An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children With Autism Spectrum Disorder

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

***Abstract*— Autism Spectrum Disorder (ASD) affects 1 in 54 children in the United States. A core social communication skill negatively impacted by ASD is joint attention (JA), which influences the development of language, cognitive, and social skills from infancy onward. Although several technology-based JA studies have shown potential, they primarily focus on response to joint attention (RJA). The other important component of JA, the initiation of joint attention (IJA), has received less attention from a technology-based intervention perspective. In this work, we present an immersive Computer-mediated Caregiver-Child Interaction (C3I) system to help children with ASD practice IJA skills. C3I is a novel computerized intervention system that integrates a caregiver in the teaching loop, thereby preserving the advantages of both human and computer-administered intervention. A feasibility study with 6 dyads (caregiver-child with ASD) was conducted. A near significant increase with medium effect size on IJA performance was observed. Meanwhile, physiology-based stress analysis showed that C3I did not increase stress of the caregivers over the course of the study. To the best of our knowledge, this is the first autonomous system designed for teaching IJA skills to children with ASD incorporating caregivers within the loop to enhance the potential for generalization in real-world.**

Index Terms— **Human computer interaction, autism spectrum disorder, initiation of joint attention, caregiver-child interaction.**

Manuscript received December 14, 2020; revised March 22, 2021; accepted April 29, 2021. Date of publication May 4, 2021; date of current version May 18, 2021. The work of Guangtao Nie was supported by the U.S. NIH under Grant 5R33MH103518-04 and Grant 1R21MH111567. *(Corresponding author: Guangtao Nie.)*

Guangtao Nie and Akshith Ullal are with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37212 USA (e-mail: guangtao.nie@vanderbilt.edu; akshith.ullal@vanderbilt.edu).

Zhi Zheng is with the Department of Biomedical Engineering, Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: zhzbme@rit.edu).

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

Amy S. Weitlauf and Zachary E. Warren are with the Vanderbilt University Medical Center, Department of Pediatrics, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Nashville, TN 37212 USA (e-mail: amy.s.weitlauf@vumc.org; zachary.e.warren@vumc.org).

Nilanjan Sarkar is with the Department of Mechanical Engineering and the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37212 USA (e-mail: nilanjan.sarkar@vanderbilt.edu).

Digital Object Identifier 10.1109/TNSRE.2021.3077480

I. INTRODUCTION

AUTISM Spectrum Disorders (ASD) are characterized by difficulties in social communication and repetitive or restricted behaviors and interests [1]. The most recent prevalence estimates indicate that 1 in 54 children in the US is estimated has ASD [2]. The estimated annual cost of caring for children with ASD in the United States, in terms of dollars as well as lost caregiver productivity, was $268 billion in 2015 and is expected to be $461 billion in 2025 [3]. Thus, there is tremendous ongoing need for systems that support care for children and families.

Assessment and behavioral intervention in early childhood have shown promise to help alleviate heterogeneous symptoms in young children with ASD [4]. A core social communication symptom of early childhood that is commonly targeted during treatment is deficits in joint attention (*JA*), which is defined as shared attention with others through pointing [5], [6], showing, and coordinated looks between objects and people [7]. It is related to the development of language, cognition, social skills and behavioral competence [8]. Initiation of joint attention (*IJA*) and response to joint attention (*RJA*) are the initiative and receptive form of joint attention, respectively. Specifically, IJA of young children with ASD is defined as the ability to spontaneously create or indicate a shared object of reference by the use of gestures, such as eye contact. Literature suggests that joint attention bids occur less frequently among children with ASD, and the deficit in *IJA* may reflect differences in inherent social motivation [9]. Meanwhile, eye-tracking studies have shown that *IJA* contains more discriminative information to distinguish ASD from TD children, as their gaze patterns are different when they are expected to initiate joint attention and not when they respond to joint attention [10].

Traditional assessment of joint attention skill and intervention necessarily rely heavily on behavioral therapists [11], [12]. However, the accessibility to these professional services are limited due to a lack of trained professionals and high expense [13]. By contrast, caregivers interact with their children with ASD every day, but without explicit interaction protocols, effective training methods, and consistent measurement of social skill performance. Additionally, although caregiver-child interaction during early childhood is important to later developmental outcomes [14], caregivers of children with ASD have reported experiencing higher stress during this interaction compared to the caregivers of typically developing (TD) children [15], [16]. Severity of symptoms associated with ASD, such as language and communication difficulties, have been found to be highly correlated to caregivers' stress [17].

During the past decade, researchers have designed human-robot interaction (HRI) and human-computer interaction (HCI) systems to help young children with ASD develop their fundamental social skills. Many children with ASD are attracted to computer and robotic systems, which can be designed to provide uniform intervention delivery and objective, quantitative performance tracking [5], [6], [16], [17]. Several autonomous systems have been designed for training *RJA* skill for young children with ASD by several researchers including the authors [5], [19]–[23]. *However, to the best of our knowledge, there has been no autonomous system reported that teaches IJA skills for children with ASD, incorporating parents within the loop, within which IJA performance can be consistently and automatically tracked (i.e., no need for video coding of the sessions)*. As such, we propose a computer-mediated caregiver-child interaction system (C3I) in this work to fill this gap. There are other technology-based IJA studies that are relevant for the current work, all of which differ significantly from ours: In a study [20] Simut *et al.* compared differences in *IJA* performance between children with ASD interacting with an experimenter or interacting with the robot [24]. In this work, they manually coded the performance under these two conditions. Boccanfuso *et al.* reported significantly increased joint attention frequency for children with ASD during interaction with CHARLIE the robot [25]. Their work on joint attention seemed to be a mixture of *RJA* and *IJA* and they measured *JA* occurrence frequency instead of social skill performance. In other works, virtual agents or embodied robots have been adopted in various studies to investigate the psychological or cognitive mechanics of *IJA*, but their participants were typically developing (TD) children and the studies were not for teaching social skills [26]–[29].

Compared to the training of *RJA* where a child is expected to respond to a joint attention bid (e.g., turn to follow someone's point) and which can be automated relatively easily, the training of *IJA* for children with ASD is more difficult because it requires creating scenarios compelling enough to prompt a child initiate a joint attention bid [30].

It is possible that children will be more likely to engage in *IJA* when in the presence of a familiar caregiver, such as a parent, as opposed to a purely technical system (i.e., computer or robot alone). Moreover, incorporating caregivers within such automated systems might help reduce the so-called "isolation effect," which happens when children learn social skills from a system lacking of real life components and thus reduces their social interest in real life interaction after system intervention [31]. Incorporating caregivers as part of teaching could increase the likelihood that children will show those skills with those same caregivers outside of the system itself, increasing real-world applicability. In addition, the autonomous part of the system would not increase caregiver stress during such teaching interactions by reducing their workload. Nonetheless, the involvement of caregiver within such a system presents an interesting technological challenge, as it will unavoidably complicate the system's interaction protocol and system design and has not been reported in the literature in the context of ASD intervention.

