Interactional coordination between conversation partners with autism using non-verbal cues in dialogues

Tahiya Chowdhury, Veronica Romero and **Amanda Stent**
Davis Institute for Artificial Intelligence, Colby College
4000 Mayflower Hill Dr, Waterville, ME 04901
tchowdhu, ajstent, vcromero@colby.edu

Abstract

The diagnosis of autism spectrum disorder (ASD) is a complex, challenging task as it depends on the analysis of dynamic interactional behaviors during diagnostic conversations, including the degree to and ways in which the individual being assessed coordinates their verbal and non-verbal behaviors with their interlocutor (interpersonal coordination), and the degree to which and ways in which they engage in repetitive behaviors (intrapersonal coordination). In this paper, we look at interactional coordination during diagnostic conversations between a psychologist and children who either are typically developing (TD) or have a diagnosis of ASD. Using Cross-Recurrence Quantification Analysis, a method developed for investigating the behavior of dynamic systems, we measure the coordination of non-verbal behaviors between child and psychologist and test whether these measures can be predictive of diagnosis outcome.

1 Introduction

Autism spectrum disorder (ASD) refers to a range of developmental disabilities that affect people's communication, interaction, learning, and other social behaviors. Adolescents with ASD generally exhibit impairments in social interaction (American Psychiatric Organization, 2013), including difficulty in reciprocating verbal and non-verbal behaviors appropriately as well as repetitive behaviors (Tager-Flusberg and Caronna, 2007; Tager-Flusberg, 1999; Mundy and Markus, 1997; Landa, 2000). Previous research on ASD has examined characteristic difficulties in understanding both verbal and non-verbal communication behaviors including following eye gaze (Baron-Cohen et al., 1997), recognizing and imitating gestures (Hobson and Lee, 1999; Williams et al., 2004) and facial expressions (Drimalla et al., 2021), as well as proper use of language pragmatics and verbal reciprocity (Norbury and Bishop, 2002).

A standard diagnostic tool for autism, the Autism Diagnostic Observation Schedule (Lord et al., 2000), relies on qualitative coding by expert assessors for the presence or absence of certain behavioral markers across multiple structured and naturalistic conversational scenarios. The assessor has to simultaneously engage the child in conversation, monitor their own conversational behavior, and make diagnostic notes based on their observations. Understanding the cognitive demands and subjective nature of this process, previous research has explored the efficacy of machine learning methods for identifying behavioral signals of ASD in conversation data (Fusaroli et al., 2019, 2017, 2022). Recent years have seen much more work on computational tools for providing fine-grained, quantitative measurements of conversational behaviors in autism diagnosis conversations. This includes the use of acoustic-prosodic features such as pitch (Kiss et al., 2012), intonation, and rhythm (Bone et al., 2015); language features such as word usage (Song et al., 2021; Prud'hommeaux et al., 2011), discourse expressions (Yang et al., 2021; Chowdhury et al., 2023), social and cognitive linguistic word counts (Kumar et al., 2016), semantic similarity (Goodkind et al., 2018); pose (Kojovic et al., 2021) and mouth movement (Parish-Morris et al., 2018). However, most of this work has examined behaviors either at the individual utterance level, or via conversation-level aggregate statistics. This means that the moment-to-moment dynamic aspects of coordination in conversation are not well captured. Prior work suggests that typically developing children have been shown to spontaneously modify their interaction patterns more than children with ASD to achieve coordination (Marsh et al., 2013; Drimalla et al., 2021). Quantifying interactional coordination under different conversation contexts during autism diagnosis could thus provide insights into an individual's behavioral flexibility to adapt across conversational

Proceedings of the First ICARD Workshop, pages 24–34
September 12, 2023

contexts and its influence on diagnostic outcomes.

Contributions: First, we introduce recurrence quantification analysis (RQA) and cross-recurrence quantification analysis (CRQA), techniques used in a variety of other fields but almost never used in NLP. RQA and CRQA permit fine-grained modeling of the dynamic systems reflected in one or two time series. Second, we use RQA and CRQA to analyze the dynamic synchronization of non-verbal conversational cues exhibited through the body movements of conversational interlocutors, focusing on conversational diagnostics for autism. We explore the following questions:

- Does interactional coordination evolve differently for typically developing (TD) children and children with ASD?
- Does interactional coordination during autism diagnostic conversations differ by conversational context?
- Can we classify children with ASD and TD children using interactional coordination measures as indicators?

2 Background

Generally speaking, in a conversation or interaction, the interlocutors will coordinate their verbal and non-verbal behaviors (Brennan and Hanna, 2009; Reitter and Moore, 2014; Rasenberg et al., 2020). This dynamic process of coordination is difficult to model or analyze using computational approaches, which generally require fixed-length representations. Traditionally, computational researchers extract summary statistics over the conversation or segments of the conversation (Stenchikova and Stent, 2007; Danescu-Niculescu-Mizil and Lee, 2011; Jones et al., 2014; Dubuisson Duplessis et al., 2021). By contrast, with CRQA it is possible to computationally model the fine-grained patterning of moment-to-moment coordination in conversation.

