Can Therapists Design Robot-Mediated Interventions and Teleoperate Robots Using VR to Deliver Interventions for ASD?

Roman Kulikovskiy[1], Megan Sochanski[1], Ala'aldin Hijaz[1], Matteson Eaton[2], Jessica Korneder[2], and Wing-Yue Geoffrey Louie[1]

***Abstract*— Socially Assistive Robots (SARs) have demonstrated success in the delivery of interventions to individuals with Autism Spectrum Disorder (ASD). To date, these robot-mediated interventions have primarily been designed and implemented by robotics researchers. It remains unclear whether therapists could independently utilize robots to deliver therapies in clinical settings. In this paper, we conducted a study to investigate whether therapists could design and implement robot-mediated interventions for children with ASD. Furthermore, we compared therapists' performance, efficiency, and perceptions towards using a Virtual Reality (VR) and kinesthetic-based interface for delivering robot-mediated interventions. Overall, our results demonstrated therapists could independently design and implement interventions with a SAR. They were faster at designing a new intervention using VR than a kinesthetic interface. Therapists also had similar performance to delivering in-person interventions when utilizing VR to deliver interventions with the robot. Therapists reported moderate workload using the VR interface and perceived VR to be usable.**

I. Introduction

A major application area for Socially Assistive Robots (SARs) is providing therapeutic treatment to individuals with Autism Spectrum Disorder (ASD). According to the Center for Disease Control, 1 in 54 individuals are identified with ASD [1]. ASD is a condition that affects an individual's social, emotional, adaptive and communication skills [1]. Early intervention for ASD has long-term positive impacts for these individuals [2], making it a current focus for SAR research. In general, these robot-mediated interventions (RMIs) have demonstrated positive outcomes [3]–[10].

While research on SARs for the delivery of therapies to individuals with ASD has primarily focused on those with ASD, the end-users (e.g., therapists) operating the robots have not received the same level of exploration [11]. End-user perceptions are important because if they do not feel they have the capability and/or knowledge to operate the technology, it will not be used. A study with experienced and future professionals demonstrated positive feedback and interest in using SARs for therapies but the participants felt they did not have the knowledge to use the technology [12]. The study in [12] was limited because end-users were provided only a brief demonstration of the robot's capabilities but did not operate the robot. Furthermore, RMIs have primarily been designed and implemented by researchers [13]. It remains unclear whether end-users could design interventions and operate robots to deliver an intervention. It is important to enable end-users to operate SARs so they can evaluate the clinical applicability of RMIs [13], [14]. Hence, there is presently an open opportunity to identify: 1) whether end-users could be trained to operate a SAR to deliver new therapies and 2) what tools would be most effective for end-users to operate a SAR.

Herein, we will focus on end-users operating SARs via teleoperation as it is a valuable tool for: 1) rapidly prototyping, evaluating, and implementing interventions delivered by a SAR; 2) data collection for development of the autonomy of a robot; and 3) evaluating models for human-robot interactions (HRIs) [15]–[18]. Putting these technologies directly in the hands of end-users, via teleoperation, will ensure that RMI can be evaluated for their clinical validity, which addresses an existing gap in RMIs [13]. Teleoperation also enables end-users to understand the capabilities and limitations of SARs as they are developed for interventions.

In this study, we investigate whether therapists can be trained to design and operate SARs to effectively, as well as efficiently, deliver new therapies to individuals with ASD. We also investigate therapist performance, efficiency, and perceptions with two different interfaces for designing RMIs and teleoperating a SAR to deliver the intervention. Namely, studies have shown that due to the rapidly changing needs of individuals with ASD, end-users require simple, fast, flexible, and usable controls for a robot [19]. Furthermore, we focus on therapies requiring a humanoid SAR to communicate both verbally and nonverbally as a majority of clinically relevant therapies for individuals with ASD require human-like verbal and nonverbal communication skills [20]–[22].

II. Literature Review

To date, SARs delivering therapies to individuals with ASD have primarily used pre-scripted social behaviors while

This work was supported by the National Science Foundation grant #1948224 and partially by the Michigan Space Grant Consortium, NASA grant #NNX15AJ20H

[1] Intelligent Robotics Laboratory, Oakland University, Michigan, USA (e-mail: louie@oakland.edu)

[2] Applied Behavior Analysis Clinic, Oakland University, Michigan, USA

the robot is teleoperated [3], [4], [15], [17], [23]. A common method for pre-scripting motions of robot behaviors is kinesthetic teaching [24]. Using pre-scripted behaviors requires the user to create all the behaviors a robot should utilize prior to an interaction, which can result in repetitive interactions. Furthermore, kinesthetic teaching can lead to unnatural body language and can be a difficult process for creating social behaviors as interpersonal communication is an automatic process that humans find difficult to describe explicitly [25], [26]. Professionals working with this population have indicated such repetitive and unnatural robot behaviors must be used with care as the goal of therapy is not to teach individuals with ASD to interact with a robot, but to enable them to generalize skills to human-human interactions [27], [28]. To effectively teach individuals with ASD skills used in human interactions, robots should model human-like variation during communication. Virtual Reality (VR) motion tracking could be a technology that enables therapists to naturally teleoperate a robot during the delivery of therapies.