Traditionally, computer-mediated human-human interaction systems have mostly been developed for education, business, media, entertainment, and social communication [32]–[34]. Recently, a few computer-mediated systems have been designed for caregiver-child interaction, but not in the context of ASD intervention. Tscholl *et al.* investigated the effectiveness of an interactive, immersive, and full-body simulation, using augmented reality, to empower children's science learning process [35]. Children were instructed to move and use their bodies to enact their understanding of a scientific system, during which caregivers discussed game strategies with them. Results indicated that productive forms of social interaction were observed during this process. Kucirkova *et al.* investigated the interaction between a 33-month-old daughter and her mother, when they shared a self-centered, audio-visual 'iPad story' [36]. Analysis showed that harmonic and smooth interaction was observed within such a story sharing context. A Technology-Enhanced Storytelling (TES) activity, Jeffy's Journey, was developed to support caregiver–child interaction and vocabulary development in preschool children through real-time visual, auditory, and textual prompts on a tablet computer [37]. It showed that more time spent on the story and usage of prompts were associated with higher quality of caregiver-child interaction. They also demonstrated the effectiveness of TES on children's vocabulary development. In recent years, collaborative virtual environments [38], [39] have been explored, where children with ASD learn skills though cooperating with other children to accomplish tasks. However, these systems were not designed for caregiver-child interaction.

C3I is derived from and significantly expanded on our previously developed platform called Autonomous Social Orienting Training System (ASOTS) [6]. ASOTS used pre-recorded video and audio clips to provide social cues (i.e., name call) and provided a closed-loop *Response to Name* (*RTN)* learning protocol to manage the interaction between the children and the system. ASOTS did not include caregivers; it was a fully autonomous interaction with the computer system. In the current work, C3I expanded ASOTS' closed-loop interaction to accommodate *IJA* training and included the caregiver into the interaction loop. C3I allows the caregiver to interact with the child in a flexible manner where a caregiver can decide how much or how little interaction one desires. More importantly, C3I created a framework to include caregivers into the loop such that social response registration of children could be made by the caregivers, which separates it from other existing systems in this field. C3I helped to evoke, guide and reinforce children's social behaviors, as well as quantitatively measured aspects of real-time behaviors of both the caregiver and the child.

C3I's interaction with both the children and the caregivers is designed to be easy to understand and follow. In particular, the caregiver could control each interaction step. In C3I, caregivers are integrally embedded to provide human inputs, such as recording the response and inferring children's mental states, such as frustration. Timing children's social responsiveness is critical in caregiver-child interaction, as children are more likely to respond synchronously if their communicative signals can be recognized, understood, and responded appropriately [40]. In this work, social responsiveness of children is defined as the response of children to a paused video clip. Caregivers' timing of children's social responsiveness can be used as ground truth as it is both personal and contextual [41]. By contrast, timing the social responsiveness of children is a challenging task for existing technology; with regard to *IJA*, it involves recognition of children's body movements, gestures, and facial expressions and their temporal coordination [41]. Additionally, children's mental states (e.g., frustration) reported by caregivers can help C3I tune system parameters to provide adaptive training for the children. As the severity of symptoms vary among children with ASD, some may find it too difficult to follow the prompts of C3I and get frustrated. Caregivers' report of children's frustration can help adapt C3I to better facilitate their interaction.

The main contribution of this work is the design and development of the novel C3I system that includes caregivers in the HCI-based intervention loop to teach IJA to children with ASD. C3I, as it is presently designed, can provide intervention that promotes *IJA* – a foundational social interaction building block, and also provide both quantitative system measures and qualitative caregiver measures. It thus combines the benefits of automated technology-based and human-based intervention. We also provide the feasibility of the C3I system through a small study with children with ASD. This study is intended to demonstrate the usability of C3I. This work is a substantial extension of our conference paper [42] in terms of detailed discussion on system design and analyses of caregivers' physiology-based stress and survey results.

In this work, we present an immersive computer-mediated system designed for caregiver-child interaction, with an explicit interaction protocol, aiming to improve *IJA* skill for young children with ASD. By immersive we mean that the caregiver is integrated within the closed-loop interaction of the human-computer system. Moreover, we anticipated that within this immersive system environment, an explicitly defined interaction protocol, an intuitive system assistance to evoke, guide and reinforce children's social behaviors, and an consistent, quantitative tracking of *IJA* performance would promote children's social skills and would not increase the stress of caregivers during interaction.

The remainder of this paper is organized as follows: system design details are presented in Section II; a feasibility study is presented in Section III; Section IV presents and discusses the experimental results, and Section V summarizes the contributions and discusses future work.

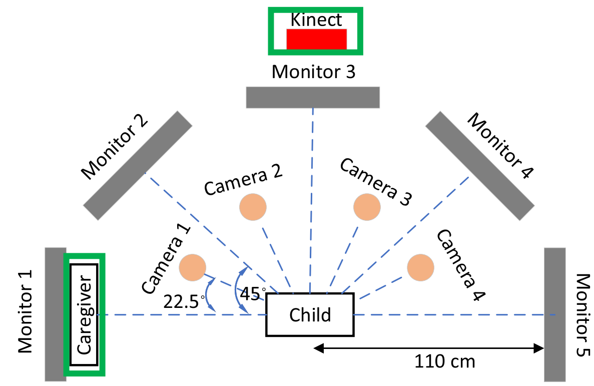

Fig. 1. ASOTS and C3I System setup (top view, components that are not green rectangles are also components of ASOTS).

II. System Development

As C3I expanded the ASOTS platform [6], ASOTS is briefly introduced in section II. The ASOTS system setup is displayed in Fig. 1. Note that the rectangles that are not green are ASOTS components.

To enable the social orientation training capacity of ASOTS, two types of *RTN* tasks were defined: one using only audio clips and the other using video (with embedded audio) clips, both of which were pre-recorded by a human therapist calling the child's name. ASOTS started a trial by playing an audio or video clip on a target monitor and rewarded children with a video clip as soon as a gaze point was detected falling onto the target monitor. For detailed information, specifications, and results of ASOTS, please refer to [6].

C3I expanded ASOTS architecture in three major ways. First, it extended the interaction from purely autonomous child-system interaction to child-caregiver-system interaction, where a caregiver was seamlessly integrated within the system interaction. Second, C3I was designed to train *IJA* tasks as opposed to *RTN* tasks for children with ASD, which were relatively more difficult to design. Third, C3I was capable of measuring the stress of the caregivers from their physiological responses in order to both assess the impact of the interaction on caregivers and how to improve C3I to reduce stress in the future.

A. C3I Design Steps

The functionalities of C3I were created using the following design steps:

1) In order to organize and coordinate peripheral modules, as well as provide uniform and consistent rule-based training, the central controller of C3I was designed as a closed-loop Finite State Machine (FSM).

2) To track the performance of children, a camera array consisting of four webcams encircled the 180° space in front of the children that tracked their head orientation.

