score-social affect. *$p < 0.05$, **$p < 0.001$.

treatment condition described above remained the same. Post-hoc pairwise comparisons with Bonferroni correction revealed that the child latency mean for the Montefiore site was significantly lower than that of the ISMMS and CADB sites at pre-intervention, and significantly lower than that of the CADB site at post-intervention (all $p < 0.05$).

We also calculated change scores by subtracting pre-assessment values from post-assessment values of both child and examiner intra-topic latency. Change scores for child intra-topic latency were significantly different between treatment and TAU groups ($t[347] = 4.05$, $p < 0.001$, $d = 0.43$). Similarly, change scores for examiner intra-topic latency were also significantly different between treatment and TAU groups ($t[347] = 2.8$, $p = 0.005$, $d = 0.30$).

Finally, regression analyses predicting Post-Intervention latency while controlling for the Pre-Intervention latency as well as NVIQ, autism symptom severity, age, expressive language level, and gender revealed differences between the treatment groups (child latency, $B = -0.530$, $p = 0.001$ for the treatment condition, $R^2 = 0.26$, $F[7, 325] = 16.49$, $p < 0.001$ for the overall model; examiner latency, $B = -0.405$, $p = 0.008$ for the treatment condition, $R^2 = 0.58$, $F[7, 325] = 63.32$, $p < 0.001$ for the overall model), with the treatment group showing significantly shorter latency compared to the TAU group for both children and examiners.

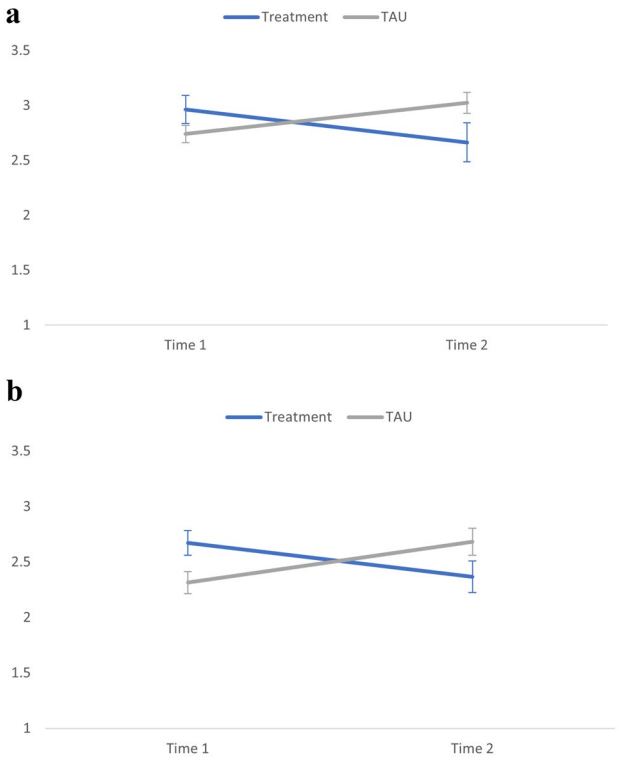

Figure 1. (**a**) Estimated means for child intra-topic latency. Error bars depict ± 2 standard errors. (**b**) Estimated means for examiner intra-topic latency. Error bars depict ± 2 standard errors.

Discussion

The present study used a novel automated measure to identify a behavioral marker of treatment response, specifically intra-topic latency, within a matched sample of children who did and did not receive focused short-term treatments to improve social communication symptoms. The finding of a significant interaction between time and treatment condition for the automated measure of child intra-topic latency reveals that children showed significantly larger reduction in latency over the course of a relatively short-term (i.e., 3 to 4 months' duration) intervention designed to improve social skills. At the same time, there was also a significant interaction between time and treatment condition for examiner intra-topic latency. Specifically, examiners interacting with children who were not receiving interventions showed increased latency over time, whereas those interacting with children who were receiving intervention did not show any difference in latency from pre- to post- intervention. These results imply a reciprocal relation between the behavior of the child and the adult in the context of a brief social interaction.