VR motion tracking systems have demonstrated success for teleoperating industrial robots [29]–[31] and humanoid robots [32]–[34] for physical manipulation tasks. The VR systems developed for industrial robots have enabled users to teleoperate the position and orientation of robot end effectors while they were immersed in the robot's perspective [29]–[31]. In [31], users were more successful and efficient in completing manipulations when using VR rather than a joystick to control a robot. Additionally, VR was perceived as intuitive and easier to use. VR has also been utilized for joint-based teleoperation of humanoid robot upper torsos [32]–[34]. These systems either utilized skeleton tracking to determine a user's joint positions [32], [33] or infer joint positions according to limited user pose information [34]. Joint positions are used to compute joint angles of the user's arm and mapped to a robot's joint angles, so the user can teleoperate the robot by moving their own body. A study with novice users demonstrated that joint-based VR teleoperation of a humanoid robot for manipulation tasks was preferred over kinesthetic guidance [34]. Users also had a lower perceived workload and were more efficient performing tasks requiring two arms when using VR.

Despite the positive results with physical tasks, VR-based control has yet to be evaluated in teleoperating a robot for social tasks. Social tasks focus on clearly communicating a message to a communication partner through verbal, nonverbal, and affective cues. Teleoperating a robot to communicate effectively requires the simultaneous control of a robot's voice as well as the dynamics, position, and orientation of all the joints due to nonverbal communication being a whole-body effort [35]. In contrast, users teleoperating a robot for a manipulation task will primarily focus on only the robot's end effectors [29]–[34]. Prior works on evaluating users' experiences utilizing VR systems to teleoperate a robot have focused on only physical tasks where the environment is static and can only be altered by the robot [31], [34]. However, during a social task the communication partner has their own beliefs, affect, goals, and intentions, therefore requiring a robot to rapidly adapt to an individual. Such adaptation is important because humans will have negative attitudes towards a robot if it fails to follow social norms or produces socially inappropriate behaviors within the context of an interaction [36]–[39].

In this paper, we uniquely investigate the efficacy, efficiency, and user perceptions of a VR-based system to teleoperate a humanoid SAR to accomplish a social task requiring verbal and nonverbal behaviors. Our hypothesis is that a VR-based teleoperation system will be more effective, efficient, and perceived more positively than other modalities. We hypothesize that since interpersonal communication is a process humans find difficult to describe, a socially immersive interface will enable a user to be engaged in the HRI and naturally control robot communication behavior.

III. Virtual Reality-based Teleoperation System

A VR-based system was developed for teleoperating a humanoid robot to perform social tasks while immersed in a robot's perspective. We focused on humanoids because they can exhibit human-like verbal and nonverbal communication. This enables individuals with ASD that interact with robots in RMIs to potentially better transfer learned skills to real life human-human contexts [11]. The VR system developed is presented in Fig. 1a. The system consisted of three main modules: 1) a VR-based teleoperation interface; 2) a VR renderer; and 3) a kinematics solver. The VR-based teleoperation interface consisted of a head mounted display (HMD), hand-held controllers, and a microphone so that a user could teleoperate the robot. The VR rendering module generated the graphical output displayed on a HMD to provide visual feedback to the user and create an immersive experience from the perspective of the robot. The kinematics solver utilized the sensed HMD and hand-held controller positions to determine the joint angles of the user so that they could be mapped to the humanoid's joints.

A. *Pepper Humanoid Robot*

We utilized the Pepper robot as an example use case for our VR system. Pepper can exhibit human-like upper body movements using two degrees of freedom (DOF) in the neck, five DOF in each arm, and one DOF in each hand. The two DOF in the robot's neck provides yaw and pitch rotations of the head. The five DOF in each of the robot's arms allow shoulder pitch and roll, elbow yaw and roll, and wrist yaw movements. The one DOF in each hand open and closes the robot's hands. The robot can monitor the environment around itself via an RGB camera and microphone sensor. The NaoQi SDK was used to interface with the robot [40].