3) A monitor and a speaker array were used to display the social targets, social stimuli, and reward videos; and transfer a child's attention by straightforward and intuitive non-social cues. As the gaze direction of child was the fundamental behavioral signal that guided the interaction, a non-social cue

was designed as a 3D animation ball which bounced from the current gaze direction (see Fig. 3) of the child to the target monitor. The bouncing trajectory was a sinusoidal wave with adjustable bouncing speed. When a bouncing ball was triggered, a simulated bouncing sound clip was also played by the 5.1 surround speakers. It was shown to be effective to guide a child's attention within this environment [5].

4) Caregivers wore a physiological sensor (the E4 wristband [43]) to measure their signals for physiology-based stress analyses using affective computing.

5) A new tablet app was designed for caregivers as an I/O interface to access the system information and input human decisions. Caregivers could use the app simply by following real-time instructions.

6) Every state of the central controller needed an input or confirmation from the caregiver to trigger the transition between the interaction states.

B. Overview of System Architecture

Overall, C3I functions in the following manner: C3I displayed short video clips on each of the possible target monitors to orient the child to the experimental space. It then showed a longer video clip to distract and keep the child's attention away from the caregiver. When the child was distracted, the caregiver pressed a button on the tablet and told the system to abruptly pause the video. At this point, the expectation was that the child would then look back at the caregiver and initiate joint attention. If the child did not look, at either the caregiver or the target monitor, the caregiver could press another button to alert the system. The system then displayed a non-social audio-visual cue (a bouncing ball with sound effects). This bouncing ball started at whichever point the child was looking, as detected by the gaze detection system, and then moved across the monitor array, guiding the child's attention in the direction of the caregiver or the target monitor.

The setup and architecture of C3I are shown in Fig. 1 and Fig. 2, respectively. In Fig. 2, the blue modules were adopted directly from ASOTS, such as the gaze tracking subsystem. Orange rectangles are modules upgraded from ASOTS, with different content or functionalities, such as the monitor and speaker array and the central controller. Green modules are newly added modules, including the body tracker, the caregiver tablet, and the E4 sensor, due to new interaction protocol and task design.

Fig. 3 shows a snapshot of a session with C31. C3I integrates a caregiver within the closed-loop interaction of ASOTS and thus allows a flexible interaction among the children, caregivers, and the autonomous components of the system. Compared to ASOTS, the differences in module components are: 1) a caregiver was seated under Monitor 1, using a tablet app as an I/O interface and wearing a E4 wristband to measure his/her physiological signals for offline stress analysis; and 2) a Kinect sensor was put right behind Monitor 3 as a body tracker to record behavioral data (i.e., gesture, skeleton, facial expression) of the children.

The information flow within C3I is presented in Fig. 4. The system provided interaction information for the caregivers and measured their physiological responses in terms of galvanic skin response and heart rate. The caregivers, in turn, reported the timing of social responsiveness and the mental state (e.g., frustration) of the children based on their observation, experience, and insight. On the other hand, C3I provided children with task components (e.g., target display and reward video display), non-social cues and tracked their real time behavioral data in terms of gaze, gesture and facial expression.

All modules were implemented in Unity3D [44] with C#.

C. Peripheral Modules

The four peripheral modules communicated with the central controller through TCP/IP asynchronous sockets.

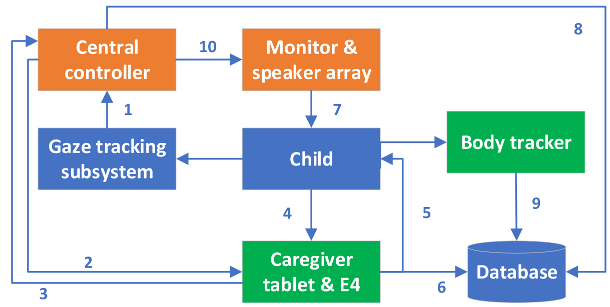

Fig. 2. ASOTS & C3I System architecture (Green rectangles label newly added modules compared to ASOTS. Orange rectangles are redesigned modules based on the central controller of ASOTS). 1. Gaze; 2. System information; 3. Caregiver heuristics; 4. Mental state; 5. Social cues; 6. Physiological signal; 7. Non-social cues, videos and audios; 8. Collective caregiver heuristics, Child gaze and System logging; 9. RGBD images, gestures and facial information; 10. Program instruction.

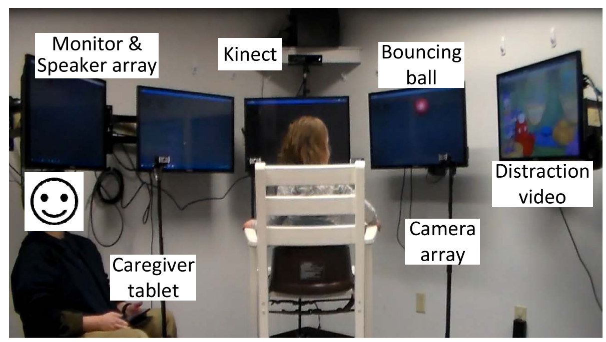

Fig. 3. Snapshot of the C3I environment. This is a video snapshot of a session in Distraction state, where distraction video was played on Monitor 5 but the child was looking at somewhere else. A bouncing ball was triggered in this scenario by a caregiver to transfer the child's attention onto Distraction video.

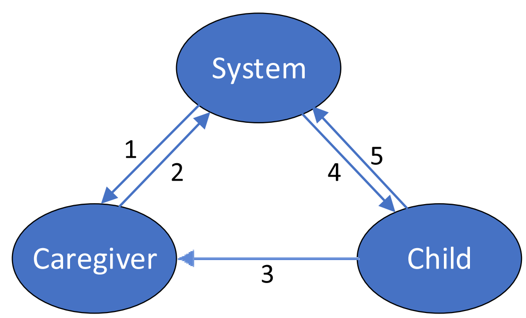

Fig. 4. C3I information flow (1. System information; 2. Decisions & Physiological signal; 3. Social responsiveness & Mental state; 4. Non-social cues, Target & Reward video; 5. Gaze, Skeleton, gesture & facial information).