Previous work by Bone et al.[3,18] using conversational samples taken from the ADOS found that the psychologist's behavior was reflective of the child's social-communicative behavior, in that the psychologist's prosody was more predictive of the child's ASD severity than the child's own prosody. Similarly, in a follow-up study, Bone et al.[4] found in cross-sectional studies that as ASD severity increases, psychologists varied their speech and language strategies in attempts to engage children in social interaction. These results, along with those of the present study on observed changes in latency of both children with ASD and their social partners, fit well within the literature on Communication Accommodation Theory[19,20]. This theory views interpersonal conversation as a dynamic adaptive exchange in which a speaker's speech signal features are tailored to their conversational partner to maximize intelligibility and efficiency. Given the potential for interpersonal synchrony to be a marker of the quality of social interactions[21–23] as well as a potential avenue for intervention[24,25] among ASD populations, intra-topic latency could provide an objective and quantifiable measure of conversational coordination and synchrony.

Additionally, child intra-topic latency was significantly correlated with the ADOS-2 Calibrated Severity Score, also indicating that these features may be useful indicators of clinical measures of overall autism severity. Importantly, neither latency variable was correlated with IQ, suggesting that these features may be separable from cognitive functioning in verbal children with ASD.

There are a few limitations of the study. First, this study was not a randomized controlled trial (RCT). Participants in the treatment and TAU conditions were not evenly represented across sites, with the Center for Autism and the Developing Brain site having a greater number of participants in the TAU condition than would be expected. In the future, we hope to extend our work to validate the automated method in a more controlled setting. We also observed a site effect, although the results remained the same when the effect of site was controlled for. An RCT design across sites will help address the potential confound further. In addition, the lack of extant

literature regarding the distribution of intra-topic latency within non-ASD populations across different ages precludes our ability to determine normative values of this measure. Comparison of this clinical feature across typical and atypical populations will reveal greater insight into what might constitute clinically abnormal latency values. Our work was also limited to children and adolescents who were verbal, and thus the results cannot be generalized to verbal adults or minimally verbal individuals. Finally, it is important to note that 65 out of 140 dyads (46.4%) had different social partners at pre- and post-interventions. Importantly, there was no significant difference between the treatment and TAU groups in the number of participants who had the same examiner at both time points, $\chi^2(1, n = 210) = 0.083$, $p = 0.773$. When the analyses were limited to the dyads who had same examiners at pre- and post-interventions (n = 75 out of 140 dyads; 53.6%), the interaction effect between time and treatment condition remained marginally significant, potentially due to limited statistical power ($F[1, 183] = 2.94$, $p = 0.088$ for child latency; $F[1, 183] = 2.83$, $p = 0.095$ for examiner latency).

Conclusion

The present study was the first to apply signal processing and machine learning techniques to derive an automated measure of treatment response in children with ASD. Specifically, the results demonstrate that automatically derived intra-topic latency can serve as an objective and scalable treatment response measure in youth with ASD who have received social skills training or other forms of interventions. Given the current lack of standard autism-specific treatment response measures that are sensitive enough to capture change[7], our results may push the field to implement automated measures of audio–video signals, which have the potential to augment human perception and judgment[1]. To this end, we leveraged the BOSCC recordings intended to provide the standardized settings of natural social interactions between a child with ASD and a social partner for automated analyses of audio and video data. Our current findings suggest that intra-topic latency in youth with ASD has potential as an indicator of treatment response. Forthcoming research would do well to use latency and other automatically extracted conversational features to examine treatment effects in the context of a broader range of ASD interventions and clinical trials.

