2024 33rd IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 26-30, 2024. Pasadena, California, USA.

2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN) | 979-8-3503-7502-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/RO-MAN60168.2024.10731459

An Exploration into the Design of Multi-Session Robot-Mediated Joint Attention Intervention for Young Children with Autism

Guangtao Nie, Zhi Zheng, *Member IEEE*, Amy R. Swanson, Amy S. Weitlauf, Zachary E. Warren, and Nilanjan Sarkar, *Senior Member, IEEE*

Abstract— **One in 36 children in the United States has autism. Numerous robotic intervention systems have been proposed for children with autism. It is widely acknowledged that engagement with the robotic system is important for intervention success. Visual attention toward the robot can be used as a proxy for engagement. However, less is known about how to maintain and enhance visual attention within a robotic system, and how the variation of visual attention may influence adaptation of dynamic intervention protocols to improve outcomes. Therefore, in this work, we propose a new metric, System Capture Ratio (SCR), that can be automatically quantified in real-time, to measure a participant's visual attention within a robotic system designed for joint attention intervention for children with autism. Then, we demonstrate that compared to a static intervention protocol, a dynamic intervention protocol can help sustain visual attention and thus achieve a significant performance improvement. The results support the implementation of adaptive and dynamic robotic intervention protocols for autistic children that are based on SCR, offering suggestions on designing effective multi-session studies for autism intervention.**

I. Introduction

According to the latest Centers for Disease Control and Prevention (CDC) report, 1 in 36 children in the U. S. has autism [1]. Children with autism show core differences in social reciprocity and communication [2]. In early childhood, this includes difficulties following another person's cue to share attention, which is defined as "response to joint attention" (RJA). Joint attention plays a critical and generative role in language development and later social responsiveness as well as adult outcomes [3], [4]. Early assessment and behavioral intervention have shown promising results to help alleviate heterogeneous symptoms in young children with autism [5]. In the technological realm, the Socially Assistive Robot (SAR) has gained momentum due to its ability to engage and evoke various social behaviors in autistic children as well as its capacity of providing uniform intervention delivery and objective, quantitative performance tracking [6]-[8].

Although several robotic studies have shown the potential of robotic intervention for autistic children [9]-[11], Controlled, multi-session studies with well-defined participants are necessary [12] to explore their experiences with robotic systems over time [13]. In a longitudinal robot-mediated intervention study, changes in participants' performance may result from both the robot's influence and natural maturation. To accurately assess the robot's intervention effect, it's essential to include the natural maturation of children throughout the study period in the data analyses [14]. Therefore, we conducted multi-session studies with delayed participation groups to investigate the effects of robot intervention and natural maturation.

A. Sustaining engagement in Socially Assistive Robots

Drawing the attention and interest of children with autism onto both a robot as well as social tasks are fundamental requirements for robotic therapy, especially when dealing with lengthy sessions [6]. Several works have investigated how to engage participants in human-robot interaction (HRI): Tielman et al. [15] introduced adaptive emotional expressions to promote the engagement of children during HRI with the NAO robot. Breazeal et al. [16] developed an emotion recognizer to perceive the emotion of participants and showed that HRI were quite engaging if the robot's expressive behaviors were adaptive, timely, and appropriate.

This work was supported in part by the U.S. NIH under Grant 5R33MH103518-04 and 1R21MH111567

G. Nie is with the JD.com, Beijing, China; email: nieguangtao1@jd.com. This work was done when he was a PhD student at Vanderbilt University.