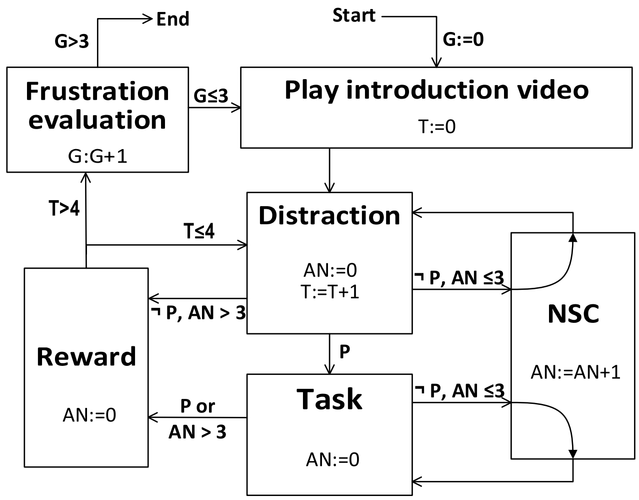

Fig. 5. Finite sate machine (FSM) of the central controller.
Annotation:
P ∈ {*true, false*}: The caregiver's decision. In Distraction state, P is caregiver confirming the child watching distraction video and in Task state, P is caregiver confirming the child looking back at him/her.
G ∈ *N*: Trial-group number. x ∈ {2,3,4,5}: a random integer number generator
T ∈ *N*: Trial number in a trial-group.
Timer ∈ *R*. Target ∈ *N*: The desired direction (indicated by the monitor #)
AN ∈ *N*: number of active questions for caregiver.
Initialization: ¬P, G = 0, T = 0, Timer = 0, AN = 0
Explanation:
Distraction state: Distraction video playing
Task: Caregiver paused the distraction video and waited for response
NSC: The bouncing ball was triggered either toward distraction video (Distraction) or caregiver (Task)
Reward: Distraction video resumed playing on Monitor 1
Frustration evaluation: Caregiver evaluated the level of frustration of the child during the trial-group.

1) Central Controller: In order to coordinate peripheral modules as well as deliver uniform and consistent intervention, the central controller was designed to be a Finite State Machine (FSM), as depicted in Fig. 5. To explicitly and accurately model different states of C3I, 6 states were designed in the FSM: Play introduction video (PIV), Distraction, Task, Non-social cues (NSC), Reward, and Frustration evaluation (FE). In PIV, an introduction video clip was played to inform the child about all monitors. Next, in the Distraction state, a muted distraction video was played on a randomly selected monitor to distract the child's attention away from the caregiver. Monitor 1 was excluded from the selection as the caregiver was seated underneath it. Then, when the caregiver determined the child was distracted by the video and decided to start a trial, the caregiver recorded the timing on the tablet (by clicking a button) and was instructed to pause the distraction video. The duration between the time the caregiver determined the child was distracted and the time the caregiver decided to trigger a trial was called *distraction duration* and it was designed to be controlled by the caregiver in order to vary the trial difficulty. Lastly, the caregiver needed to confirm the response of the child to trigger a reward video on Monitor 1.

During the states of Distraction and Task, if the desired behavior was not observed by the caregiver within a 7 second time window, which was decided based on discussion with behavioral therapists, the central controller would switch into the NSC state, where the distraction video might be unmuted, or bouncing balls might be triggered. When a trial-group of 5 *IJA* trials was completed, the central controller entered the state FE, where a question on trial difficulty (frustration) was evaluated by the caregiver to adapt a parameter of C3I: if judged too easy, bouncing speed of the animation ball would increase by 10%; if too hard, bouncing speed would decrease by 10% in the next trial-group; otherwise there would be no change.

2) Monitor and Speaker Array: The Monitor and Speaker array consisted of five monitors (70 cm × 43 cm in size) with five surround sound speakers mounted onto the back of each monitor, which were used to present picture/audio/video during interaction, provide non-social cue (bouncing ball) to transfer the child's attention, and display reward videos. This combination setup created an immersive audiovisual environment. As shown in Fig. 3, monitors encircled the child along a semi-circle of 110 cm radius.

3) Gaze Tracking Subsystem: The gaze tracking subsystem was used to determine the starting position of the bouncing ball, and record the trajectory of children's gaze during interaction.

The gaze tracking subsystem was a camera array consisting of four 720p Logitech C920 web cameras, placed on a semi-circle of 70 cm around the child to cover the entire 180° horizontal space in front of the child to track one's frontal face. The frontal face images were processed by a Supervised Descent Algorithm [45], predicting the 3D head pose (i.e., yaw, pitch, and roll angles) on each frame, which was then projected to estimate the eye gaze direction [6]. This subsystem ran at 15 fps, which was adequate for real-time gaze tracking. In the environment shown in Fig. 3, when a child's gaze fell within the pitch range of [−30°, 21°], and its yaw angle fell within [107°, 73°], [62°, 28°], [17°, −17°], [−28°, −62°], [−73°, −107°], C3I registered the child as watching monitors 1 to 5, respectively. Caregivers were seated underneath Monitor 1 and C3I registered children's gaze as looking at caregivers when their gaze approximately falling on Monitor 1. The registration results were only used for initializing the starting position of the bouncing ball. In other situations, they only served as assistive references for caregivers to make decisions.

4) Body Tracker: In order to track the skeletal and facial information of the child, a Kinect v2 [46] was put right behind Monitor 3, at a height of 1.8 m, tilted towards the child. Kinect v2 provided RGBD data through a high-resolution RGB camera (1920 × 1080 p) and an infrared depth sensor (512 × 512 p). It could detect up to 6 people simultaneously at the speed of 30 fps. With its proprietary SDK, the detection results of facial expression (11 descriptors) and skeletal joint coordinates (25 key points) were recorded. To ensure that the caregiver would not be incorrectly detected as the child, the detection range was limited to the participant chair. All the RGBD, facial and skeletal data were written into a SQL database at 15 fps. Although the body tracking data wasn't used in this version of C3I, these data provided quantitative information about the gestures associated with *IJA* and can be correlated with caregivers' assessment in the future.

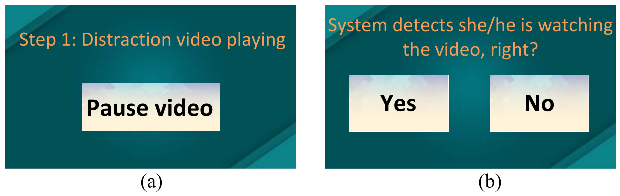

Fig. 6. Caregiver tablet app screen shots: (a) A button in the active mode; (b) A confirmation question in the passive mode.

5) Caregiver Tablet App: The caregiver tablet app was designed for caregivers to receive information from C3I and input their decisions into C3I with regard to the timing of social responsiveness and mental states of children. The app was designed to be easy to use simply by following real-time instructions. It allowed the caregiver to observe the child for as much time as possible and proceed with interaction whenever the desired behavior from the child was observed.

The app had both an active mode (Fig. 6-(a)) and a passive mode (Fig. 6-(b)). These two modes functioned equivalently, i.e., if choosing "Yes" in Fig. 6-(b), its effect was the same as pressing "Pause video" in Fig.6-(a); and the effect of choosing "No" in Fig. 6-(b) was the same as idle state in Fig. 6-(a).

The active mode was designed for the caregiver to actively proceed with the interaction during the response time window. The passive mode was designed to remind the caregiver to make a decision when response time window exceeded the designed limit but there was no input from the caregiver. For both the Distraction and Task states of the FSM in Fig. 5, the active mode came first and lasted for 7 seconds (i.e., response time window) for the caregiver to actively input that the desired behavior was observed; if there was no input in the 7 second response time window, system would switch into passive mode with a tablet vibration, where a question on the desired behavior would be asked for the caregiver as a reminder. Questions were displayed in two forms based on whether the desired behavior was detected by the gaze tracking subsystem (confirmation) or not (reminder).