B. *VR-based Teleoperation Interface*

The VR-based interface was developed to take user inputs for teleoperating a robot and immersing a user in the robot's perspective. The OpenVR SDK and SteamVR runtime were used for all VR software development [41] [42]. The commercially available HTC Vive HMD, hand-held controllers, and microphone were used as input devices by the user to control the robot's head, arms, and voice

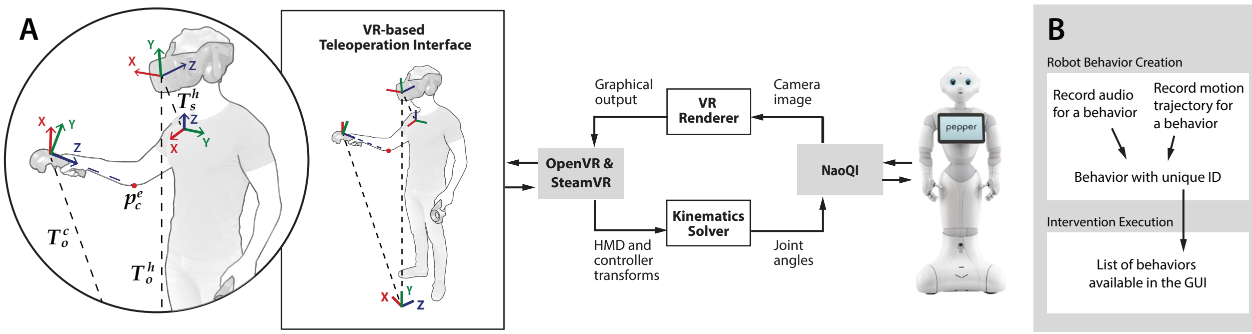

Fig. 1. a) VR-based Teleoperation System Architecture and b) Kinesthetic Interface

respectively. Users could naturally move their head while the HMD tracked the position and orientation. Users could also move their arms while the hand-held controllers tracked the user's hand position and orientation. These head, arm, and hand teleoperation inputs were provided to the kinematics solver module to map user to robot motions. The user could also speak at the robot's location through a microphone streaming audio to the robot's speaker. The HMD was also used to immerse a user within a social interaction from the perspective of the robot. Audio from the robot's microphone was transmitted to the user's headphones and the egocentric view of the visual information perceived by the robot was transmitted to the HMD. This is achieved by the VR renderer module generating graphics from the RGB camera in the robot's head to be displayed to the teleoperator in the HMD.

C. VR Renderer

An OpenGL-based VR rendering engine displayed a virtual dashboard and a first-person perspective of the robot's viewpoint in the HMD. Once equipped with the HMD, the user was immersed in a virtual environment with a dashboard of instructions on how to activate the robot teleoperation. When activated, the user was immersed in a first-person perspective of the robot's viewpoint and received control of the robot's upper body.

A monoscopic view of the RGB camera in the robot's head was presented in the HMD to provide a user the robot's viewpoint. A monoscopic view was utilized because unlike physical manipulation tasks, depth perception is not required for non-contact socially interactive tasks. Stereoscopic views also have the potential to increase the risk of cybersickness if the rendered views do not match a user's physiology [43], which can be amplified by latency during head motions [44]. To reduce these discrepancies, a virtual screen was created in the VR environment so that the robot's RGB camera stream could be displayed on it. The virtual screen was a fixed distance from the user. If the user moved their head, the virtual screen would follow their head motions so that it remained centered in their view. This technique of projecting a robot's view onto a screen, instead of directly into each of the user's eyes, has been shown to reduce motion sickness and improve immersion [45].

D. Kinematics Solver

The primary goal was to enable a user to naturally demonstrate motions and map these motions to a humanoid. The kinematics solver module enabled a user to control a robot's joint movements using a HMD and two hand-held controllers. The HTC Vive tracked and provided a transformation matrix for each of the devices relative to an origin frame (i.e., the ground). We used a modified version of the kinematics presented in [34] to map user motions to robot motions based on the HMD and controller transformations.

The user's head motions were mapped to the robot by controlling the robot's neck yaw (ψ_n) and pitch (θ_n). This was accomplished by directly mapping the yaw and pitch components from the transformation matrix of the tracked HMD (T_o^h) to the robot's neck yaw and pitch rotations:

$$\psi_n = yaw(T_o^h),\ \theta_n = -pitch(T_o^h) \tag{1}$$

User arm motions were mapped to the robot motions by first determining the elbow position of the user. Based on empirical evaluation, we approximated the user's wrists remained fixed when holding a controller due to the ergonomics of it, and for adults their elbow position was offset by 35 cm in the positive z direction from the controller: $p_c^e = \begin{bmatrix} 0 & 0 & 0.35 \end{bmatrix}$. The elbow position was then transformed into the shoulder reference frame where the shoulder pitch and roll were calculated. The transformation matrix from the shoulder to the head was taken from [34], which was derived from anthropomorphic measurements of adults. Formally, the HMD to shoulder transformation matrix (T_s^h) was defined as:

$$T_s^h = \begin{bmatrix} 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0.21 \\ 0 & 1 & 0 & 0.225 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} & R_o^h & & 0 \\ & & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where R_o^h is the rotation matrix of the HMD from the origin and used to ensure the shoulder frame was not affected by the rotation of a user's head. The position of the elbow can then be transformed into the shoulder frame (p_s^e) by:

$$p_s^e = T_s^h (T_o^h)^{-1} T_o^c p_c^e \tag{3}$$

Given the position of the elbow in the shoulder frame, the shoulder pitch (θ_s) can be calculated by:

$$\theta_s = \operatorname{atan2}(-p_{s,z}^e, p_{s,x}^e) \tag{4}$$

Prior to calculating the shoulder roll, the elbow position in the shoulder frame needs to account for the shoulder pitch rotation. The elbow position in the shoulder frame after the shoulder pitch rotation ($p^e_{s'}$) can be determined by:

$$p^e_{s'} = \begin{bmatrix} \cos\theta_s & 0 & -\sin\theta_s \\ 0 & 1 & 0 \\ \sin\theta_s & 0 & \cos\theta_s \end{bmatrix} p^e_s \tag{5}$$

The shoulder roll angle (ϕ_s) can then be calculated by:

$$\phi_s = \mathrm{atan2}(p^e_{s',y}, p^e_{s',x}) \tag{6}$$

To calculate the elbow angles, it was necessary to infer the direction of the forearm. We approximated the forearm to be in the -z direction relative to the controller: $e^f_c = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}$. We can determine the direction of the forearm in the elbow frame (e^f_e) after the shoulder pitch and roll rotations by:

$$e^f_e = R^s_e * R^h_s * (R^h_o)^{-1} * R^c_o * e^f_c \tag{7}$$

where the rotation matrix between the shoulder and elbow (R^s_e) was defined as:

$$R^s_e = \begin{bmatrix} \cos\phi_s & \sin\phi_s & 0 \\ -\sin\phi_s & \cos\phi_s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_s & 0 & -\sin\theta_s \\ 0 & 1 & 0 \\ \sin\theta_s & 0 & \cos\theta_s \end{bmatrix} \tag{8}$$

Once the forearm direction is in the the elbow frame of reference we can then calculate the elbow yaw (ψ_e) by:

$$\psi_e = \mathrm{atan2}(e^f_{e,z}, e^f_{e,y}) \tag{9}$$

The forearm direction in the elbow frame needs to account for the elbow yaw prior to calculating the elbow roll. The forearm in the elbow frame after the elbow yaw ($e^f_{e'}$) is:

$$e^f_{e'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi_e & \sin\psi_e \\ 0 & -\sin\psi_e & \cos\psi_e \end{bmatrix} e^f_e \tag{10}$$

The elbow roll (ϕ_e) can then be calculated by:

$$\phi_e = \mathrm{atan2}(e^f_{e',y}, e^f_{e',x}) \tag{11}$$

The wrist joints for the robot remained fixed and a fist could be made by the user pressing the trigger on the controllers.

IV. User Study with Therapists

We conducted a study with therapists from an Applied Behavior Analysis (ABA) clinic for children with ASD to evaluate whether therapists could plan, design, and implement an RMI. We also compared our developed VR interface with an existing kinesthetic-based interface, SoftBank Robotics Choregraphe [46], available for Pepper. We evaluated therapist training time, efficiency, performance, and perceptions for the two methods of designing and teleoperating a robot to deliver an intervention. The hypotheses we evaluated include:

H1: Therapists require less time to plan and design an intervention using the VR rather than the kinesthetic interface.

H2: There will be a difference in the number of errors produced by a therapist during in-person, VR, and kinesthetic intervention delivery. Namely, therapists will produce the most errors with the kinesthetic interface.

H3: There will be a difference in the time it takes a therapist for in-person, VR, and kinesthetic intervention delivery. Namely, therapists will require the most time with the kinesthetic interface.

H4: Therapists will rate the VR interface to have a higher usability than the kinesthetic interface.

H5: Therapists will rate the VR interface to have a lower workload than the kinesthetic interface.

A. Participants

The participants in this study included the therapists delivering the RMI and the children receiving the interventions. All participants were recruited from an ABA autism clinic.

The inclusion criteria for therapists were: 1) working at an ABA clinic; and 2) no prior history of seizures with VR. We recruited ten participants but only eight participants (one male and seven females) with ages ranging from 22-33 (μ=25.13, σ=4.05) completed the study. Two participants withdrew due to lack of availability. Our male to female ratio was representative of the population as a large proportion of individuals practicing ABA are female [47]. The inclusion criteria for the children were: 1) 3-8 years old; 2) DSM-V diagnosis of ASD [48]; and 3) has not mastered the skill of emotion recognition. Prior to the study, skill mastery was evaluated by the BCBA-D presenting to the children the six emotions to be taught in the interventions and having them name the emotion. Only children unable to name the emotions correctly were included in the study. In total, we had 4 child participants (3 males and 1 female) with an age range of 4-6 (μ=5, σ=0.95).