2.1 Recurrence Quantification Analysis

Recurrence Quantification Analysis (RQA) and Cross-Recurrence Quantification Analysis (CRQA) are non-parametric, non-linear techniques that can be used to analyze any (set of) time series (Zbilut et al., 1998). CRQA has been used by cognitive scientists and psychologists to model the coordination of behaviors by conversational interlocutors (e.g. Dale and Spivey, 2006; Fusaroli et al., 2014; Kodama et al., 2021; Romero and Paxton, 2023), but this technique is so far almost unknown to the NLP community[1].

RQA converts an input time series (with some measure along the y axis and time along the x axis) into a phase-space representation of an estimate of the underlying dynamics of the system that generated it (Webber Jr and Zbilut, 2005). CRQA, an extension of RQA, is used for two time-series (Wallot and Leonardi, 2018). For example, if our only measure of a conversational interlocutor's behavior is the position of their nose (an estimate of directionality of gaze), then our time series would be the x and y coordinates of the positions of each participant's nose over time, but the underlying system would include much more information.

CRQA has been applied to quantify interactions between people in a wide range of modalities. It has been used to quantify heart rate coordination during performances (Konvalinka et al., 2011) and while completing joint construction tasks (Fusaroli et al., 2016). Ramenzoni et al. (2011) found that interpersonal coordination in motor behaviors varies due to the nature of the task performed and can affect individual and joint performance differently. Similarly, Wallot et al. (2016) observed that movement coordination in joint construction tasks depends on the task context, and coordination can affect performance positively or negatively depending on the type of interactions demanded by the task.

CRQA has also been used to explore conversational scenarios to measure the level of coordination through behavior matching in speech, gaze (Richardson and Dale, 2005), and gestures (Louwerse et al., 2012). Shared knowledge between interlocutors that work as a common ground is found to influence coordination achieved during dialogue (Richardson et al., 2007). Richardson et al. (2009) showed that the conversation partner's belief about the contextual information available to the other influences their language usage and coordination. Leonardi (2012) posed conversation as a coordination task where alignment in the form of recurrence can happen in verbal and non-verbal interaction involved, including lexical, syntactic, and movement levels.

[1]We could find only one paper in the ACL Anthology where RQA is used, (Chinaei et al., 2017); other NLP-related papers that use RQA are (Allen et al., 2017; Dale et al., 2018; Song et al., 2023), two of which are unpublished preprints.

The complex dynamics of child language and speech usage in dyadic interaction with adults has been studied using recurrence quantification measures (Cox and van Dijk, 2013) where they found increased dynamic adaptation as the child's language developed with age. Similarly, Dale and Spivey (2006) studied conversations between child and caregiver, and found that the child's ability to coordinate reflects their language acquisition and development. Warlaumont et al. (2010) studied interactional dynamics during conversations between child and caregiver and found recurrent delayed response as an indicator of autism. In a similar study using recurrence analysis, Romero et al. (2016) found differences in interpersonal coordination patterns between children with autism and typically developing across a variety of tasks.

In prior research, several CRQA-based metrics have been used to measure coordination (e.g. Fusaroli et al., 2014; Reuzel et al., 2013; Richardson and Dale, 2005; Louwerse et al., 2012). The ones we used here are:

- **Recurrence Rate (RR)** measures the amount of similarity between the trajectories of the two systems, or the amount of time in which interlocutors showed any kind of interactional coordination during their conversation.
- **Determinism (DET)** measures the determinism or stability of the coordination.
- **Longest Line Length (MaxLine)** is another measure of the stability of the coordination.
- **Entropy (ENTR)** in this context provides a measure of the regularity or irregularity of the coordination over time. Where low ENTR implies regularity of movement and high ENTR means more complex, chaotic movement.
- **Trapping time (TT)** measures the permanence of coordination between the two series.

3 Data

We used video data collected during sessions of the Autism Diagnostic Observation Schedule - Second Edition (ADOS-2), an assessment tool used to categorize ASD (Lord et al., 2000). In this assessment, a child and a certified adult assessor (usually a psychologist) engage in a sequence of semi-structured activities (subtasks). Our data includes fourteen different subtasks from Module 3 of the ADOS-2, which is designed for verbally fluent children and adolescents. Depending on the subtask, the child may be asked to engage in a spontaneous conversation (tell a story, play with toys with the assessor, act out a cartoon, or simply chat) or participate in a structured interview on topics such as social life, friends, or emotions.

Our data involved 29 sessions, each with a different child, administered by a single psychologist assessor. Each session lasted on average 40-60 minutes. 14 children had been previously diagnosed with autism (3 Female) and 15 were age-matched typically developing (TD) children (5 Female) who had not received any diagnosis of a mental disorder in the past. All the children were between the ages of 10 and 15. Those in the ASD group had a mean age of 12.36 years (SD = 1.60) and typically developing children were on average 12.20 years old (SD = 1.93). Of the 14 children with ASD, nine of them were white (no Hispanics) and 5 of them were African American. For the TD group, there were 13 white (2 Hispanics) children, 1 African American child, and 1 Asian child. We split each session recording by subtask using annotations done by a research assistant; the average length of these videos is 5 minutes. We cropped each video into left (child) and right (assessor) participant videos, each of resolution 640x720 pixels. We downsampled the videos to 10 frames per second for efficiency in analysis[2]. We processed the left and right participant videos for each subtask using *OpenPose* (Cao et al., 2017; Cao et al., 2019), obtaining time series of x- and y-coordinates for 25 skeletal keypoints of the person present in the video. In the experiments reported here, we used only four of these key points: nose (head), neck (body), and wrists (hand). This allowed us to capture the temporal dynamics of, and relationships between, the child's and assessor's non-verbal behaviors, without using on-body sensors.