D. Interaction Protocol

In ASOTS, there was no caregiver involved. By contrast, in the interaction protocol of C3I, caregivers controlled the interaction flow and children followed prompts from C3I. The flowchart in Fig. 7 depicts the flow of a trial.

Before the start of a trial, a muted distraction video clip was played on any monitor other than Monitor 1 to distract a child's attention away from the caregiver. Monitor 1 was excluded because it was right above the caregiver (see Fig. 1 and Fig. 3) and thus it would be hard to tell if a child was distracted or not.

Then, the caregiver was expected to pause the video to see the child's reaction. The child was expected to look toward the caregiver to find out why the video was paused. Once the child looked at the caregiver, the caregiver would immediately record the response (by pressing the "Reward" button) and hereby trigger a reward video clip. The reward was set as resuming playing the paused video for 10 seconds on Monitor 1. The reward was presented by C3I rather than caregivers in order to provide consistent rewards to children.

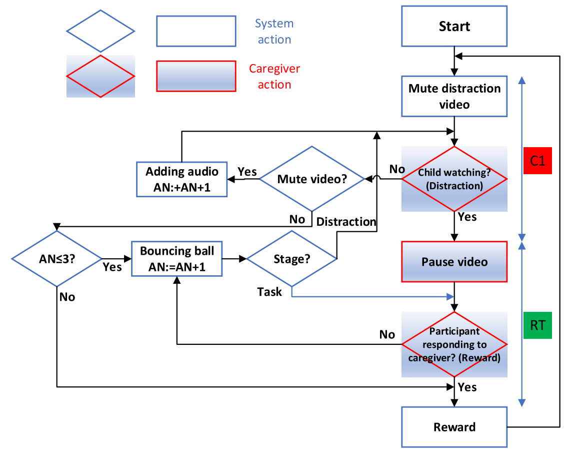

Fig. 7. Trial flowchart. Covariates: {C1: distraction duration}; Performance measurement: {RT: response time}.

If a child wasn't distracted by the muted video clip, sound might be added, or bouncing could be triggered to help transfer the child's attention onto the distraction video, per the caregiver's request through the caregiver tablet.

If a child didn't respond to a video pause, a bouncing ball could be triggered by the caregiver to transfer their attention to Monitor 1, which was the closest one to the caregiver.

III. Experiments

We conducted a feasibility study to observe the functioning and feasibility of C3I, which is presented below.

A. Study Design

In this feasibility study, *every caregiver-child dyad came for one session, which consisted of two trial-groups, in a total of 10 trials of IJA*. Each trial-group started with an introduction video, followed by 5 trials (a trial as shown in Fig. 7) and ended with a trial difficulty question for caregivers to evaluate children's frustration status.

B. Recruitment

Research participants were recruited from the CRUX clinical registry, a large database of patients who have consented to being contacted for research and had confirmed medical diagnosis of Autism Spectrum Disorder.

10 young children with ASD (5 male, 5 female, age: 18-32 months, mean = 26.7 months, SD = 4.1) with their caregivers were recruited for one session per dyad for the feasibility study. Two dyads dropped because the children were too fatigued and dysregulated to participate and attend to the task, resulting in an attrition rate of 20%. In addition, 2 of the remaining 8 dyads were excluded in the following data analyses due to data loss. Therefore, the data analyses in this work are based on 6 caregiver-child dyads (3 male, 3 female, age: 18-32 months, mean = 27.3 months, SD = 4.8). This study was approved by Vanderbilt University's Institutional Review Board under IRB Protocol 120922 (Date of approval 12/20/2018). Informed consent and assent procedures were conducted with participating families prior to enrollment by key study personnel.

As the caregiver played a significant role in this study, prior to the start of a session, a research assistant taught caregivers about the system interaction procedures and the tablet app. Additionally, caregivers were recommended to control the distraction duration (neither too long nor too short) according to their own judgement to explore different task difficulties.

C. Measurements

C3I was evaluated based on the following criteria: 1) the performance of children within C3I; 2) the stress of caregivers within C3I; and 3) the quality of C3I design, namely, covariates control and user experience. We used both quantitative and qualitative evaluation metrics, which are described below.

1) Response Time and Distraction Duration: Response time was used to objectively evaluate the performance of the children within C3I. It was the time duration between the moment a caregiver paused the video and the moment the caregiver confirmed the response from the child (see Fig. 7). A shorter response time was considered to indicate better child performance (more rapid *IJA*). Moreover, whenever a bouncing ball was triggered as a child not orienting to the parent, the response time was extended for approximately 3 seconds, after which the caregiver was allowed to confirm response. In other words, every time a bouncing ball was triggered, the response time was penalized for about 3 seconds. As such, response time was a comprehensive measurement of performance that included the need for additional support (bouncing ball). We did not separate independent looking from gaze following the bouncing ball as they both can be reflected by the response time.

The distraction duration (see Fig. 7) was chosen randomly by the caregiver with prompting from C3I. For example, if the child was detected watching the distraction video but the caregiver did not actively pause the video during the response time window, C3I would prompt the caregiver to make a decision, whether to keep waiting or to pause the video.

Giving caregivers the choice as to how long the duration lasted was an important part of including them in the system loop. The distraction duration was a covariate because it might also influence the response time of the child. For example, watching the distraction video for too long might get the child attached to the distraction video and thus might get detached from the C3I environment, influencing the response time of triggering IJA.

2) Caregiver Stress: Caregivers wore E4 wristbands to measure their Electrodermal Activity (EDA) and Photoplethysmography (PPG) signals. The detection frequency of EDA is 4 Hz and that of PPG is 64 Hz. Heart rate was extracted from the PPG, provided by the E4 embedded software, with a frequency of 1 Hz. Both EDA and heart rate signal were used to infer caregiver stress. EDA has been widely adopted to measure emotional processing and sympathetic activity [47]. Generally, the EDA signal contains two kinds of information: a slowly changing tonic skin conductance level (*tonic-SCL*) and a rapidly changing phasic skin conductance responses (*phasic-SCRs*) [47]. Here we used the variation of EDA as a measurement of caregiver stress based on the pattern summarized in [48] that decreased tonic-SCL and decreased frequency of phasic-SCRs indicate decreased stress compared to baseline.

The EDA signal was processed as follows: First, we filtered the raw EDA signal with a Gaussian filter of $window\ size = 32$. Then a deconvolution method proposed in [49] was adopted to extract the *tonic-SCL* and frequency of the *phasic-SCRs* for each trial. Next, the *tonic-SCL* was normalized by Equation (1) below to make it comparable across individuals [50]. Any *phasic-SCRs* with amplitude less than the threshold of 0.01 μS were excluded as noise [51]. Finally, a one-minute resting period before the session was used as baseline to measure within-individual variations [51]. The baseline period was used to measure the stress of caregivers interacting directly with children, which served as a snapshot of their daily lives.

$$SCL = (SCL - SCL_{min})/(SCL_{max} - SCL_{min}) \qquad (1)$$

Similar to the analysis of EDA signal, the heart rate data of caregivers was first normalized in a similar way as shown in Equation (1) and then the average heart rate in resting period was subtracted from the heart rate signal segments during trials. The commonly accepted pattern that decreased heart rate stand for decreased stress [48] was adopted to analyze stress variation.