Z. Zheng is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA; email: zzheng3@nd.edu

A. R. Swanson is with the Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University Medical Center, Nashville, TN 37212 USA; email: amy.r.swanson@vanderbilt.edu

A. S. Weitlauf and Z. E. Warren are with the Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN 37212 USA; email: amy.s.weitlauf@vanderbilt.edu; zachary.e.warren@vanderbilt.edu

N. Sarkar is with the Mechanical Engineering Department, Electrical Engineering and Computer Science Department, Vanderbilt University, Nashville, TN 37212 USA; email: nilanjan.sarkar@vanderbilt.edu

B. Visual attention tracking

Although the role of visual attention has been explored [17], [18], works on how to maintain or increase the visual attention of autistic children during HRI has been limited: De Silva et al. [19] analyzed visual attention patterns of 5 autistic children with a therapeutic assistive robot in time segments through an unsupervised method. Joint attention was successfully established in 3 attempts for all 5 participants. However, this was not a controlled multi-session study. Schadenberg et al. [20] conducted a single-session study for autistic children aged 6-11 years with a control group. Manual video coding on participants' visual attention demonstrated that lower predictability of a robot may lead to less activity-related visual attention and negative learning influence. Based on a study with 17 children with autism (3-7 years old), Clabaugh et al. [21] reported that for long-term intervention in home settings, adaptive instruction and feedback were critical to performance improvements and retention. In this work, reinforcement learning was adopted to challenge the participants without pushing them to make too many mistakes. Scassellati et al. [22] demonstrated that a SAR deployed at home could improve RJA skills of 12 autistic children (7-12 years old) as the robot encouraged engagement of participants during triadic play by adapting the difficulties of the activities. Works [21] and [22] did not include control groups.

C. Innovations and contributions of the current work

In our study, we introduce the System Capture Ratio (SCR) as a novel, automatically tracked metric for measuring visual attention in a robotic intervention system for children with autism. We show, for the first time, that SCR can be used to sustain visual attention during joint attention interventions, leading to performance improvements not extensively explored in prior research. Our experimental design involved a detailed multi-session protocol to monitor changes in the children's attention and performance, with participants randomized into an immediate intervention group and a waitlist control group to account for natural development. This research is among the pioneering efforts to empirically demonstrate that a robotic system can maintain the visual attention of children with ASD and that such sustained attention can significantly enhance performance outcomes in a structured multi-session context.

The remainder of this paper is organized as follows: Section II describes the robotic system developed for the presented experiments and an initial study without an adaptive interaction protocol. Section III critically analyzes the limitations of this initial study thorough the lens of decreased visual attention. Section IV introduces an improved study with an adaptive intervention protocol designed to improve visual attention. Section V present a detailed comparative analysis between the two studies. Finally, in Section VI we summarize the contributions of the work and discuss current limitations and future work.

II. NORRIS OVERVIEW

A. Experimental robotic system

Noncontact Responsive Robot-mediated Intervention System (NORRIS) [23] was designed to be a closed-loop HRI system that could automatically initiate RJA trials, track participant performance, and respond accordingly. The experimental setup and the architecture of NORRIS are shown in Fig. 1. NORRIS includes five components: supervisory controller, robot administrator, gaze tracking subsystem, target monitors, and reward videos. These components together constituted a closed-loop HRI process to initiate, train, and reinforce RJA skills.

B. Intervention task design

The interaction task was designed according to the hierarchical RJA skill assessment embedded within the Autism Diagnostic Observation Schedule-2 (ADOS-2) [24], the gold standard autism diagnostic instrument.

To accommodate training experiences across various participant skill levels, trials of differing complexities were provided: single-target, double-target, and triple-target trials. In single-target trials, participants were guided by the robot to focus on either the left or right target monitor. Double-target trials involved a sequence of prompts, either left to right or right to left. Similarly, triple-target trials featured a sequence such as left-right-left or right-left-right. A hierarchy of prompts from least to most complex [23] in single and double-target trials used different stimuli to direct participants towards the common target(s). Following each prompt, participants had a 5-second (for single-target trials) or 7-second (for double and triple-target trials) window to respond by looking at the target(s).

A trial was deemed successful when the participant gazed at the intended target(s) in response to the robot's cues within a specified time frame (target hit). The final cue index required for the success of this trial was recorded, succeeded by the presentation of a reward video on the target monitor to conclude the trial.