The collection and use of this data were approved by the IRBs at the institutions of the corresponding authors and where the data was collected.

4 Method

In this section, we describe how we fit a CRQA model to our data.

[2] People's movements over time periods of less than 1/10 second are not typically trackable by AI pose tracking software.

Vertical Head Movement

Indicator	F	$p-$value
DET	21.668	< 0.001
TT	6.082	0.05
ENT	21.173	< 0.001
L_{max} (V)	4.918	< 0.05
Delay	84.343	< 0.0001

Horizontal Body Movement

Indicator	F	$p-$value
DET	17.615	< 0.001
TT	14.810	< 0.0001
ENT	37.200	< 0.0001
L_{max} (V)	11.363	< 0.01
Delay	100.395	< 0.0001

Horizontal Hand Movement

Indicator	F	$p-$value
DET	30.069	< 0.0001
TT	20.374	< 0.005
ENT	32.645	< 0.0001
L_{max} (V)	18.810	< 0.005
Delay	78.911	< 0.0001

Vertical Hand Movement

Indicator	F	$p-$value
DET	4.748	< 0.05
TT	1.007	0.3247
ENT	17.712	< 0.001
L_{max} (V)	17.228	< 0.001
Delay	36.080	< 0.0001

Table 1: (Section 5.1) Interpersonal coordination varies between first and last subtasks (results of mixed ANOVA).

4.1 Parameter Selection

For quantifying interactional dynamics between time series pairs of non-verbal conversational behavior, we need to first estimate a set of parameters to reconstruct the phase space dynamics of the time series of interest (Takens, 2006): the embedding dimension m, the delay d, and the radius r.

Delay: The delay d is used to recover the latent dimensions by using embedded copies of the time series at different delays. We estimated delay d using the approach described in prior work (Wallot and Leonardi, 2018). Specifically, we used the delay value at the first local minimum of the average mutual information function (AMI) of the component time-series, since the time-series is most independent of itself at that delay (Abarbanel, 2012).

Embedding dimension: The embedding dimension m is an estimate of the number of latent dimensions responsible for the dynamics. We used the false nearest neighbor (FNN) method (Kennel et al., 1992) to estimate the embedding dimension m. Using the delay parameter to embed the time series, we used the embedding obtained at the first local minimum of FNN.

Radius: The radius r specifies the interval space within which two values are counted as recurrent, as continuous-valued time series usually never repeat at exactly the same value. We chose r by incrementally increasing it until RR reached 4%, which is within the recommended range of 1-5% for behavioral data (Webber Jr and Zbilut, 2005) to balance stochastic and deterministic components of the signal.

As we separately estimated d, m, and r for each analysis, this did not allow us to compare RR across participants and tasks. The dynamics of each conversation were too different to use a fixed set of parameters for all. We used PyRQA (Rawald et al., 2017) for the recurrence analysis and teaspoon (Munch and Khasawneh, 2022) for parameter selection for phase space reconstruction.

We limited our analysis to vertical movement of the head (e.g. nodding), horizontal movement of the body (e.g. postural sway or proximity due to moving closer or apart from the other), and horizontal and vertical movement of hands (gestures), resulting in four pairs of time series of non-verbal cues per conversation. Note that using raw movement as non-verbal behavioral cues is advantageous as it allows us to compare conversations that are seemingly different activity-wise yet involve non-verbal behaviors that are universal to conversations in general.

5 Results

5.1 How does interpersonal coordination change over time and by diagnostic group?

In our first experiment, we examined which movements exhibit greater coordination at the end of ADOS diagnostic conversations vs the beginning. In each of our conversations, participants completed subtasks from Module 3 of ADOS-2 in order; we compared coordination in the last subtask vs the first one. This gave $n = 29$ observations (one pair of subtasks per participant). We used a mixed

Vertical Head Movement

Indicator	Estimates	$p-$value
DET	0.001	0.919
TT	−0.004	0.628
ENT	0.001	0.847
L_{max} (V)	−0.473	0.537
Delay	−3.552	0.082

Horizontal Body Movement

Indicator	Estimates	$p-$value
DET	−0.011	0.429
TT	−0.005	0.548
ENT	0.002	0.843
L_{max} (V)	−2.288	0.278
Delay	−8.679	< 0.01

Horizontal Hand Movement

Indicator	Estimates	$p-$value
DET	−0.004	0.781
TT	−0.003	0.706
ENT	0.004	0.279
L_{max} (V)	−0.530	0.723
Delay	−5.965	< 0.05

Vertical Hand Movement

Indicator	Estimates	$p-$value
DET	−0.007	0.670
TT	0.001	0.934
ENT	0.003	0.668
L_{max} (V)	1.189	0.149
Delay	−2.900	0.204

Table 2: (Section 5.2.1) Interpersonal coordination generally does not vary by diagnostic group (results of linear mixed model).