3) User Experience Questionnaire: Immediately after the session, every caregiver responded to the system evaluation survey below:

1. How hard was it to use the app? (0: Very difficult, 4: Very easy)
2. How much help did you need to use the app? (0: Very much, 4: Not at all)
3. Overall, what do you think of the app? (0: Very bad, 4: Excellent)
4. How much do you think the system could help you teach your child? (0: Not at all, 4: Very much)
5. How hard were the tasks for your child? (0: Very easy, 4: Very difficult)
6. Did you like the activities where you paused the video to see if your child looked at you? (0: Not at all, 4: Very much)
7. What do you think of the system's video and sound? (0: Very bad, 4: Excellent)
8. How much do you think your child liked the system? (0: Not at all, 4: Very much)
9. What do you think of the system in general? (0: Very bad, 4: Excellent)
10. By the end of the second set of activities, do you think your child looked at you more quickly when the video was paused? (0: Not at all, 4: Very much)

IV. Results and Discussion

A. Quantitative Metrics

1) Response Time and Distraction Duration: To analyze the effect offered by C3I, pre-post trial-group comparison between Trial 1 and Trial 5 and between Trial 6 and Trial 10, and pre-post task comparison between Trial 1 and Trial 10 on response time are presented in Table II. The average response time gradually decreased from Trial 1 to Trial 6 and then

TABLE I
CAREGIVER TABLET APP DESIGN

Trial Step	Step behavior	Caregiver tablet app		
		Active mode	Passive mode	
			reminder	confirmation
1	Distraction video playing	Button: Pause video	Q1	Q2
2	Video pause	Button: Reward	Q3	Q4

Q1: Is he/she watching the video? Q2: System detects that she is watching the video, right? Q3: Is he/she looking back at you? Q4: System detects that she is looking back at you, right?

TABLE II
PRE-POST TRIAL-GROUP COMPARISON AND PRE-POST TASK COMPARISON ON RESPONSE TIME, MEAN WITH SD

IJA response time, units: second			
Trial 1	Trial 5	Trial 6	Trial 10
15.18 (10.99)	11.59 (7.08)	8.07 (6.24)	9.64 (9.11)
p_1=0.60		p_2=0.79	
p_3=0.052, Cohen's d=0.55			

p_1, p_2, p_3: *p*-values between Trial 1 and Trial 5, Trial 6 and Trial 10, Trial 1 and Trial 10, respectively, using paired Wilcoxon signed rank test.

increased by a small amount by the end of session (Trial 10). We believe that the reasons for this observed trend are: first, all children had initial difficulty with IJA as part of their existing ASD diagnoses, which improved with training as they gradually learned what to do within the system. Second, as the dyads were not familiarized with the system before the session, they likely required extra initial time to understand its components and the expectations. However, the time required for familiarization became less as the dyad participated more in the trials. The increased response time of trial 10 as compared to trial 6 might have come from the combination of minor perturbation and tiredness.

There was a near significant decrease of *IJA* response time for children with ASD in pre-post task comparison, with a medium effect size indicating the potential of C3I to help train *IJA* skills for children with ASD. In the meantime, paired Wilcoxon signed rank test on the covariate of distraction duration between Trial 1 (8.83 $\pm$ 12.28) and Trial 10 (9.04 $\pm$ 4.80) showed no significant difference ($p = 0.84$). The results showed that C3I was both technologically feasible to facilitate caregiver-child interaction and acceptable to the participants. In addition, we believe that the improvement observed in the IJA skill of children with ASD during this feasibility study indicated the potential of C3I and paved the way for a large scale study in the future.

2) Stress of Caregivers: As discussed in section III.C, baseline values were subtracted and therefore the baseline was labeled as 0 in Fig. 8-Fig. 10. "Group split" in these figures is the label between two trial-groups.

Tonic-SCL variations of *IJA* trials are displayed in Fig. 8. The vertical dashed lines denote the boundary between two trial-groups. We observed that Tonic-SCL started higher than baseline, indicating stress at the start of the session, when introduction video was played. It gradually reduced to the baseline thereafter, till the end of the session, which indicates that C3I did not increase caregiver stress over the course of the study. Because the *Tonic-SCL* is essentially as lowly

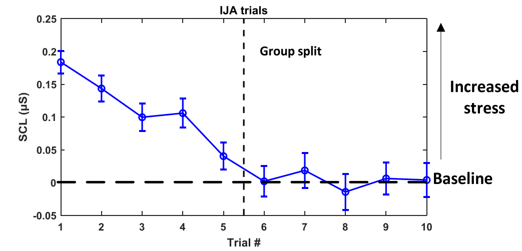

Fig. 8. Tonic-SCL variation of caregivers in *IJA* trials. (Mean value with 10% SD as error bar).

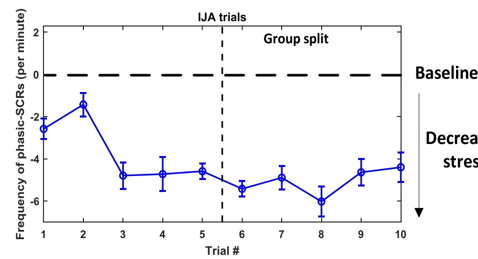

Fig. 9. Frequency of phasic-SCRs variation of caregivers in *IJA* trials (Mean value with 10% SD as bar).

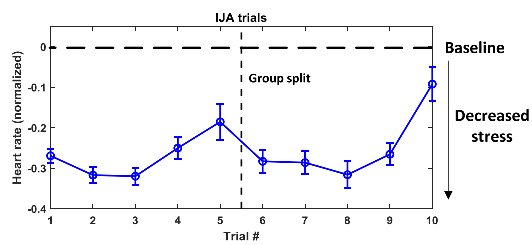

Fig. 10. Heart rate variation of caregivers in *IJA* trials (Mean value with 10% SD as error bar).

changing response, we might have seen further reduction if the experiment had more trials.

Variations of frequency of *phasic-SCRs* of *IJA* trials are shown in Fig. 9. Due to its fast-changing nature, the frequency of *phasic-SCRs* decreased immediately after the introduction video finished and reached their minimum starting at trial 3. Overall, compared to baseline, caregivers of ASD showed non-increasing physiological stress by the end of the session.

From Fig. 10, when measured by normalized heart rate, we can see that for *IJA* trials, stress of caregivers of ASD fluctuated with trials going on and were lower than baseline.

The statistics of tonic-SCL, frequency of phasic-SCRs, and normalized heart rate of caregivers are summarized in Table III.

To conclude, EDA and heart rate data were used as proxy for caregiver stress in this study. Results show that the system did not increase the stress of caregivers. It is important to document that participation in a C3I session did not inadvertently *increase* caregiver stress, which is always a risk when introducing a new technology to a vulnerable population.