Methods

Participants. Data were obtained from a study of the sensitivity and reliability of new outcome measures in capturing treatment responses for children with ASD who were receiving specific interventions at four autism clinics in university-based centers in the New York metropolitan area, compared to a "treatment as usual" (TAU) group who did not receive any interventions at the clinics. Thus, this was not a randomized clinical trial. Participants included 210 children with ASD between the ages of 4 and 17 years from Center for Autism and the Developing Brain (CADB) at Weill Cornell Medical College (n = 62), Child Study Center at the New York University Langone Medical Center (NYU; n = 62), the Seaver Autism Center for Research and Treatment at the Icahn School of Medicine at Mount Sinai (ISMMS; n = 37), and Montefiore Medical Center/Albert Einstein College of Medicine (Montefiore; n = 49). All participants had phrase speech (i.e., flexible 3- to 4-word combinations; n = 22) or verbally fluent language (i.e., use complex sentences and can speak about the past and future; n = 188). Among these children, half of them (n = 105) were participating in social groups or other forms of treatments (see below) at the four university-based centers listed above. Specifically, participants at CADB received individual and group behavior interventions focused on improving social communication in children with ASD, such as the Program for the Education and Enrichment of Relational Skills (PEERS®[26]) and the Secret Agent Society (SAS)[27] as well as a clinical trial testing the use of intranasal oxytocin for the treatment of autism symptoms[28]. Participants at NYU received parent-mediated social skills interventions in a group setting, including PEERS®[26] and the Children's Friendship Program (CFP)[29,30]. Participants at ISMMS took part in a clinical trial testing the use of intranasal oxytocin for the treatment of autism symptoms, as well as family peer advocate services (FPA)[31,32]. Participants at Montefiore received individual and group social skills interventions using the Social Thinking curriculum[33,34]. Children were asked to delay medication changes, when possible. Mean treatment duration was 3.25 months (SD = 2.05 months, range = 1–16 months).

In addition, the other 105 children with ASD from the same sites who were not actively receiving these or other similar treatments served as a "treatment as usual" (TAU) group. Participants in the TAU group included children who received diagnostic assessments but no treatment, as well as those on waiting lists for services, at the four sites. This comparison group included children of equivalent age, IQ, autism symptom severity, gender, race, ethnicity, and maternal education level (see Table 3). We conducted a chi-square test of independence to examine the relation between treatment condition and site. The association was significant, $\chi^2(3, N = 210) = 8.58$, $p = 0.036$, with the CADB site having a larger proportion of TAU cases compared to treatment cases. Specifically, observed count for the TAU condition at CADB was 40, versus an observed count of 22 in the treatment group (expected count of 31 for both treatment and TAU groups).

Ethics. This study was approved by the Institutional Review Boards of Weill Cornell Medicine (#1507016398), New York University Langone Medical Center (#i15-00168 and #i15-00804), the Icahn School of Medicine at Mount Sinai (#15-0325), and the Albert Einstein College of Medicine (#2015-5508). This work was carried out in accordance with the ethical standards of the institutions named above and with the Declaration of Helsinki. Written informed consent was obtained from parents of all participants.

Measures. *BOSCC.* The Brief Observation of Social Communication Change (BOSCC)[5] is a measure of treatment response for social communication behaviors. The BOSCC consists of specific items designed to identify changes in social communication behaviors over relatively short periods of time by quantifying information about the behaviors' quality and frequency. The BOSCC was initially developed by expanding and adding to

	Treatment ($n = 105$) M (SD)	Treatment-as-usual ($n = 105$) M (SD)	t statistic
Age (months)	113.4 (36.8)	108.7 (38.6)	− 0.885
Cognitive skills			
VIQ	99.9 (19.1)	99.4 (18.8)	− 0.206
NVIQ	100.7 (16.7)	101.7 (18.1)	0.390
ADOS-2			
CSS SA	7.5 (1.9)	7.7 (1.7)	0.739
CSS RRB	6.8 (2.4)	6.4 (2.7)	− 1.03
	Treatment n **(%)**	**Treatment-as-usual** n **(%)**	X^2 **statistic**
Sex			0.147
Males	90 (85.7)	88 (83.8)	
Race			5.36
White	58 (55.2)	51 (48.6)	
African American	10 (9.5)	16 (15.2)	
Asian/Pacific Islander	7 (6.7)	5 (4.8)	
American Indian	2 (1.9)	1 (1.0)	
Biracial	9 (8.6)	9 (8.6)	
Other/unspecified	16 (15.3)	14 (13.3)	
Ethnicity			0.934
Hispanic	30 (28.6)	22 (21.0)	
Maternal education			5.50
High school diploma	4 (3.8)	4 (3.8)	
Some college/vocational	12 (11.4)	5 (4.8)	
Associate degree	6 (5.7)	5 (4.8)	
Four-year college degree	22 (21.0)	11 (10.5)	
Graduate/professional degree	26 (24.8)	21 (20.0)	

Table 3. Baseline demographic characteristics. Race was not reported for 3 participants in the treatment group and 9 participants in the treatment-as-usual group. Ethnicity was not reported for 22 participants in the treatment group and 29 participants in the treatment-as-usual group. Maternal education level was not available for 30 participants in the treatment group and 54 participants in the treatment-as-usual group.