Vertical Head Movement

Indicator	Estimates	$p-$value
DET	0.002	0.802
TT	0.005	0.532
ENT	0.002	0.708
L_{max} (V)	1.460	< 0.05
Delay	−3.649	< 0.05

Horizontal Body Movement

Indicator	Estimates	$p-$value
DET	0.011	0.443
TT	0.005	0.451
ENT	0.001	0.872
L_{max} (V)	3.225	0.129
Delay	−10.092	< 0.001

Horizontal Hand Movement

Indicator	Estimates	$p-$value
DET	0.013	0.283
TT	−0.010	0.082
ENT	0.003	0.337
L_{max} (V)	2.733	0.051
Delay	−6.942	< 0.005

Vertical Hand Movement

Indicator	Estimates	$p-$value
DET	0.001	0.979
TT	−0.021	< 0.05
ENT	−0.002	0.676
L_{max} (V)	0.788	0.551
Delay	−4.496	< 0.05

Table 3: (Section 5.2.1) Interpersonal coordination varies by subtask type (results of linear mixed model).

2(diagnostic group: TD vs. ASD) x 2(task order: first vs. last) ANOVA for our experiment.

Our results are shown in Table 1. Interpersonal coordination between child and psychologist differs significantly ($p < 0.05$) between the first and last subtask for all modalities, for all metrics other than Trapping Time (which indicates the proportion of time the interlocutors stay in a coordinated state).

For typically developing children, Trapping Time improved between the first task (construction with a puzzle) and the last task (creating a story using props), while for the children with ASD, it remained similar. However, we did not observe a significant difference between the two diagnostic groups for any of the movement modalities, perhaps due to the small number of observations. Prior work on synchronization during joint work made a similar observation: practice and task difficulty improved coordination over time (Louwerse et al., 2012). Here, the subtask type can be considered analogous to task difficulty, so we next look at how coordination changes depending on the subtask type.

5.2 How does interpersonal coordination change with subtask type and diagnostic group?

In our second experiment, we grouped the ADOS-2 module 3 conversation-centric tasks into two types:

Vertical Head Movement

Indicator	Estimates	$p-$value
DET	0.023	< 0.05
TT	0.010	0.059
ENT	0.000	0.829
L_{max} (V)	2.078	< 0.01
Delay	-1.644	0.399

Horizontal Body Movement

Indicator	Estimates	$p-$value
DET	-0.027	0.111
TT	-0.017	0.171
ENT	0.000	0.866
L_{max} (V)	-0.622	0.613
Delay	-6.444	< 0.05

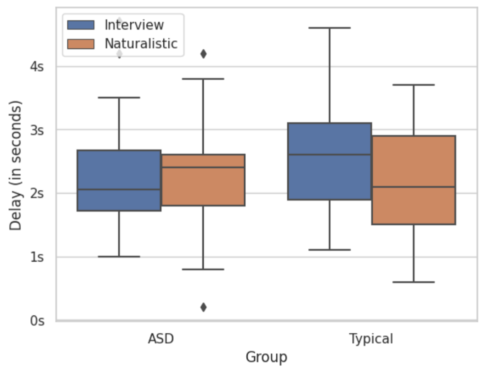
Horizontal Hand Movement

Indicator	Estimates	$p-$value
DET	-0.018	0.439
TT	-0.013	0.454
ENT	0.169	0.909
L_{max} (V)	-0.622	0.613
Delay	-5.563	< 0.05

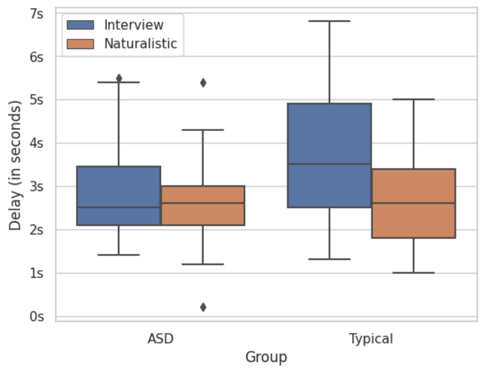
Vertical Hand Movement

Indicator	Estimates	$p-$value
DET	-0.010	0.576
TT	-0.007	0.650
ENT	0.000	0.872
L_{max} (V)	-0.130	0.917
Delay	-2.051	0.396

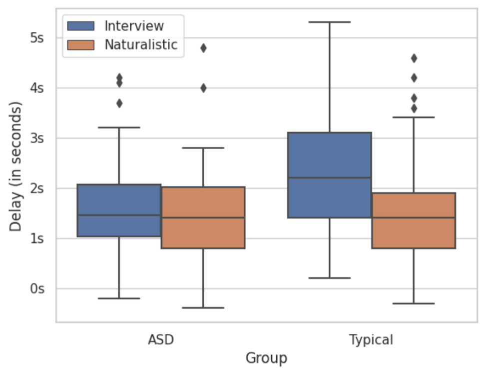
Table 4: (Section 5.2.2) Intra-personal coordination generally does not vary by diagnostic group (results of linear mixed model).