B. Emotion Recognition Intervention

A board-certified behavior analyst-doctoral (BCBA-D) developed an ABA emotion recognition intervention. The children were taught to recognize emotions only from an individual's body language, without facial expressions nor sound effects. We chose this intervention because a challenge faced by individuals with ASD is recognizing emotions [49], and recognizing them from body language is relevant during COVID-19 as facial expressions are occluded. Furthermore, recent research suggests that reading body language may be a more effective approach for an individual with ASD to recognize emotions than other facial cues [50]. This intervention also allowed therapists to experience teleoperating the robot to interact using verbal/nonverbal communication.

The interventions followed standard ABA clinical procedures and were broken down into three components. First, the therapist teleoperator would ask the child how they (i.e. the robot) are feeling while presenting an emotion using only the robot's movements. Initially, a vocal prompt of the correct emotion was provided until the child was able to respond correctly without the prompt. Second, the child would then be provided an opportunity to respond to the question. Third, the therapist teleoperator would then respond to the child either with: 1) social praise if they answered correctly or 2) follow-up with a prompt for the correct emotion if the child answered incorrectly or provided no response. This three step sequence defines a single discrete trial and a complete intervention consisted of nine trials. Each intervention aimed to teach three emotions by presenting each emotion three times in a randomized order. The three emotions taught in an intervention were randomly chosen from six emotions: happy, scared, sad, surprised, angry, and tired.

C. Study Design and Procedure

A within-subjects experiment was designed for each therapist to be trained to control Pepper using VR and kinesthetic interfaces and then independently plan, design, and implement a robot-based emotion recognition intervention, Fig. 2. The experiment was reviewed by an IRB and consent was obtained from all participants. The study was divided into three days for each participant: 1) the therapist delivering the intervention to the child in-person; 2) the therapist designing the RMI and delivering mock interventions with each of the interfaces; and 3) the therapist delivering real interventions to a child by controlling the robot using each of the interfaces. Participants were video recorded for post study analysis.

1) Day 1: The goal of day 1 was for the therapists to familiarize themselves with the intervention. Following typical clinic procedures for new interventions, the BCBA-D provided instructions on the intervention design and randomly assigned three emotions to the therapist to teach. The therapist then implemented the intervention with a child.

2) Day 2: The goal on day 2 was to train the therapists to utilize the two interfaces and allow them to experience the process needed to deliver a new intervention with a SAR. After being trained with each interface, the participants planned and designed their own robot-based emotion recognition intervention. They then delivered a mock intervention with an adult stand-in that simulated responses from a child. The order the interfaces were presented was counterbalanced.

VR Interface - The training session for the VR interface consisted of a researcher explaining and demonstrating that the VR equipment would allow the participant to perceive (i.e., hear or see) what the robot perceives and control the robot's arms, head, and speech. The participant practiced controlling the robot while it faced a mirror, which provided visual feedback of the robot's movements. Once the participant indicated they were ready to deliver an intervention, they delivered their mock intervention.

Kinesthetic Interface - The kinesthetic interface training session began with a walkthrough of Choregraphe, Fig. 1b. The participants were taught how to create a robot behavior (i.e., motion trajectories and/or speech) and were guided by the researcher through a creation of a sample behavior. Namely, the creation of each behavior consisted of two steps: 1) recording a motion trajectory by physically guiding the robot's joints through the desired path via kinesthetic teaching and 2) recording an audio file with a microphone so it can be played back on the robot's speakers for the behavior. Once the components of a behavior have been recorded, the behavior is made available as a list item with a unique ID on a graphical user interface containing all available robot behaviors. Selecting the behavior with a mouse click will result in the recorded behavior being replayed on the robot.

Once confident from the walkthrough, the participant was asked to create robot behaviors for their emotion recognition intervention without assistance. Participants then ran a practice intervention to confirm they had all the behaviors to run an intervention, with the chance to add or change behaviors. A mock intervention was then run with a researcher. The

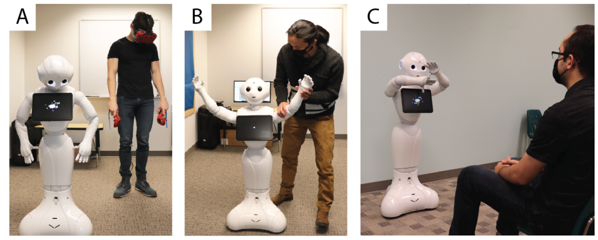

Fig. 2. a) VR Interface, b) Kinesthetic Interface, c) Mock Intervention

participant could hear and see from the robot's perspective by wearing headphones and viewing a stream of the robot's camera from a monitor. The participant conducted the intervention by selecting with a mouse the appropriate behaviors from the list of robot behaviors they created.

3) Day 3: The goal of day 3 was to have the participants use the two teleoperation interfaces to deliver the interventions they prepared on day 2, this time conducted with a child with ASD. The child receiving the treatment was chosen according to schedule availability. The order of the interfaces used by each participant was the same as on day 2.