ADOS-2 items, but it has now been substantially revised. It is coded from a videotaped observation of a child during a naturalistic social and play interaction. The BOSCC has also been applied to segments of the ADOS in a group of young, minimally verbal children with ASD, and was found to be more sensitive in monitoring changes over the course of treatment as compared to the ADOS Calibrated Severity Score[35]. This study was an early component of a larger effort to develop behavioral coding schemes across language and age levels that is still underway. Tasks from this initial version of the BOSCC for verbal children provide optimal opportunities to obtain language samples of conversation between a child with ASD and an adult social partner within the 12-min, semi-structured, natural interactions. The social partners (examiners) were blind to the treatment condition of the children. The tasks consisted of 4 min of play with various age-appropriate materials (e.g., pinball game, puzzles, transformers, picture books, scratch art, foreign money, conversational cards), followed by 2 min of conversation without any materials present, and another 4 min of play with a different set of materials and 2 min of conversation. The BOSCC was administered prior to treatment entry and again following treatment completion for children in the treatment group. Children in the TAU group were administered the BOSCC before and after a time determined by yoking by treatment length.

Social communication symptom levels. For descriptive purposes, research reliable clinicians administered and scored the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2)[17] for all children prior to treatment entry. The ADOS-2 provides an overall Calibrated Severity Score for social communication symptoms under the Social Affect domain (CSS SA)[36]. The overall level of language item from the ADOS-2 was also used to quantify the language levels, following the new metric that has been validated in a large sample across different ADOS-2 Modules[37]. Higher scores indicate higher language levels (e.g., a score of 5 = regular use of utterances with two or more words; a score of 6 = speech is mostly utterances of at least three words, but without complex language; a score of 7 = some relatively complex speech but with recurrent grammatical errors not associated with use of dialect; a score of 8 = uses sentences in a largely correct fashion). Of the 210 participants, 22 were administered ADOS-2 Module 2, 185 were administered Module 3, and 3 were administered Module 4 (for adolescents ages 16–17 years).