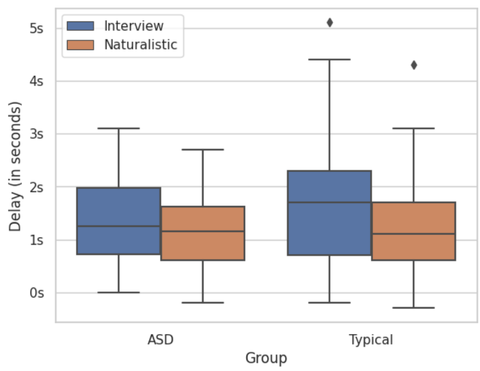
(a) Vertical Head Movement



(b) Horizontal Body Movement



(c) Horizontal Hand Movement



(d) Vertical Hand Movement

Figure 1: Difference in latency for interpersonal coordination in movement

naturalistic and interview. Naturalistic subtasks are those in which the interlocutors engage in unstructured conversations (discussing a picture, talking about a topic of interest, silent play, or unstructured conversation). Interview subtasks are those in which the assessor uses a structured sequence of questions (interviews about social life, friends, or emotions). This gave $n = 169$ observations (three subtasks per participant per subtask type).

Repetitive behavior is a symptom of autism; therefore, we looked at both interpersonal coordination (coordination between interlocutors) us-

Model	Acc.	Prec.	Recall	F1
Naive Bayes	0.48	0.50	0.48	0.40
Decision Tree	0.61	0.68	0.61	0.62
Random Forest	0.62	0.68	0.62	0.63
Adaboost	0.60	0.76	0.60	0.63

Model	Acc.	Prec.	Recall	F1
Naive Bayes	0.30	0.11	0.30	0.15
Decision Tree	0.66	0.72	0.66	0.66
Random Forest	0.61	0.65	0.61	0.61
Adaboost	0.61	0.66	0.61	0.61

Table 5: Classification of autism diagnosis using CRQA metrics of interpersonal coordination (left) and RQA metrics of intrapersonal coordination (right).

ing CRQA, and intrapersonal coordination (self-coordination of the child's behaviors) using RQA. We used linear mixed-effects models with the diagnostic group as the between-group factor and conversation type as the within-group factor.

5.2.1 Interpersonal coordination

Our results are shown in Tables 2 and 3.

Coordination measures decreased or remain nearly unchanged in all 4 behavioral modalities for children with ASD compared to children who are typically developing (Table 2). However, Delay (the latency at which coordination happens) is significantly lower for head movement, hand horizontal movement, and vertical hand movement ($p < 0.05$) in naturalistic subtasks compared to interview subtasks (Table 3). The negative coefficients for all three cases (-3.649, -6.942, -4.496) suggest that children respond more immediately in naturalistic subtasks, perhaps because of the relaxed nature of the conversations. Furthermore, there are significant diagnostic group differences in horizontal body movement and horizontal hand movement for Delay ($p < 0.05$). The negative coefficients for both (-8.679, -5.965) suggest that movement latency is reduced for children with ASD compared to children who are typically developing.

5.2.2 Intrapersonal coordination

Our results are shown in Table 4. We found no significant differences in intrapersonal coordination between naturalistic and interview subtask types. Delay (the latency at which coordination happens) is significantly lower for children with ASD than for children who are typically developing. For vertical head movement (which captures behaviors such as nodding), intrapersonal coordination shows a higher deterministic pattern for children with ASD. This is also evident from Figure 1, which depicts the distribution of coordination delay for the two subtask types between the two diagnostic groups. For children with ASD, in all 4 behavioral modalities, delay either increased or remain unchanged.

5.3 Can information from RQA and CRQA analyses be successfully used in diagnostic classification models?

Recurrence analysis metrics from time series generated from wearable sensors have been used for detecting repetitive motor movement (Großekathöfer et al., 2017). In our third experiment, we used recurrence analysis metrics from time series generated via AI-based human body-skeleton detection in ADOS conversational assessments as features for the classification of autism diagnosis.

As features, we used the same five recurrence analysis metrics as above (DET, TT, ENT, L_{max}, and Delay) obtained from vertical head movement, horizontal body movement, and horizontal and vertical hand movement data from each subtask, plus the task itself, This gave 6 features for each of $n = 1568$ observations (four movements for each of 14 subtasks for each of 29 participating conversational pairs, with 14 subtasks missing because the children declined to participate).

To ensure generalization for out-of-sample testing, we performed cross-validation by using leave-n-user-out [3]. We report results averaged over 10 runs, where 80% (23) of children were randomly selected for training and 20% (6) for testing in each run. We experimented with simple, relatively interpretable classification approaches: Naive Bayes, Decision Tree, Random Forest, and Adaboost[4]. All experiments were run using the scikit-learn and sciPy libraries with default parameter settings. We report accuracy, precision, recall, and F1-score.

We trained one set of models using RQA metrics capturing intrapersonal coordination and another using CRQA metrics capturing interpersonal

[3]We chose this over 10-fold cross-validation, as this ensures our training set does not include information from a child who is also present in our test set.