C. Performance measurements

The RJA performance of a session that contained multiple trials was evaluated by two measurements: *hit rate* and *prompt level*, defined as follows:

hit rate: *# of successful trials / # of all trials in the session*; (1)

prompt level: average of the last prompt indices across trials. (2)

A higher hit rate signifies more successful trials, while a lower prompt level indicates trials were concluded with fewer cues. Consequently, a higher hit rate and a lower prompt level suggest improved RJA performance. The hit rate is a more essential measure because it reveals if a participant can identify the shared targets, central to RJA abilities; in contrast, the prompt level assesses the extent of assistance required by a participant to locate the shared targets.

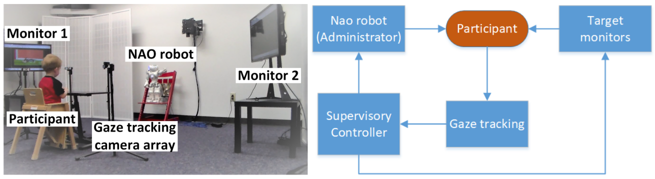

Fig. 1. System environment A) NORRIS experimental setup; B) NORRIS architecture

D. *Analysis of children's task engagement*

Observations from a study team indicated a significant drop in engagement among participants with NORRIS during an initial study with a simple, repetitive interaction protocol (Study I, Section III). This decline in engagement was attributed to the loss of novelty in repeated sessions, leading to disinterest [7], [25]. Disengaged children often physically turned away from the system or covered their faces, behaviors that were noted when their gaze shifted away from or was blocked within the interaction zone. To quantify engagement, a new metric called System Capture Ratio (SCR) was introduced, based on gaze distribution. SCR, enabled by NORRIS's gaze tracking covering the entire interaction area, measures the percentage of time a participant's gaze is captured by the system during a session or trial.

Using SCR, we anaylized chlidren's engagement in Study I, followed by the motivations and proposed improvements that led to an improved study (Study II, descrived in Section IV).

III. Study I-Initial study failed to engage participants

A. *Participants Characteristics*

Twelve young children with ASD (9 males, 3 females, age range = 2.03-4.64 years; Mean age = 2.62 years, standard deviation = 0.33) were enrolled in study I. Each participant scored on the ADOS-2 above the thresholds for autism spectrum classification. Participants were allocated randomly to either an immediate participation group (IPG) or a waitlist control group (WCG). No significant statistical differences were observed in diagnostic features (e.g., age at recruitment, ADOS-2 total raw score, SRS-2 total raw score, SRS-2 T-score, SCQ Current Total Score, MSEL ELC) or in the duration since diagnosis between both groups ($p > .35$, two-sided Wilcoxon rank sum test). The research received approval from the Vanderbilt University Institutional Review Board, IRB Protocol 170022. Informed consent was obtained from all participating families before enrollment by principal study personnel.

B. *Study timeline and protocol*

Table I illustrates the use of a waitlist control design to examine two variables potentially affecting RJA performance: natural maturation and the NORRIS intervention. Assessments are denoted by A, while interventions are indicated by S. The IPG underwent stages of assessment-intervention-assessment, whereas the WCG went through assessment-wait-assessment-intervention-assessment phases. On average, the time span from IPG-A1 to IPG-A2 was 6.25 weeks (SD = 1.65), and from WCG-A0 to WCG-A1 was 6.40 weeks (SD = 1.70). Wilcoxon rank sum test ($p > .10$) shows no statistically significant difference between the two time periods.

The study design offered two opportunities. Firstly, the WCG, during its waiting phase, acted as the control for natural maturation for the IPG. Secondly, as the WCG underwent the intervention post WCG-A1, both groups could be merged to evaluate the intervention's impact by contrasting A1 with A2. This approach allowed for a greater sample size than solely utilizing the IPG's data. S1, S2, S3-A, and S4-A constituted single-target sessions, with 8 trials each. S3-B and S4-B, being optional, aimed to mitigate a potential ceiling effect, meaning if participants consistently succeeded at hitting targets with minimal prompts, further skill enhancement would be unattainable. Part B involved 5-minute double-target trials if participants achieved success in at least 4 consecutive trials with the first two prompts or completed all 8 trials successfully with the first four prompts in Part A. If not, Part B included another set of 8 single-target trials. During each session (for both Study I and Study II), an experimenter observed the entire session from a separate monitoring room.