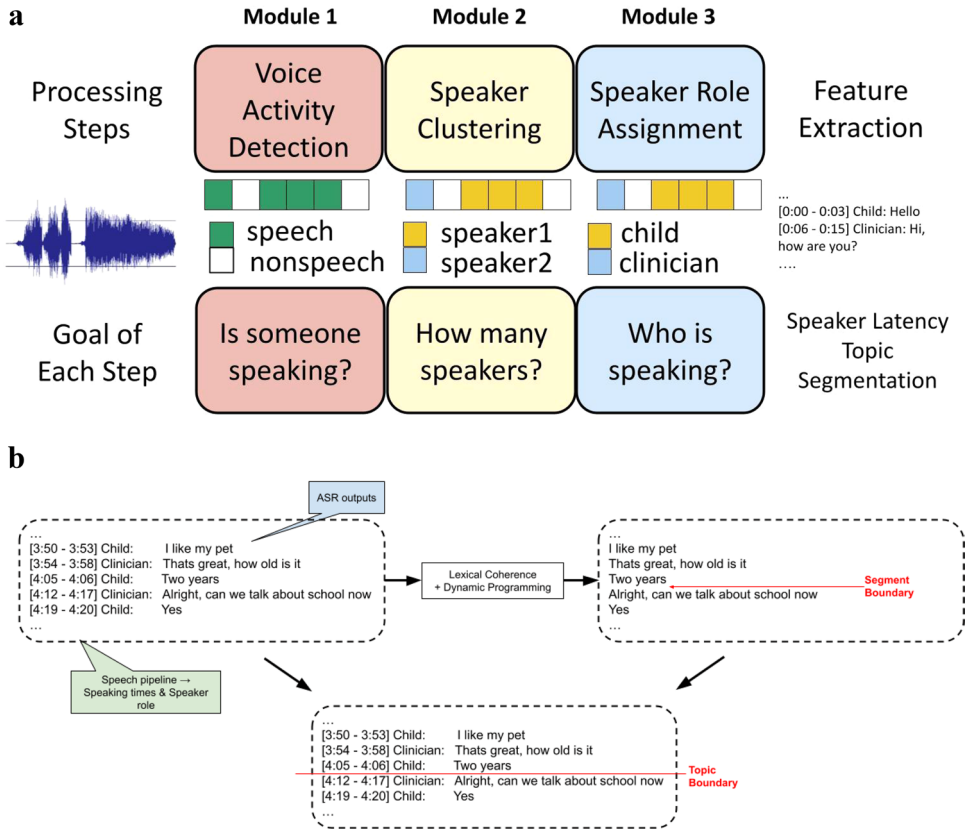

Figure 2. (**a**) Automated speech processing pipeline. Processing steps are depicted in the top row, with the goal of each step depicted in the bottom row. (**b**) Identification of topic boundaries. Topic boundaries (topic start and topic end) were computed automatically using the outputs from automatic speech recognition (ASR) system. Dynamic programming was used to divide the ASR transcript into contiguous segments.

Cognitive levels. For descriptive purposes, examiners blind to the purposes of the study administered cognitive assessments to participants at baseline, including the Differential Ability Scales, Second Edition (DAS-II)[38], Stanford Binet Intelligence Scales, Fifth Edition (SB-5)[39], Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV)[40], Wechsler Intelligence Scales for Children, Fifth Edition (WISC-V)[41], Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV)[42], and Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)[43]. Mean verbal IQ and nonverbal IQ values are reported in Table 3.

Audio data collection. A portable audio–video recording set-up was standardized across all four sites to eliminate unwanted differences in data due to site-specific biases. Specifically, video of the child's face, movements of participants, and the overall interaction was captured from one high-definition camcorder mounted in the corner of the clinical space. Audio was recorded from four high-quality microphones: one far-field shotgun microphone, mounted on the camcorder, recorded the voices of participants along with any ambient noise. Two lapel/lavalier microphones, one attached to the clothing of the child and the other to the examiner, recorded near-field high quality speech signals of each participant. A boundary/surface microphone, mounted on or near the interaction table, recorded the speech of both child and examiner. Lapel microphones were wireless, so that participants were able to move freely around the room; in the case that the lapel microphone was distracting or bothersome for the child, the microphone was not used, and the boundary microphone was used instead to capture near-field speech of the child.

Automated measure of latency. We extracted conversational features for participants in each dyad using an automated speech processing pipeline. The pipeline consists of multiple modules that process the conversational audio to obtain speaking times for the child and the adult interviewer. Figure 2a depicts our novel technological approach.

For the first module, the speech activity detector, we used a two-hidden layer feed-forward deep neural network (DNN) to classify audio into speech and non-speech regions (background silence is treated as non-speech). The input consists of spliced (± 15 frames) 13-dimensional Mel frequency cepstral coefficients (MFCCs) extracted from the audio at each time frame to provide context, while the output nodes are binary labels (speech/non-speech). The DNN is trained to minimize cross-entropy loss. During testing, the class posteriors are smoothed and passed through a threshold to ascertain frame-level decisions about involving speech segments, or not.

In the second module, audio speaker embeddings (x-vectors[44]) were computed at uniform intervals using the extracted speech segments. X-vector speech embeddings have been shown to capture speaker information from audio in multiple use-case scenarios[45,46]. Within each conversation, we clustered x-vectors into two clusters (child and adult) using agglomerative hierarchical clustering. At the final module, speaker roles (child and adult) were assigned to the two clusters based on the number of questions asked. For each cluster, an automatic speech recognition system (ASR) was used to convert the audio into text for obtaining a question count. The cluster with a higher number of questions was assigned as the adult, as expected in a natural social interaction between an adult and a child with ASD.