D. Measures and Data Analysis

Therapist performance and efficiency in planning, designing, and implementing an RMI was evaluated using three measures: intervention planning and design time, treatment integrity, and intervention time. These measures were used for the VR and kinesthetic interfaces and were defined as:

Intervention planning and design time refers to any time the participant put towards creating and/or practicing their intervention. For the kinesthetic interface, this consisted of the participant planning and creating their behaviors using Choregraphe. For the VR interface, it was the amount of time the participant practiced their intervention with the SAR in front of a mirror.

Treatment Integrity refers to the extent to which the intervention was implemented successfully [51]. In the emotion recognition intervention this was evaluated according to the correct application of the components of an ABA discrete trial. The components included: establishing a child is looking at the participant before presenting an instruction; presenting an instruction using the same actions defined by the intervention protocol provided by the BCBA-D; providing a prompt by presenting the correct emotion after the instruction; providing a prompt with the correct delay after an instruction; re-presenting an instruction when a child makes an error; reinforcing a child with social praise (e.g., "great job!") within 5 seconds of a correct response; and ending a trial by providing verbal feedback that a child's response was incorrect (e.g., "let's try again). Not all components are required in each trial because the requirement of a component is dependent on a child's response. Treatment integrity was calculated as the total number of correctly implemented components divided by the total required components and converted to a percentage. Treatment integrity was collected for interventions implemented with the child on days 1 and 3. Treatment integrity was coded by two independent researchers on 50% of the data with an interobserver agreement of 100% using the trial-by-trial method.

Intervention times refers to the time required to implement an intervention. Only behaviors that pertained to the intervention were included in measuring intervention time for both interfaces. Again, intervention time was only collected on days with the child, days 1 and 3.

Participant perceptions of usability and workload towards the interfaces were also measured using the System Usability Scale (SUS) [52] and NASA Task Load Index (NASA-TLX) [53], respectively. These post-task questionnaires were administered on days 2 and 3, after the participants utilized each interface to implement an intervention. In total, four questionnaires were administered to each participant. Open-ended questions were also administered to investigate participants' user experience with the interfaces.

After collecting the data, a two-tailed paired t-test was conducted to test hypotheses H1, H4, and H5 with the VR and kinesthetic interfaces as the conditions for each dependent variable. Prior to running the two-tailed paired t-test, we confirmed the dependent variables were normally distributed using the Shaprio-Wilk test. A repeated measures ANOVA was conducted for H2 and H3 with the in-person, VR, and kinesthetic interventions as the conditions for the dependent variables. Shapiro-Wilk and Mauchly's tests were used to test for normality and sphericity, respectively. When sphericity was violated, a Greenhouse-Geiser correction was applied. An $\alpha = 0.05$ was set for all tests.

V. Results & Discussion

All participants were capable of designing and planning an RMI using both the VR and kinesthetic interfaces. On average, participants designed and planned the intervention faster with the VR interface (μ=523.4s, σ=235.9) than the kinesthetic interface (μ=1631.6s, σ=1068.5). There was a statistically significant difference in time required by therapists for designing and planning an intervention using the VR compared to the kinesthetic interface ($t(7)$=-3.093, p=0.017), which supports H1. This finding aligns with the short answer responses from participants stating that the kinesthetic design and planning were more labor intensive and required significant preparation.

The average treatment integrity for the in-person intervention was 84.0% (σ=17.5). The VR interface used to teleoperate the robot in delivering the intervention to the child resulted in a similar average treatment integrity, 83.7% (σ=12.7). When participants used the kinesthetic interface with the child there was a decrease in the average treatment integrity to 52.9% (σ=22.5). There was a statistically significant difference in treatment integrity between in-person, VR, and kinesthetic intervention delivery ($F(1.208, 8.453)$=15.374, p=0.003). Post hoc tests with a Bonferroni correction revealed that the difference in treatment integrity between in-person and VR was not statistically significant (p=1.000), but between in-person and kinesthetic was statistically significant (p=0.021). The difference in treatment integrity between VR and kinesthetic intervention delivery was also statistically significant (p=0.008), which supports H2. Participants reported it was easier to implement the components of an intervention using VR and participants found it easier to adapt to a child's changing needs.

The participants required on average 83.4s (σ=22.4) to deliver the intervention in-person, 148.4s (σ=20.8) using VR, and 235.8s (σ=122.4) using the kinesthetic interface. There was a statistically significant difference between in-person, VR, and kinesthetic intervention delivery ($F(1.072, 7.504)$=10.491, p=0.012). Post hoc tests with Bonferroni corrections revealed there was a statistically significant difference in intervention time between in-person and VR (p=0.002), as well as in-person and kinesthetic (p=0.013). There was no statistically significant difference in intervention time between VR and kinesthetic delivery (p=0.247). Therefore, H3 was not supported. The rejection of H3 is likely explained by participants indicating they would need more practice with VR to use the robot more efficiently. Additionally, one survey indicated that it was more difficult to tell the system what to do instead of doing it themselves.