C. *Results of Study I*

In Study I, we examined the relationship between SCR and RJA performance, incorporating data from both IPG and WCG. The correlation between session-SCR and hit rate ($r = 0.40$, $p < .01$), as well as between session-SCR and prompt level ($r = -0.36$, $p < .01$), were both moderate and statistically significant. These findings imply that children who were more engaged with the robot (as evidenced by higher SCR) exhibited improved performance in the robot-mediated intervention.

In Study I, across the 624 trials conducted, the mean trial-SCR for successful trials was 0.83 (SD = 0.21), while for unsuccessful trials it was 0.74 (SD = 0.31). The Wilcoxon rank sum test revealed a statistically significant difference between these values ($p < .0001$). Hence, a high trial-SCR signals a trial's success.

Given that SCR was not influenced by interaction protocols or task designs, it could serve as a metric to evaluate participant performance across varied study protocols executed with NORRIS.

To evaluate the impact of robot-mediated intervention

TABLE I
Timeline of Study I

<table>
<tr><td rowspan="2">IPG</td><td rowspan="2">A1</td><td rowspan="2">S1</td><td rowspan="2">S2</td><td colspan="2">S3</td><td colspan="2">S4</td><td rowspan="2">A2</td><td colspan="8" rowspan="2"></td></tr>
<tr><td>A</td><td>B</td><td>A</td><td>B</td></tr>
<tr><td rowspan="2">WCG</td><td rowspan="2">A0</td><td colspan="6" rowspan="2">Waiting period</td><td rowspan="2">A1</td><td rowspan="2">S1</td><td rowspan="2">S2</td><td colspan="2">S3</td><td colspan="2">S4</td><td rowspan="2">A2</td></tr>
<tr><td>A</td><td>B</td><td>A</td><td>B</td></tr>
</table>

IPG: immediate participation group
WCG: waitlist control group

versus natural development on hit rate and prompt level, we first checked for a significant interaction between time and group for each variable through a two-way ANOVA, finding none. Further analysis showed no significant improvements within groups (IPG-A1 to IPG-A2 or WCG-A0 to WCG-A1) or between groups (IPG and WCG) during the waiting phase, as indicated by the Wilcoxon rank sum test. Similarly, comparing pre-post RJA performances of IPG and WCG revealed no notable changes. An empirical penalty was applied in our analysis by assigning a value of 10 to the last prompt index in unsuccessful trials..

The reduction in hit rate and elevation in prompt level from WCG-A0 to WCG-A1 could indicate a decline in RJA abilities throughout the waiting period, potentially resulting from a diminished novelty effect and consequently lower engagement with NORRIS.

D. *Analysis of session-SCR in Study I*

To comprehend the variations in participants' engagement over different sessions, we examined the trends of average session-SCR for both IPG and WCG. Fig. 2 A) presents the baseline-normalized outcomes. The session-SCRs are depicted in dashed lines when combining Part-As and Part-Bs. Wilcoxon rank sum tests were utilized to evaluate the significance of differences observed:

1). T No significant alterations were detected between two consecutive sessions in either IPG or WCG, irrespective of the inclusion of Part-Bs.

2). The reductions from post-assessment sessions (A2) to pre-assessment sessions (A0, A1) in both groups were not statistically significant.

E. *Analysis of trial-SCR of sessions in Study I*

To delve into participant engagement with greater detail, we analyzed the trend of average trial-SCR in Study I. Fig. 2 B) separates the trial-SCRs into two segments: the "required phase" based on all mandatory trials in Study I (Part-Bs in S3 and S4 excluded); the "optional phase" from the voluntary trials in Part-Bs during S3 and S4. The data were normalized to baseline, revealing key trends:

1). From Trial 1, a marked decrease in trial-SCR was observed by Trial 4 to 8 ($p < .03$), indicating a diminishing engagement with NORRIS due to the repetitive nature of single-target trials in Study I**.**

2). Following a pause (in Part-Bs), the average trial-SCR saw an uptick at Trial 9 ($p = .38$) and Trial 10 ($p = .16$), though the increases were not statistically significant.