Once speaking roles (child vs. adult) were obtained for each participant, we computed latency as the duration (in seconds) of non-speech regions between speaker turns. We computed latency for the child and the examiner separately. As opposed to using the entire session to compute latency measures for the child and examiner, we restricted our computation to only times within a conversational topic. We defined a topic as a semantically coherent segment of conversation and estimated topic boundaries in an automated manner using the ASR outputs (text format). We segmented the ASR transcript for a BOSCC session into contiguous chunks (i.e., topics) using dynamic programming[47] to optimize for the overall lexical coherence[48] of the session that is computed as distance measures between word embedding representations. Combining this segmentation output with speaking times obtained from the speech processing pipeline, we obtained topic boundaries in the audio recording (see Fig. 2b). The topic boundaries allowed us to compute latency within each detected topic category, referred to as *intra-topic* latency. We were primarily interested in examining the time it takes for the child or examiner to respond when they are having back-and-forth exchanges *within* each conversational topic, since there is no clear expectation for social responses as one moves from one conversational topic to another.

Data analysis. Data were analyzed using SPSS® Statistics Version 26[49]. Pearson correlations were computed between intra-topic latency and baseline cognitive levels and autism symptom severity. Generalized Linear Mixed Models (GLMM) were used to determine whether significant improvements in conversational features were observed while controlling for baseline cognitive level, symptom severity, language level, age, and gender. GLMM allowed us to determine the effects of the predictor variables on the response variable while accounting for the repeated testing of the same participants (i.e., pre- to post-treatment). GLMM control for missing data[50]; therefore, we included cases with missing timepoints (n = 70 for participants with Time 1 data only). Mixed models included time, treatment condition, and the interaction between time and treatment condition as fixed effects. Additional covariates entered in each model as fixed effects included gender, baseline age, NVIQ, ADOS-2 CSS SA, and ADOS-2 expressive language level. Treatment duration and site effects were also added to secondary models to confirm that the results were not affected by confounding factors. Post-hoc paired sample *t* tests (two-tailed) were performed to examine the significant differences between T1 and T2 for each of the groups. Next, change scores for child and examiner intra-topic latency were examined using independent-samples *t* tests (two-tailed) across treatment and TAU groups. In addition, although independent samples *t* tests revealed that there were no significant differences in latency between treatment and TAU groups at Pre-Intervention (child latency, $t[206] = -0.915$, $p = 0.361$, $d = -0.127$; examiner latency, $t[206] = -1.17$, $p = 0.242$, $d = -0.163$), we conducted additional regression analyses to confirm there were still significant differences in latency between the treatment and TAU groups at Post-Intervention even after controlling for the Pre-Intervention latency as well as NVIQ, autism symptom severity, age, expressive language level, and gender.

Data availability

The data that support the findings of this study are available from the corresponding author (S.H.K.) upon reasonable request.

Received: 6 August 2021; Accepted: 7 February 2022
Published online: 28 February 2022

References

1. Narayanan, S. & Georgiou, P. G. Behavioral signal processing: Deriving human behavioral informatics from speech and language. *Proc. IEEE.* **101**, 1203–1233 (2013).
2. Bone, D., Lee, C.-C., Chaspari, T., Gibson, J. & Narayanan, S. Signal processing and machine learning for mental health research and clinical applications [Perspectives]. *IEEE Signal Process. Mag.* **34**, 195–196 (2017).
3. Bone, D. *et al.* Acoustic-prosodic, turn-taking, and language cues in child-psychologist interactions for varying social demand. In *Proceedings of Interspeech*, 2400–2404 (2013).
4. Bone, D., Bishop, S., Gupta, R., Lee, S., & Narayanan, S. S. Acoustic-prosodic and turn-taking features in interactions with children with neurodevelopmental disorders. In *Proceedings of Interspeech*, 1185–1189, (2016).
5. Grzadzinski, R. *et al.* Measuring changes in social communication behaviors: Preliminary development of the Brief Observation of Social Communication Change (BOSCC). *J. Autism Dev. Disord.* **46**, 2464–2479 (2016).
6. Bolte, E. E. & Diehl, J. J. Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. *J Autism Dev Disord.* **43**(11), 2491–2501 (2013).
7. Anagnostou, E. *et al.* Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism* **19**, 622–636 (2015).
8. Scahill, L. *et al.* Measuring repetitive behaviors as a treatment endpoint in youth with autism spectrum disorder. *Autism* **19**, 38–52 (2015).
9. Grzadzinski, R., Janvier, D. & Kim, S. H. Recent developments in treatment outcome measures for young children with autism spectrum disorder (ASD). *Semin. Pediatr. Neurol.* **34**, 100806 (2020).
10. Vivanti, G. *et al.* Implementing and evaluating early intervention for children with autism: Where are the gaps and what should we do?. *Autism Res.* **11**, 16–23 (2018).