[4]Our goal in this work is to give assessors *assessment support tools*, not to support automated diagnosis. For this reason, the interpretability of model decisions is important. In addition to the results presented here, we tried multi-layer perceptrons and support vector machines, which performed worse than a random baseline.

coordination. Our results are shown in Table 5. Decision trees were the best-performing approach with RQA across all metrics; tree-based models (Adaboost, Random Forests, and decision trees) performed similarly with CRQA metrics. In both cases, Naive Bayes performs worse than a random baseline; this can be attributed to its strong feature independence assumption. No modeling approach gave results good enough likely to make it a useful assessment support tool for ADOS assessors; however, it is possible that by combining CRQA or RQA-derived features with acoustic/prosodic and language-derived features, we could obtain better results (see Chowdhury et al., 2023).

6 Limitations

We would like to emphasize that this study is preliminary. The sample set is relatively small, and the number of non-verbal behaviors we had the opportunity to evaluate is also relatively small.

In addition, although our results and those of others cited in this paper show that automated measures extracted from autism assessments may be somewhat predictive of autism diagnosis, we *in no way* mean to imply that it is now or will soon be possible (or even desirable) to automate autism diagnosis. Especially since the tools available to us at the moment, like *OpenPose* need to be improved to consistently extract movement data from video (Chung et al., 2022). We would like to drive this point home:

- There is an element of subjectivity in manual assessment diagnosis of autism; yet, all the diagnosis data that we have comes from these manual assessments. This means the labels are noisy.

- There are demographic limitations in the available data. The relatively small amount of data available from manual assessments for autism is not balanced for important factors including sex, gender, ethnicity, race, country of origin, language, age, educational status, or income status. This means the data is biased.

- Given these concerns, and the growing literature on biased outcomes of automated assessments for marginalized populations, it is our position that *any decision* involving a significant outcome for a human being should have a human involved.

Our long-term goal is therefore not to provide a machine learning-based "autism test", but to provide machine learning-based automatic measures that an assessor can use to examine the acceptability of assessment sessions and to inform their own diagnostic decisions.

7 Conclusions and Future Work

In this work, we explore a method for quantifying interactional coordination in autism assessment. We use non-verbal movement exhibited through head, body, and hand positions to capture movement dynamics during conversation and measure coordination over time. We show that coordination between interlocutors changes over time for both children who are typically developing and children who exhibit symptoms of ASD.

Importantly, we find that the level and stability (as measured by L_{max} and TT) of both inter-and intra-personal coordination do not generally differ by diagnostic group. We also find that coordination delay was significantly lower for the ASD group in both interpersonal and intra-personal coordination, which conforms to the existing literature on response delay as a symptom of autism. Interestingly, this finding does provide some insight into how children in each diagnostic group exhibit different dynamics, even though these differences were not captured by ENT. Contextual information such as subtask type (interview vs. naturalistic) does influence the degree of coordination between the interlocutors, but does not affect the child's coordination within their own behavior.

In future research, we plan to extend our analysis to other interactional behaviors including non-verbal cues such as facial expressions, eye gaze and verbal cues such as acoustic-prosodic behaviors (pitch, intonation), word usage, and discourse usage for measuring interpersonal and intra-personal coordination during diagnostic conversations.

8 Acknowledgements

We would like to thank and acknowledge all our participants and their families. The data set was obtained thanks to support from the National Institute of Mental Health and the National Institutes of Health: awards R21MH094659 and F31MH108331.

References

Henry Abarbanel. 2012. *Analysis of observed chaotic data*. Springer Science & Business Media.

Laura K Allen, Aaron D Likens, and Danielle McNamara. 2017. Recurrence quantification analysis: A technique for the dynamical analysis of student writing. In *Proceedings of the International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pages 240–245.

American Psychiatric Organization. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Publishing, Washington, DC.

Simon Baron-Cohen, Sally Wheelwright, , and Therese Jolliffe. 1997. Is there a "language of the eyes"? evidence from normal adults, and adults with autism or asperger syndrome. *Visual Cognition*, 4(3):311–331.

Daniel Bone, Matthew P Black, Anil Ramakrishna, Ruth B Grossman, and Shrikanth S Narayanan. 2015. Acoustic-prosodic correlates of 'awkward'prosody in story retellings from adolescents with autism. In *Proceedings of INTERSPEECH*, pages 1616–1620.

Susan E Brennan and Joy E Hanna. 2009. Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2):274–291.

Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of CVPR*.

Hamidreza Chinaei, Leila Chan Currie, Andrew Danks, Hubert Lin, Tejas Mehta, and Frank Rudzicz. 2017. Identifying and avoiding confusion in dialogue with people with alzheimer's disease. *Computational Linguistics*, 43(2):377–406.

Tahiya Chowdhury, Veronica Romero, and Amanda Stent. 2023. Parameter selection for analyzing conversations with autism spectrum disorder. In *Proceedings of INTERSPEECH*.

Jen-Li Chung, Lee-Yeng Ong, and Meng-Chew Leow. 2022. Comparative analysis of skeleton-based human pose estimation. *Future Internet*, 14(12).

Ralf F. A. Cox and Marijn van Dijk. 2013. Microdevelopment in parent-child conversations: From global changes to flexibility. *Ecological Psychology*, 25(3):304–315.