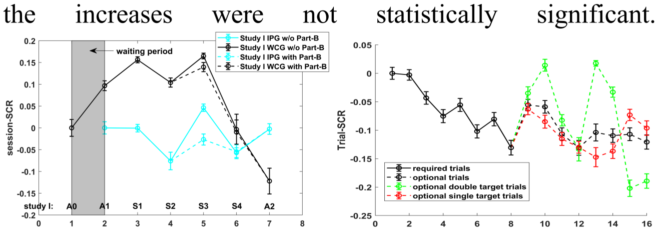

Fig. 2. **A)** Baseline normalized trends of average session-SCR of Study I. B) Baseline normalized trends of average trial-SCRs of Study I.

F. *Summary of findings from Study I*

Initially, we observed a decline in participants' engagement due to an excessive number of sessions. Consequently, we hypothesized that reducing the session count could sustain engagement over time. Furthermore, within a session, engagement dipped progressively with the repetition of monotonous single-target trials. Therefore, we proposed that diversifying trial designs could preserve engagement during a session. Lastly, introducing breaks appeared to rapidly enhance participants' engagement.

IV. Study II-Improved study with adaptive intervention protocol to enhance engagement

We revised the study design for a subsequent multi-session experiment, Study II. To accurately compare intervention outcomes between Study I and Study II, we maintained the initial assessment session structure with 8 single-target trials. The trials of intervention session were upgraded to be performance adaptive.

A. *Intervention timeline*

Study II's timeline is depicted in Table II. We decreased the intervention sessions from 4 to 2 in Study II, and correspondingly shortened the waiting period for the WCG. In Study II, the mean gap between IPG-A1 and IPG-A2 was 3.21 weeks (SD = 0.77), and between WCG-A0 and WCG-A1 was 3.21 weeks (SD = 0.32), with no notable difference observed between the two ($p = 1.0$).

B. *Performance-based adaptive intervention session*

Table III outlines the revised protocol for the performance-based adaptive intervention session designed to sustain participant engagement. The indices for single-, double-, and triple-target trials remained consistent between Study I and II.

In Table III, six potential pathways are identified, with participants in Study II undergoing three specific ones: S-D-T, S-S-D, and S-D-D.

C. *Participant Characteristics*

The inclusion criteria for participants were identical to those

TABLE II
TIMELINE OF STUDY II

IPG	A1	S1	S2	A2			
WCG	A0	Waiting period		A1	S1	S2	A2

Labels similar to Table I

TABLE III
PERFORMANCE-BASED ADAPTIVE INTERVENTION SESSION PROTOCOL

<table>
<tr><th>Stage 1</th><th>Criteria 1</th><th>Stage 2</th><th>Criteria 2</th><th>Stage 3</th></tr>
<tr><td rowspan="6">S</td><td rowspan="3"># of target hit ≥ 2</td><td>D</td><td># of target hit ≥ 2</td><td>T</td></tr>
<tr><td>D</td><td># of target hit = 1</td><td>D</td></tr>
<tr><td>D</td><td># of target hit = 0</td><td>S</td></tr>
<tr><td rowspan="3"># of target hit < 2</td><td>S</td><td># of target hit ≥ 2</td><td>D</td></tr>
<tr><td>S</td><td># of target hit = 1</td><td>S</td></tr>
<tr><td>S</td><td># of target hit = 0</td><td>N/A</td></tr>
</table>

S: 4 single-target trials; D: 3 double-target trials;T: 2 triple-target trials

in Study I (refer to section III.A). Twelve participants (8 males and 4 females, aged 2.03-4.64 years; average age = 2.62 years, standard deviation = 0.33) were enrolled and randomized into the IPG and WCG, showing no significant differences in age at recruitment (p = .62), ADOS-2 total raw score (p = .16), SRS-2 total raw score (p = .93), SRS-2 T-score (p = 1.0), SCQ current total score (p = .27), MSEL ELC (p = .29), and months between ADOS assessment and enrollment (p = .24). The study received approval from the Vanderbilt University Institutional Review Board under IRB Protocol 170022. Informed consent was obtained from all participating families before enrollment by principal study staff.