B. Qualitative Metrics—User Experience Questionnaire

The results of the survey are summarized in Table IV. In general, caregivers were satisfied with the tablet app (Q3-3.33/4), task design (Q6-3.5/4) and the whole system

TABLE III
AVERAGE *Tonic-SCL*, FREQUENCY OF *Phasic-SCRS* AND HR OF CAREGIVERS IN *IJA* TRIALS, WITH SD IN THE PARENTHESIS

	IJA
Tonic-SCL	0.06(±0.42)
Phasic-SCRs	-4.4(±5.6)
Heart rate	-0.26(±0.28)

TABLE IV
AVERAGE SCALE WITH SD OF EACH QUESTION IN THE QUESTIONNAIRE

Question	Scale
Q1	3.33 (0.52)
Q2	3.17 (1.17)
Q3	3.33 (0.52)
Q4	3.67 (0.52)
Q5	2 (0.89)
Q6	3.5 (0.84)
Q7	3.5 (0.55)
Q8	3.17 (0.98)
Q9	3.5 (0.55)
Q10	3.17 (0.75)

(Q9-3.5/4). Regarding change in child behavior, caregivers were satisfied with how much *IJA* teaching impacted their children (Q10-3.17/4).

V. CONCLUSION & FUTURE WORK

A. Conclusion

We designed an immersive Computer-mediated Caregiver-Child Interaction (C3I) system for young children with ASD. To the best of our knowledge, this is the first system designed to promote *IJA* in children with ASD that 1) incorporates a caregiver for inputting decisions on social responsiveness and children's mental states; and 2) consistently and automatically tracks *IJA* performance. A feasibility study with 6 ASD caregiver-child dyads showed that C31 has potential to create situations in which children are likely to initiate joint attention tasks without creating additional stress for the caregivers. Decreased pre-post session *Response Time* in children indicated they were able to initiate joint attention more quickly after training. We also observed patterns of several physiological responses consistent with non-increasing stress. We believe that C3I's ability to help with various aspects of the *IJA* tasks did not increase workload on the caregivers, which led to non-increasing their stress.

C3I differed from existing works in that this is the first work trying to teach *IJA* skills to children with ASD in an immersive environment with caregiver in-the-loop, an explicit interaction protocol, and consistent *IJA* performance tracking and comparison. Given the high prevalence of children with ASD, the importance of caregivers to teaching new skills, the limited availabilities of trained therapists, and the high level of stress that caregivers experience on a daily basis, we believe our work lays the foundation for future studies to consider how technological support can also include caregivers to support children with ASD and their families. Including caregivers in this way may increase the chance of skill transfer outside of the experimental setting; meanwhile, the uniform stimuli and reward provision offered by the system make the *IJA* performance measurement consistent and comparable. Moreover, the presence of caregiver might help reduce "isolation effect", a phenomenon describing children's reduced social interests after interaction with autonomous system, which happens mainly due to a lack of real life components in autonomous systems designed for children with ASD [31].

B. Limitations

Although C3I showed promising results and feasibility, there were several limitations. First, the feasibility study had only one session per dyad with repetitive trials. In addition, the sample size was small and there was no control group. As such, the results of this feasibility study need to be considered with caution until a larger study verifies its generalizability. Moreover, while results are promising future work must evaluate whether improvement is a practice effect, true improvement, and most importantly whether the system translates into improvements outside of the intervention system.

Second, the stress of caregiver was analyzed based on physiological signals and assumptions of their relationships with stress. Although those assumptions were widely acknowledged [48], the results would be stronger if self-reports on caregiver's stress were also collected as an additional evidence. Finally, caregivers needed to operate the tablet in C3I, which occasionally distracted children's attention.

C. Future Work

In order to address the aforementioned limitations, in the future, we plan to conduct randomized controlled trials with larger sample sizes. The intervention effect may also be investigated with a control group of children who interact with ASOTS on *IJA* training, i.e., without a caregiver involved. With such a control group and multiple visits, isolation effect may also be quantitatively investigated to see if the incorporation of caregivers can help children on transferring learnt social skills back to daily lives or increasing their social interests [31]. Moreover, the tablet may be replaced by a Bluetooth remote controller, which senses the gesture along with ear buds which send system instruction to caregivers and collect their voice to trigger system events.

The scope of C31 may also be extended to include other tasks, such as response to joint attention, triadic play, and imitation leading to a more comprehensive intervention environment for children with ASD.

REFERENCES

[1] C. Wong *et al.*, "Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review," *J. Autism Develop. Disorders*, vol. 45, no. 7, pp. 1951–1966, Jul. 2015.
[2] M. J. Maenner, "Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR Surveill. Summaries*, vol. 69, no. 4, pp. 1–12, 2020.
[3] J. P. Leigh and J. Du, "Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the United States," *J. Autism Develop. Disorders*, vol. 45, no. 12, pp. 4135–4139, Dec. 2015.
[4] Z. E. Warren and W. L. Stone, "Best practices: Early diagnosis and psychological assessment," *Autism Spectr. Disorders*, pp. 1271–1282, May 2011.
[5] 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. Human-Machine Syst.*, vol. 48, no. 2, pp. 125–135, Apr. 2018.