11. Guastella, A. J. *et al.* The effects of a course of intranasal oxytocin on social behaviors in youth diagnosed with autism spectrum disorders: A randomized controlled trial. *J. Child Psychol. Psychiatry.* **56**, 444–452 (2015).
12. Nuernberger, J. E., Ringdahl, J. E., Vargo, K. K., Crumpecker, A. C. & Gunnarsson, K. F. Using a behavioral skills training package to teach conversation skills to young adults with autism spectrum disorders. *Res. Autism Spectr. Disord.* **7**, 411–417 (2013).
13. Mason, R. A., Rispoli, M., Ganz, J. B., Boles, M. B. & Orr, K. Effects of video modeling on communicative social skills of college students with Asperger syndrome. *Dev. Neurorehabil.* **15**, 425–434 (2012).
14. Mann, C. C. & Karsten, A. M. Efficacy and social validity of procedures for improving conversational skills of college students with autism. *J. Appl. Behav. Anal.* **53**, 402–421 (2020).
15. Heeman, P., Lunsford, R., Selfridge, E., Black, L. & van Santen, J. Autism and interactional aspects of dialogue. In *Proceedings of the SIGDIAL 2010 Conference*, 249–252 (2010).
16. Parish-Morris, J. *et al.* Exploring autism spectrum disorders using HLT. In *Proc. Conf. Assoc. Comput. Linguist. Meet.* 74–84 (2016).
17. Lord, C. *et al. Autism Diagnostic Observation Schedule–2nd Edition (ADOS-2)* (Western Psychological Corporation, 2012).
18. Bone, D. *et al.* The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *J. Speech Lang. Hear. Res.* **57**, 1162–1177 (2014).
19. Giles, H. Communication accommodation theory. In *The International Encyclopedia of Communication Theory and Philosophy* (eds Jensen, K. B. & Craig, R. T.) (Wiley, 2016).
20. Dragojevic, M., Gasiorek, J. & Giles, H. Communication accommodation theory. In *The International Encyclopedia of Interpersonal Communication* (eds Berger, C. R. & Roloff, M. E.) 1–21 (Wiley, 2015).
21. Bloch, C., Vogeley, K., Georgescu, A. L. & Falter-Wagner, C. M. INTRApersonal synchrony as constituent of INTERpersonal synchrony and its relevance for autism spectrum disorder. *Front. Robot. AI* **6**, 73 (2019).
22. McNaughton, K. A. & Redcay, E. Interpersonal synchrony in autism. *Curr. Psychiatry Rep.* **22**, 12 (2020).
23. Peper, C., Lieke, E., van der Wal, S. J. & Begeer, S. Autism in action: Reduced bodily connectedness during social interactions?. *Front. Psychol.* **7**, 1862 (2016).
24. Landa, R. J., Holman, K. C., O'Neill, A. H. & Stuart, E. A. Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial. *J. Child Psychol. Psychiatry.* **52**, 13–21 (2011).
25. Srinivasan, S. M. *et al.* The effects of rhythm and robotic interventions on the imitation/praxis, interpersonal synchrony, and motor performance of children with autism spectrum disorder (ASD): A pilot randomized controlled trial. *Autism Res. Treat.* **2015**, e736516 (2015).
26. Laugeson, E. A. & Frankel, F. *Social Skills for Teenagers with Developmental and Autism Spectrum Disorders: The PEERS Treatment Manual* (Routledge, 2011).
27. Beaumont, R. & Sofronoff, K. A multi-component social skills intervention for children with Asperger syndrome: The junior detective training program. *J. Child Psychol. Psychiatry.* **49**, 743–753 (2008).
28. Sikich, L. *et al.* Intranasal oxytocin in children and adolescents with autism spectrum disorder. *N. Engl. J. Med.* **385**, 1462–2147 (2021).
29. Frankel, F. D. & Myatt, R. J. *Children's Friendship Training* (Routledge, 2002). https://doi.org/10.4324/9780203009154.
30. Frankel, F. *et al.* A randomized controlled study of parent-assisted children's friendship training with children having autism spectrum disorders. *J. Autism Dev. Disord.* **40**, 827–842 (2010).
31. Wisdom, J. P., Olin, S., Shorter, P., Burton, G. & Hoagwood, K. Family peer advocates: A pilot study of the content and process of service provision. *J. Child Fam. Stud.* **20**, 833–843 (2011).