D. *RJA performance analysis*

Fig. 3 A) and B) illustrate the outcomes of robot-mediated intervention versus natural maturation in Study II. Insights from Fig. 3 and Table IV include:

1). The hit rate, where higher scores are preferable, showed significant improvement over time, indicated by a large effect size. A significant difference was also observed between groups, with a medium effect size，showing the intervention group outperformed the non-intervention group. Additionally, a significant time-group interaction was noted, also with a large effect size.

2). Regarding prompt level, where lower scores are advantageous, a significant reduction over time was evident, marked by a large effect size. Nonetheless, no significant differences were found concerning the group factor or the time-group interaction, both displaying small effect sizes.

Table V presents the pre-post intervention comparison, amalgamating IPG-S and WCG-S. In this analysis, both hit rate and prompt level exhibited significant enhancements, each with large effect sizes.

Collectively, these findings indicate that participants' RJA performance saw notable improvements following the NORRIS intervention, thanks to the refined intervention schedule and protocol.

E. *Validation of SCR in Study II*

In Study II, the association between session-SCR and hit rate (r = 0.35, p = .0088), as well as the association between session-SCR and prompt level (r = -0.47, p = .0004), were both moderate and statistically significant.

The mean trial-SCRs for successful and unsuccessful attempts were 0.80 (SD = 0.26) and 0.64 (SD = 0.36), respectively (p < .0001). Utilizing all 467 trials from Study II, the relationship between trial prompt levels (with unsuccessful trial index adjusted to 10) and trial-SCRs was examined. Findings reveal a moderate and significant correlation (r = -0.31, p < .0001), endorsing SCR as a measure of engagement to reliably estimate one's RJA capabilities.

The correlations and comparative analyses of differences above exhibited similarities with those from Study I (refer to section IV). These findings validated the use of SCR as an effective metric across various studies and task frameworks.

F. *Analysis of session-SCR & trial-SCR in Study II*

In our findings, as shown in Figure 4 A), there was a significant increase in session-SCR from IPG-S2 to IPG-A2 (p = .038), with no significant changes in other sessions. In the second study, the performance-based adaptive intervention protocol resulted in six possible paths, with participants following three specific pathways (S-D-T for 17 sessions, S-S-D for 5 sessions, and S-D-D for 2 sessions) across 24 sessions. Figure 4 B) illustrates the average trial-SCR trends in Study II, normalized against baseline data. For consistency, the first 8 trials of intervention sessions were combined with assessment sessions for average calculations, depicted in black in Figure 4 B). Unlike Study I, where trial-SCR significantly dropped after the fourth trial, Study II showed no significant differences in trial-SCR across all trials..

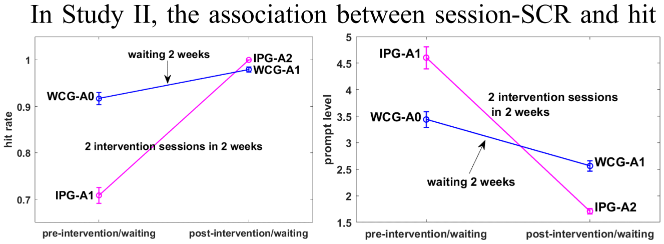

FIG. 3. PERFORMANCE CHANGES IN STUDY II
A). HIT RATE COMPARISON. B) PROMPT LEVEL COMPARISON

TABLE IV
(A) 2-WAY ANOVA TEST (TIME × GROUP) ON PERFORMANCE COMPARING PRE-POST IPG-S AND WCG-WP IN STUDY II. (P < .05 IN BOLD)

Measurement	*Wilcoxon rank sum test between IPG-A1 and WCG-A0, IPG-A2 and WCG-A1*	*Wilcoxon signed-rank test on IPG-A1 and IPG-A2, WCG-A0 and WCG-A1*
hit rate	**.0476 (1.38)**, 1 (0.57)	**.0313 (2.42)**, .5(0.64)
prompt level	.4156 (0.64), .0952(1.10)	**.0313 (1.92)**, .125(0.69)

V. DISCUSSION AND CONCLUSION

To our knowledge, this article stands as one of the initial efforts to automate the quantification of an engagement indicator and its variation's effect on intervention outcomes. Our findings suggest that consistent engagement is crucial for the success of robot-mediated interventions for young children with autism. Moreover, this research demonstrates how the intervention timeline and protocol influence engagement, as evidenced by comparing two multi-session studies (Studies I and II) utilizing a robot-mediated RJA system, NORRIS.