[6] Z. Zheng *et al.*, "Design of an autonomous social orienting training system (ASOTS) for young children with autism," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 668–678, Jun. 2017.
[7] C. Kasari, S. Freeman, and T. Paparella, "Joint attention and symbolic play in young children with autism: A randomized controlled intervention study," *J. Child Psychol. Psychiatry*, vol. 47, no. 6, pp. 611–620, Jun. 2006.
[8] T. Charman, S. Baron-Cohen, J. Swettenham, G. Baird, A. Drew, and A. Cox, "Predicting language outcome in infants with autism and pervasive developmental disorder," *Int. J. Lang. Commun. Disorders*, vol. 38, no. 3, pp. 265–285, Jan. 2003.
[9] P. Mundy, "Joint attention and social-emotional approach behavior in children with autism," *Develop. Psychopathol.*, vol. 7, no. 1, pp. 63–82, 1995.
[10] L. Billeci *et al.*, "Disentangling the initiation from the response in joint attention: An eye-tracking study in toddlers with autism spectrum disorders," *Transl. Psychiatry*, vol. 6, no. 5, p. e808, May 2016.
[11] P. Mundy, C. Delgado, J. Block, M. Venezia, A. Hogan, and J. Seibert, "Early social communication scales (ESCS)," Univ. Miami, Coral Gables, FL, USA, Tech. Rep., 2003.
[12] W. L. Stone, E. E. Coonrod, L. M. Turner, and S. L. Pozdol, "Psychometric properties of the STAT for early autism screening," *J. Autism Develop. Disorders*, vol. 34, no. 6, pp. 691–701, Dec. 2004.
[13] Z. Zheng, G. Nie, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "A randomized controlled trial of an intelligent robotic response to joint attention intervention system," *J. Autism Develop. Disorders*, vol. 50, pp. 1–13, Feb. 2020.
[14] J. A. Vu, J. T. Hustedt, W. M. Pinder, and M. Han, "Building early relationships: A review of caregiver–child interaction interventions for use in community-based early childhood programmes," *Early Child Develop. Care*, vol. 185, no. 1, pp. 138–154, Jan. 2015.
[15] C. D. Hoffman, D. P. Sweeney, D. Hodge, M. C. Lopez-Wagner, and L. Looney, "Parenting stress and closeness: Mothers of typically developing children and mothers of children with autism," *Focus Autism Other Develop. Disabilities*, vol. 24, no. 3, pp. 178–187, Sep. 2009.
[16] A. Dabrowska and E. Pisula, "Parenting stress and coping styles in mothers and fathers of pre-school children with autism and down syndrome," *J. Intellectual Disability Res.*, vol. 54, no. 3, pp. 266–280, Mar. 2010.
[17] R. L. Gabriels, M. L. Cuccaro, D. E. Hill, B. J. Ivers, and E. Goldson, "Repetitive behaviors in autism: Relationships with associated clinical features," *Res. Develop. Disabilities*, vol. 26, no. 2, pp. 169–181, Mar. 2005.
[18] J. J. Diehl, L. M. Schmitt, M. Villano, and C. R. Crowell, "The clinical use of robots for individuals with autism spectrum disorders: A critical review," *Res. Autism Spectr. Disorders*, vol. 6, no. 1, pp. 249–262, Jan. 2012.
[19] 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 *Proc. 27th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2018, pp. 1–4.
[20] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the engagement with social robots," *Int. J. Social Robot.*, vol. 7, no. 4, pp. 465–478, Aug. 2015.
[21] S. Boucenna *et al.*, "Interactive technologies for autistic children: A review," *Cognit. Comput.*, vol. 6, no. 4, pp. 722–740, Dec. 2014.
[22] S. M. Anzalone *et al.*, "How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+time) environment during a joint attention induction task with a robot," *Res. Autism Spectr. Disorders*, vol. 8, no. 7, pp. 814–826, Jul. 2014.
[23] 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.
[24] R. E. Simut, J. Vanderfaeillie, A. Peca, G. Van de Perre, and B. Vanderborght, "Children with autism spectrum disorders make a fruit salad with probo, the social robot: An interaction study," *J. Autism Develop. Disorders*, vol. 46, no. 1, pp. 113–126, Jan. 2016.
[25] L. Boccanfuso, S. Scarborough, R. K. Abramson, A. V. Hall, H. H. Wright, and J. M. O'Kane, "A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: Field trials and lessons learned," *Auton. Robots*, vol. 41, no. 3, pp. 637–655, Mar. 2017.
[26] A. P. Bayliss, E. Murphy, C. K. Naughtin, A. Kritikos, L. Schilbach, and S. I. Becker, "'Gaze leading': Initiating simulated joint attention influences eye movements and choice behavior," *J. Exp. Psychol., Gen.*, vol. 142, no. 1, p. 76, 2013.
[27] M. Dalmaso, S. G. Edwards, and A. P. Bayliss, "Re-encountering individuals who previously engaged in joint gaze modulates subsequent gaze cueing," *J. Exp. Psychol., Learn. Memory Cognition*, vol. 42, no. 2, p. 271, 2016.
[28] S. G. Edwards, L. J. Stephenson, M. Dalmaso, and A. P. Bayliss, "Social orienting in gaze leading: A mechanism for shared attention," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 282, no. 1812, Aug. 2015, Art. no. 20151141.
[29] L. Schilbach *et al.*, "Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry," *J. Cognit. Neurosci.*, vol. 22, no. 12, pp. 2702–2715, Dec. 2010.
[30] P. Mundy and M. Crowson, "Joint attention and early social communication: Implications for research on intervention with autism," *J. Autism Develop. Disorders*, vol. 27, no. 6, pp. 653–676, 1997.
[31] D. Feil-Seifer and M. Matarić, "Socially assistive robotics," *IEEE Robot. Autom. Mag.*, vol. 18, no. 1, pp. 24–31, Mar. 2011.
[32] R. Johansen, "Groupware: Computer support for business teams," The Free Press, 1988.
[33] R. Luppicini, "Review of computer mediated communication research for education," *Instruct. Sci.*, vol. 35, no. 2, pp. 141–185, Mar. 2007.
[34] Y. Baruch, "The autistic society," *Inf. Manage.*, vol. 38, no. 3, pp. 129–136, Jan. 2001.
[35] M. Tscholl and R. Lindgren, "Empowering digital interactions with real world conversation," *TechTrends*, vol. 58, no. 1, pp. 56–63, Jan. 2014.
[36] N. Kucirkova, D. Messer, K. Sheehy, and R. Flewitt, "Sharing personalised stories on iPads: A close look at one parent-child interaction," *Literacy*, vol. 47, no. 3, pp. 115–122, Nov. 2013.
[37] R. C. Teepe, I. Molenaar, and L. Verhoeven, "Technology-enhanced storytelling stimulating parent-child interaction and preschool children's vocabulary knowledge," *J. Comput. Assist. Learn.*, vol. 33, no. 2, pp. 123–136, Apr. 2017.
[38] L. Zhang, Q. Fu, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Design and evaluation of a collaborative virtual environment (CoMove) for autism spectrum disorder intervention," *ACM Trans. Access. Comput.*, vol. 11, no. 2, p. 11, 2018.
[39] H. Zhao, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar, "Hand-in-hand: A communication-enhancement collaborative virtual reality system for promoting social interaction in children with autism spectrum disorders," *IEEE Trans. Human-Machine Syst.*, vol. 48, no. 2, pp. 136–148, Apr. 2018.
[40] D. L. Olds, L. Sadler, and H. Kitzman, "Programs for parents of infants and toddlers: Recent evidence from randomized trials," *J. Child Psychol. Psychiatry*, vol. 48, nos. 3–4, pp. 355–391, Mar. 2007.
[41] M. J. Matarić, "Socially assistive robotics: Human augmentation versus automation," *Sci. Robot.*, vol. 2, no. 4, Mar. 2017, Art. no. eaam5410.
[42] 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 *Proc. Int. Conf. Hum.-Comput. Interact.*, 2019, pp. 123–132.
[43] *E4*. Accessed: Oct. 11, 2020. [Online]. Available: https://www.empatica.com/en-int/research/e4/
[44] *Unity*. Accessed: Oct. 11, 2020. [Online]. Available: https://unity.com/products
[45] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 532–539.
[46] *Kinect*. Accessed: Oct. 11, 2020. [Online]. Available: https://developer.microsoft.com/en-us/windows/kinect/
[47] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, "A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments," *Psychophysiology*, vol. 49, no. 1, pp. 1017–1034, 2013.
[48] L. Shu *et al.*, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, Jun. 2018.
[49] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, Jun. 2010.
[50] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," in *Handbook of Psychophysiology* (Cambridge Handbooks in Psychology), J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2017, pp. 217–243.
[51] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, Aug. 2012.