Rick Dale, Nicholas D Duran, and Moreno Coco. 2018. Dynamic natural language processing with recurrence quantification analysis. *arXiv preprint arXiv:1803.07136*.

Rick Dale and Michael J Spivey. 2006. Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56(3):391–430.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 76–87.

Hanna Drimalla, Irina Baskow, Behnoush Behnia, Stefan Roepke, and Isabel Dziobek. 2021. Imitation and recognition of facial emotions in autism: a computer vision approach. *Molecular Autism*, 12:1–15.

Guillaume Dubuisson Duplessis, Caroline Langlet, Chloé Clavel, and Frédéric Landragin. 2021. Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, 55:353–388.

Riccardo Fusaroli, Johanne S Bjørndahl, Andreas Roepstorff, and Kristian Tylén. 2016. A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9):1297.

Riccardo Fusaroli, Ruth Grossman, Niels Bilenberg, Cathriona Cantio, Jens Richardt Møllegaard Jepsen, and Ethan Weed. 2022. Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Research: Official Journal of the International Society for Autism Research*, 15(4):653–664.

Riccardo Fusaroli, Ivana Konvalinka, and Sebastian Wallot. 2014. Analyzing social interactions: the promises and challenges of using cross recurrence quantification analysis. In *Translational recurrences: From mathematical theory to real-world applications*, pages 137–155. Springer.

Riccardo Fusaroli, Anna Lambrechts, Dan Bang, Dermot M Bowler, and Sebastian B Gaigg. 2017. Is voice a marker for autism spectrum disorder? a systematic review and meta-analysis. *Autism Research: Official Journal of the International Society for Autism Research*, 10(3):384–407.

Riccardo Fusaroli, Ethan Weed, Deborah Fein, and Letitia Naigles. 2019. Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, 183:1–18.

Adam Goodkind, Michelle Lee, Gary E Martin, Molly Losh, and Klinton Bicknell. 2018. Detecting language impairments in autism: A computational analysis of semi-structured conversations with vector semantics. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2018*, pages 12–22.

Ulf Großekathöfer, Nikolay V Manyakov, Vojkan Mihajlović, Gahan Pandina, Andrew Skalkin, Seth Ness, Abigail Bangerter, and Matthew S Goodwin. 2017. Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. *Frontiers in neuroinformatics*, 11:9.

R Peter Hobson and Anthony Lee. 1999. Imitation and identification in autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(4):649–659.

Simon L Jones, Rachel Cotterill, Nigel Dewdney, Kate Muir, and Adam N Joinson. 2014. Finding zelig in text: A measure for normalising linguistic accommodation. In *Proceedings of COLING*.

Matthew B Kennel, Reggie Brown, and Henry DI Abarbanel. 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical review A*, 45(6):3403.

Géza Kiss, Jan PH van Santen, Emily Prud'Hommeaux, and Lois M Black. 2012. Quantitative analysis of pitch in speech of children with neurodevelopmental disorders. In *Proceedings of INTERSPEECH*.

Kentaro Kodama, Daichi Shimizu, Rick Dale, and Kazuki Sekine. 2021. An approach to aligning categorical and continuous time series for studying the dynamics of complex human behavior. *Frontiers in Psychology*, 12:614431.

Nada Kojovic, Shreyasvi Natraj, Sharada Prasanna Mohanty, Thomas Maillart, and Marie Schaer. 2021. Using 2d video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Scientific Reports*, 11(1):1–10.

Ivana Konvalinka, Dimitris Xygalatas, Joseph Bulbulia, Uffe Schjødt, Else-Marie Jegindø, Sebastian Wallot, Guy Van Orden, and Andreas Roepstorff. 2011. Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences*, 108(20):8514–8519.

Manoj Kumar, Rahul Gupta, Daniel Bone, Nikolaos Malandrakis, Somer Bishop, and Shrikanth S. Narayanan. 2016. Objective Language Feature Analysis in Children with Neurodevelopmental Disorders During Autism Assessment. In *Proceedings of INTERSPEECH*, pages 2721–2725.

Rebecca Landa. 2000. Social language use in Asperger syndrome and high-functioning autism. In *Asperger Syndrome*, pages 125–155. The Guilford Press, New York, NY, US.

Giuseppe Leonardi. 2012. The study of language and conversation with recurrence analysis methods. *Psychology of Language and Communication*, 16(2):165–183.

Catherine Lord, Susan Risi, Linda Lambrecht, Edwin H. Cook, Bennett L. Leventhal, Pamela C. DiLavore, Andrew Pickles, and Michael Rutter. 2000. The Autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of Autism. *Journal of Autism and Developmental Disorders*, 30(3):205–223.

Max M Louwerse, Rick Dale, Ellen G Bard, and Patrick Jeuniaux. 2012. Behavior matching in multimodal communication is synchronized. *Cognitive science*, 36(8):1404–1426.

L. Marsh, A. Pearson, D. Ropar, and A. Hamilton. 2013. Children with autism do not overimitate. *Current Biology*, 23(7):R266–R268.

Elizabeth Munch and Firas Khasawneh. 2022. teaspoon.