We introduced the System Capture Ratio (SCR) for measuring engagement via eye gaze in Study I, noting diminished engagement with a repetitive interaction protocol.

TABLE V
(A) PERFORMANCE COMPARING PRE-POST IPG-S+WCG-S IN STUDY II. (P < .05 IN BOLD); WILCOXON RANK SUM TEST

Measurement	A1 mean (SD)	A2 mean (SD)	*p*-value (Cohen's D) of time
hit rate	0.8438(0.19)	0.9688(0.08)	**.0352 (0.857)**
prompt level	3.5833(1.88)	2.1458(1.20)	**.0241 (0.911)**

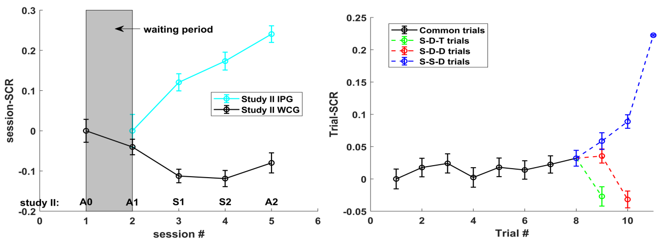

Fig. 4. **A)** Baseline normalized trends of average session-SCR of II. B) Baseline normalized trial-SCR trends in study II

By analyzing SCR, we proposed fewer sessions and a performance-adaptive intervention could improve RJA (Responsive Joint Attention) performance. This led to Study II, which confirmed increased engagement and significant RJA improvements, supporting our theory. However, the study has limitations, such as the requirement for participants to remain seated and the robot's inability to understand participant communication beyond visual cues. Future improvements should focus on integrating speech recognition, facial expression analysis, and body language interpretation to enhance the efficacy of robot-mediated interventions like NORRIS, enabling a deeper understanding of participant responses.

The study's limited sample sizes, while yielding promising effect sizes, necessitate larger future studies to fully evaluate NORRIS's effectiveness in real-world skill transfer across diverse autism profiles, settings, and over time. Expanding sample sizes will aid in optimizing intervention parameters and customizing the system for specific autism subgroups. Additionally, investigating the influence of cognitive, language, and attention skills on system interaction is recommended. While NORRIS focuses on RJA training, addressing the wider spectrum of social-communication difficulties in children with autism remains a gap. Nonetheless, this research contributes significantly to robot-mediated interventions for autism by demonstrating through randomized control trials and statistical analysis that the intervention's structure influences engagement and that maintaining this engagement is key to enhancing performance.

References

[1] M. J. Maenner, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2020," *MMWR. Surveillance Summaries*, vol. 72, 2023.

[2] C. Lord, E. H. Cook, B. L. Leventhal, and D. G. Amaral, "Autism spectrum disorders," *Autism: The Science of Mental Health*, vol. 28, no. 2, p. 217, 2013.

[3] K. Gillespie-Lynch *et al.*, "Early childhood predictors of the social competence of adults with autism," *J Autism Dev Disord*, vol. 42, no. 2, pp. 161–174, 2012.

[4] H. H. Schertz, S. L. Odom, K. M. Baggett, and J. H. Sideris, "Effects of joint attention mediated learning for toddlers with autism spectrum disorders: An initial randomized controlled study," *Early Child Res Q*, vol. 28, no. 2, pp. 249–258, 2013.

[5] Z. E. Warren and W. L. Stone, "Best practices: Early diagnosis and psychological assessment," *Autism spectrum disorders*, pp. 1271–1282, 2011.