On average, participants rated the usability of the VR interface with a SUS score of 63.75 (σ=11.1) for the mock intervention, and 53.75 (σ=16.6) for the intervention with a child. Participants rated the usability of the kinesthetic interface with a SUS score of 52.19 (σ=13.5) for the mock intervention, and 44.38 (σ=10.5) for the intervention with a child. With a higher SUS score in the mock and real interventions with a child, the VR interface demonstrated better system usability. However, there was not a statistically significant difference between therapists' SUS scores for the interfaces during the mock ($t(7)$=1.527, p=0.171) or real interventions ($t(7)$=1.309, p=0.232). This suggests that H4 was not supported in this study. A common challenge brought up by participants with the VR interface was their lack of body awareness and slight differences in embodiment. In the future, we plan to investigate the addition of a third person perspective of the robot to provide visual feedback and improve usability.

The workload between the VR and kinesthetic interface for the mock interventions were similar, but for the real interventions with the child, the VR interface had lower perceived workload. Participants, on average, rated the workload of using the VR interface with a NASA-TLX score of 55.67 (σ=17.9) for the mock intervention and 56.83 (σ=12.6) for the real intervention. Participants' average rating for workload using the kinesthetic interface was 59.67 (σ=12.1) for the mock intervention and 71.21 (σ=12.0) for the real intervention. Participants perceptions on workload between the VR and kinesthetic interfaces during the mock intervention ($t(7)$=-0.574, p=0.584) and intervention with children with ASD ($t(7)$=-1.784, p=0.118) suggests that there was no statistically significant difference. Hence, H5 was not supported. The workload scores for using the VR interface, for both the mock and real interventions, coincided with the median NASA-TLX global workload scores typically observed with robot operation tasks [54]. This is expected because therapists need to constantly adapt to the learning needs of a child which can result in high mental demand.

Overall, this study demonstrated that participants had positive perceptions towards using the robot for intervention delivery and therapists can be trained to design and implement their own RMI. In general, participants were better at utilizing VR rather than kinesthetic to deliver interventions. Namely, their performance using VR was on par with in-person interventions in terms of treatment integrity. Participants could also design and plan interventions faster using VR over the kinesthetic interface. Furthermore, all participants found VR more natural for delivering interventions and most participants found VR to be more adaptive to children's changing behavior than the kinesthetic interface.

REFERENCES

[1] K. A. Shaw, *et al.*, "Early identification of autism spectrum disorder among children aged 4 years - early autism and developmental disabilities monitoring network, six sites, united states, 2016," *Morbidity and mortality weekly report. Surveillance summaries*, vol. 69, no. 3, pp. 1–11, 2020.

[2] D. Granpeesheh, D. R. Dixon, J. Tarbox, A. M. Kaplan, and A. E. Wilke, "The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders," *Research in Autism Spectrum Disorders*, vol. 3, no. 4, pp. 1014–1022, 2009.

[3] W.-Y. G. Louie, J. Korneder, I. Abbas, and C. Pawluk, "A study on an applied behavior analysis-based robot-mediated listening comprehension intervention for ASD," *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 31–46, 2021.

[4] B. Robins, K. Dautenhahn, R. Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills?" *Universal Access in the Information Society*, vol. 4, no. 2, pp. 105–120, 2005.

[5] J. Greczek, E. Kaszubski, A. Atrash, and M. Matarić, "Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 561–566.

[6] E. Bekele, *et al.*, "Pilot clinical application of an adaptive robotic system for young children with autism," *Autism*, vol. 18, no. 5, pp. 598–608, 2014.

[7] C. A. Pop, *et al.*, "Social robots vs. computer display: Does the way social stories are delivered make a difference for their effectiveness on asd children?" *Journal of Educational Computing Research*, vol. 49, no. 3, pp. 381–401, 2013.

[8] D. Feil-Seifer and M. Mataric', "Robot-assisted therapy for children with autism spectrum disorders," in *Proceedings of the International Conference on Interaction Design and Children*, 2008, pp. 49–52.

[9] K. Jordan, M. King, S. Hellersteth, A. Wirén, and H. Mulligan, "Robotic technology for teaching adolescents with autism spectrum disorder: A feasibility study," in *Proceedings of the International Conference on Rehabilitation Engineering & Assistive Technology*, 2012, pp. 1–4.

[10] H. E. Andreae, P. M. Andreae, J. Low, and D. Brown, "A study of auti: A socially assistive robotic toy," in *Proceedings of the Conference on Interaction Design and Children*, 2014, pp. 245—248.