32. Jamison, J. M. *et al.* Examining the efficacy of a family peer advocate model for Black and Hispanic caregivers of children with autism spectrum disorder. *J. Autism Dev. Disord.* **9**, 1314–1322 (2017).
33. Winner, M. G. & Crooke, P. J. Social thinking: A training paradigm for professionals and treatment approach for individuals with social learning/social pragmatic challenges. *Perspect. Lang. Learn. Educ.* **16**, 62–69 (2009).
34. Winner, M. G. *Think Social! A Social Thinking Curriculum for School-age Students* (Think Social Publishing, 2008).
35. Kim, S. H., Grzadzinski, R., Martinez, K. & Lord, C. Measuring treatment response in children with autism spectrum disorder: Applications of the Brief Observation of Social Communication Change to the Autism Diagnostic Observation Schedule. *Autism* **23**, 1176–1185 (2019).
36. Gotham, K., Pickles, A. & Lord, C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J. Autism Dev. Disord.* **39**, 693–705 (2009).
37. Mazurek, M. O., Baker-Ericzén, M. & Kanne, S. M. Brief report: Calculation and convergent and divergent validity of a new ADOS-2 expressive language score. *Am. J. Intellect. Dev. Disabil.* **124**, 438–449 (2019).
38. Elliott, C. D. *Differential Ability Scales-II* (Pearson, 2007).
39. Roid, G. H. *Stanford-Binet Intelligence Scales* 5th edn. (Psychological Corporation, 2012).
40. Wechsler, D. *Wechsler Preschool and Primary scale of Intelligence* 4th edn. (Psychological Corporation, 2012).
41. Wechsler, D. *Wechsler Intelligence Scale for Children* 5th edn. (Pearson, 2014).
42. Wechsler, D. *Wechsler Adult Intelligence Scale* 4th edn. (Pearson, 2008).
43. Wechsler, D. *WASI-II: Wechsler Abbreviated Scale of Intelligence* (PsychCorp, 2011).
44. Snyder, D., Garcia-Romero, D., Sell, G., Povey, D. & Khudanpur, S. X-vectors: robust DNN embeddings for speaker recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5329–5333 (2018) https://doi.org/10.1109/ICASSP.2018.8461375.
45. Sell, G. *et al.* Diarization is hard: some experiences and lessons learned for the JHU team in the inaugural DIHARD challenge. In *Proceedings of Interspeech,* 2808–2812, September 2018. https://doi.org/10.21437/Interspeech.2018-1893.
46. Snyder, D. *et al.* Speaker recognition for multi-speaker conversations using X-vectors. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 5796–5800 (2019) https://doi.org/10.1109/ICASSP.2019.8683760.
47. Bellman, R. Dynamic programming. *Science* **153**, 34–37 (1966).
48. Alemi, A. A. & Ginsparg, P. Text segmentation based on semantic word embeddings. *arXiv* arXiv:1503.05543 *Cs* (2015).
49. IBM Corp. *IBM SPSS Statistics for Windows, Version 26.0* (IBM Corp, 2019).
50. Krueger, C. & Tian, L. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biol. Res. Nurs.* **6**, 151–157 (2004).

Acknowledgements

We gratefully acknowledge the participants and their families. This project was funded by the SFARI Research Award RFA #345327.

Author contributions

The authors confirm contribution to the paper as follows: study conceptualization and design: S.H.K., A.D., L.S., A.K, C.L., S.N.; data collection: S.H.K., A.D., L.S., A.K., C.L.; data analysis and interpretation of results: E.P.M., M.K., S.H.K.; draft manuscript preparation: E.P.M. and M.K. with supervision from S.N. and S.H.K. All authors reviewed the results and approved the final version of the manuscript.

Competing interests

The authors declare the following competing interests: Catherine Lord acknowledges receipt of royalties for the Autism Diagnostic Observation Schedule-2; profits related to this study were donated to charity. The remaining authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.H.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com