[6] B. Scassellati, H. Admoni, and M. Matarić, "Robots for use in autism research," *Annu Rev Biomed Eng*, vol. 14, pp. 275–294, 2012.

[7] D. Feil-Seifer and M. J. Matarić, "Socially assistive robotics," *IEEE Robot Autom Mag*, vol. 18, no. 1, pp. 24–31, 2011.

[8] Z. Zheng, S. Das, E. M. Young, A. Swanson, Z. Warren, and N. Sarkar, "Autonomous robot-mediated imitation learning for children with autism," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, IEEE, 2014, pp. 2707–2712.

[9] A. Billard, B. Robins, J. Nadel, and K. Dautenhahn, "Building Robota, a mini-humanoid robot for the rehabilitation of children with autism," *Assistive Technology*, vol. 19, no. 1, pp. 37–49, 2007.

[10] G. Nie *et al.*, "Predicting Response to Joint Attention Performance in Human-Human Interaction Based on Human-Robot Interaction for Young Children with Autism Spectrum Disorder," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE, 2018, pp. 1–4.

[11] G. Nie, A. Ullal, A. R. Swanson, A. S. Weitauf, Z. E. Warren, and N. Sarkar, "Design of an Intelligent and Immersive System to Facilitate the Social Interaction Between Caregivers and Young Children with Autism," in *International Conference on Human-Computer Interaction*, Springer, 2019, pp. 123–132.

[12] E. S. Kim, R. Paul, F. Shic, and B. Scassellati, "Bridging the research gap: Making HRI useful to individuals with autism," *J Hum Robot Interact*, vol. 1, no. 1, pp. 26–54, 2012.

[13] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *Int J Soc Robot*, vol. 5, no. 2, pp. 291–308, 2013.

[14] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[15] M. Tielman, M. Neerincx, J. Meyer, and R. Looije, *Adaptive emotional expression in robot-child interaction*. 2014. doi: 10.1145/2559636.2559663.

[16] C. Breazeal, "Emotion and Sociable Humanoid Robots," *Int J Hum Comput Stud*, vol. 59, pp. 119–155, Jul. 2003, doi: 10.1016/S1071-5819(03)00018-1.

[17] D. Riby and P. Hancock, "Looking at movies and cartoons: Eye-tracking evidence from Williams syndrome and autism," *J Intellect Disabil Res*, vol. 53, pp. 169–181, Mar. 2009, doi: 10.1111/j.1365-2788.2008.01142.x.

[18] S. M. Anzalone *et al.*, "Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment," *Pattern Recognit Lett*, vol. 118, pp. 42–50, Feb. 2019, doi: 10.1016/j.patrec.2018.03.007.

[19] P. Ravindra, S. De Silva, K. Tadano, A. Saito, S. G. Lambacher, and M. Higashi, "Therapeutic-assisted robot for children with autism," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2009, pp. 3561–3567.

[20] B. R. Schadenberg *et al.*, "Predictable Robots for Autistic Children—Variance in Robot Behaviour, Idiosyncrasies in Autistic Children's Characteristics, and Child–Robot Engagement," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 28, no. 5, pp. 1–42, 2021.

[21] C. Clabaugh *et al.*, "Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders," *Front Robot AI*, vol. 6, p. 110, 2019.

[22] B. Scassellati *et al.*, "Improving social skills in children with ASD using a long-term, in-home social robot," *Sci Robot*, vol. 3, no. 21, p. eaat7544, 2018.

[23] Z. Zheng, H. Zhao, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar, "Design, Development, and Evaluation of a Noninvasive Autonomous Robot-Mediated Joint Attention Intervention System for Young Children With ASD," *IEEE Trans Hum Mach Syst*, vol. 48, no. 2, pp. 125–135, 2018.

[24] J. R. Pruette, "Autism diagnostic observation schedule-2 (ADOS-2)," *Google Scholar*, 2013.

[25] L. Moshkina and R. C. Arkin, "Human perspective on affective robotic behavior: a longitudinal study.," in *IROS*, Citeseer, 2005, pp. 1444–1451.