Peter Mundy and Jessica Markus. 1997. On the nature of communication and language impairment in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, 3(4):343–349.

Courtenay Frazier Norbury and Dorothy VM Bishop. 2002. Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language & Communication Disorders*, 37(3):227–251.

Julia Parish-Morris, Evangelos Sariyanidi, Casey Zampella, G. Keith Bartley, Emily Ferguson, Ashley A. Pallathra, Leila Bateman, Samantha Plate, Meredith Cola, Juhi Pandey, Edward S. Brodkin, Robert T. Schultz, and Birkan Tunç. 2018. Oral-motor and lexical diversity during naturalistic conversations in adults with autism spectrum disorder. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*.

Emily T. Prud'hommeaux, Brian Roark, Lois M. Black, and Jan van Santen. 2011. Classification of atypical language in autism. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 88–96.

Verónica C Ramenzoni, Tehran J Davis, Michael A Riley, Kevin Shockley, and Aimee A Baker. 2011. Joint action in a cooperative precision task: nested processes of intrapersonal and interpersonal coordination. *Experimental Brain Research*, 211:447–457.

Marlou Rasenberg, Asli Özyürek, and Mark Dingemanse. 2020. Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, 44(11):e12911.

Tobias Rawald, Mike Sips, and Norbert Marwan. 2017. Pyrqa—conducting recurrence quantification analysis on very long time series efficiently. *Computers and Geosciences*, 104:101–108.

David Reitter and Johanna D Moore. 2014. Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76:29–46.

Ellen Reuzel, Petri JCM Embregts, Anna MT Bosman, Ralf Cox, Maroesjka van Nieuwenhuijzen, and Andrew Jahoda. 2013. Conversational synchronization in naturally occurring settings: A recurrence-based analysis of gaze directions and speech rhythms of staff and clients with intellectual disability. *Journal of Nonverbal Behavior*, 37:281–305.

Daniel C Richardson and Rick Dale. 2005. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6):1045–1060.

Daniel C Richardson, Rick Dale, and Natasha Z Kirkham. 2007. The art of conversation is coordination. *Psychological Science*, 18(5):407–413.

Daniel C Richardson, Rick Dale, and John M Tomlinson. 2009. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, 33(8):1468–1482.

Veronica Romero, Paula Fitzpatrick, RC Schmidt, and Michael J Richardson. 2016. Using cross-recurrence quantification analysis to understand social motor coordination in children with autism spectrum disorder. In *Recurrence Plots and Their Quantifications: Expanding Horizons: Proceedings of the 6th International Symposium on Recurrence Plots*, pages 227–240. Springer.

Veronica Romero and Alexandra Paxton. 2023. Stage 2: Visual information and communication context as modulators of interpersonal coordination in face-to-face and videoconference-based interactions. *Acta Psychologica*, 239:103992.

Amber Song, Meredith Cola, Samantha Plate, Victoria Petrulla, Lisa Yankowitz, Juhi Pandey, Robert T Schultz, and Julia Parish-Morris. 2021. Natural language markers of social phenotype in girls with autism. *Journal of Child Psychology and Psychiatry*, 62(8):949–960.

Yukyeong Song, Wanli Xing, Xiaoyi Tian, and Chenglu Li. 2023. Are we on the same page? modeling linguistic synchrony and math literacy in mathematical discussions.

Svetlana Stenchikova and Amanda Stent. 2007. Measuring adaptation between dialogs. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 166–173.

Helen Tager-Flusberg. 1999. A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry*, 11(4):325–334.

Helen Tager-Flusberg and Elizabeth Caronna. 2007. Language disorders: Autism and other pervasive developmental disorders. *Pediatric Clinics of North America*, 54(3):469–481.

Floris Takens. 2006. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pages 366–381. Springer.

Sebastian Wallot and Giuseppe Leonardi. 2018. Analyzing multivariate dynamics using cross-recurrence quantification analysis (CRQA), diagonal-cross-recurrence profiles (DCRP), and multidimensional recurrence quantification analysis (MdRQA) – a tutorial in R. *Frontiers in Psychology*, 9.

Sebastian Wallot, Panagiotis Mitkidis, John J McGraw, and Andreas Roepstorff. 2016. Beyond synchrony: joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony. *PloS One*, 11(12):e0168306.

Anne S Warlaumont, D Kimbrough Oller, Rick Dale, Jeffrey A Richards, Jill Gilkerson, and Dongxin Xu. 2010. Vocal interaction dynamics of children with and without autism. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32.

Charles L Webber Jr and Joseph P Zbilut. 2005. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, 94(2005):26–94.

Justin HG Williams, Andrew Whiten, and Tulika Singh. 2004. A systematic review of action imitation in autistic spectrum disorder. *Journal of autism and developmental disorders*, 34:285–299.

Christine Yang, Duanchen Liu, Qingyun Yang, Zoey Liu, and Emily Prud'hommeaux. 2021. Predicting pragmatic discourse features in the language of adults with autism spectrum disorder. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*, pages 284–291, Online. Association for Computational Linguistics.

Joseph P Zbilut, Alessandro Giuliani, and Charles L Webber Jr. 1998. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, 246(1-2):122–128.