[11] A. M. Alcorn, *et al.*, "Educators' Views on Using Humanoid Robots With Autistic Learners in Special Education Settings in England," *Frontiers in Robotics and AI*, vol. 6, no. 107, pp. 1–15, 2019.

[12] D. Conti, S. Di Nuovo, S. Buono, and A. Di Nuovo, "Robots in Education and Care of Children with Developmental Disabilities: A Study on Acceptance by Experienced and Future Professionals," *International Journal of Social Robotics*, vol. 9, no. 1, pp. 51–62, 2017.

[13] M. Begum, R. W. Serna, and H. A. Yanco, "Are Robots Ready to Deliver Autism Interventions? A Comprehensive Review," *International Journal of Social Robotics*, vol. 8, no. 2, pp. 157–181, 2016.

[14] 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," *Research in Autism Spectrum Disorders*, vol. 6, no. 1, pp. 249–262, 2012.

[15] M. Hirokawa, A. Funahashi, Y. Itoh, and K. Suzuki, "Adaptive behavior acquisition of a robot based on affective feedback and improvised teleoperation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 3, pp. 405–413, 2019.

[16] K. Tsiakas, M. Kyrarini, V. Karkaletsis, F. Makedon, and O. Korn, "A taxonomy in robot-assisted training: Current trends, needs and challenges," *Technologies*, vol. 6, no. 4, pp. 1–19, 2018.

[17] B. A. English, A. Coates, and A. Howard, "Recognition of gestural behaviors expressed by humanoid robotic platforms for teaching affect recognition to children with autism-a healthy subjects pilot study," in *International Conference on Social Robotics*, 2017, pp. 567–576.

[18] P. G. Esteban, *et al.*, "How to Build a Supervised Autonomous System for Robot-Enhanced Therapy for Children with Autism Spectrum Disorder," *Paladyn, Journal of Behavioral Robotics*, vol. 8, no. 1, pp. 18–38, 2017.

[19] T. Hughes-Roberts and D. Brown, "Implementing a Robot-Based Pedagogy in the Classroom: Initial Results from Stakeholder Interviews," in *International Conference on Interactive Technologies and Games*, 2015, pp. 49–54.

[20] L. Schreibman, *et al.*, "Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 45, pp. 2411––2428, 2015.

[21] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, "'if you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2012, pp. 125—126.

[22] R. Tamagawa, C. I. Watson, I. H. Kuo, B. A. MacDonald, and E. Broadbent, "The effects of synthesized voice accents on user perceptions of robots," *International Journal of Social Robotics*, vol. 3, no. 3, pp. 253—262, 2011.

[23] A. Taheri, A. Meghdari, M. Alemi, and H. Pouretemad, "Teaching music to children with autism: A social robotics challenge," *Scientia Iranica*, vol. 26, no. 1, pp. 40–58, 2019.

[24] K. Fischer, *et al.*, "A comparison of types of robot control for programming by demonstration," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2016, pp. 213–220.

[25] S. Calinon and A. Billard, "Incremental Learning of Gestures by Imitation in a Humanoid Robot," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2007, pp. 255–262.

[26] K. Vogeley and G. Bente, ""Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication," *Neural Networks*, vol. 23, no. 8, pp. 1077–1090, 2010.

[27] C. A. Huijnen, M. A. Lexis, R. Jansens, and L. P. de Witte, "Roles, Strengths and Challenges of Using Robots in Interventions for Children with Autism Spectrum Disorder (ASD)," *Journal of Autism and Developmental Disorders*, vol. 49, no. 1, pp. 11–21, 2019.

[28] L. McHugh, A. Bobarnac, and P. Reed, "Brief report: teaching situation-based emotions to children with autistic spectrum disorder," *Journal of autism and developmental disorders*, vol. 41, no. 10, pp. 1423–1428, 2011.

[29] J. Chen, *et al.*, "Development of an immersive interface for robot teleoperation," in *Towards Autonomous Robotic Systems*, Y. Gao, S. Fallah, Y. Jin, and C. Lekakou, Eds. Cham: Springer International Publishing, 2017, pp. 1–15.

[30] D. Whitney, E. Rosen, D. Ullman, E. Phillips, and S. Tellex, "Ros reality: A virtual reality framework using consumer-grade hardware for ros-enabled robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018, pp. 1–9.

[31] A. Jackson, B. D. Northcutt, and G. Sukthankar, "The benefits of immersive demonstrations for teaching robots," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2019, pp. 326–334.

[32] C. Stanton, A. Bogdanovych, and E. Ratanasena, "Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning," in *Proceedings of Australasian Conference on Robotics and Automation*, 2012, pp. 1–10.

[33] L. Fritsche, F. Unverzag, J. Peters, and R. Calandra, "First-person teleoperation of a humanoid robot," in *IEEE Conference on Humanoid Robots*, 2015, pp. 997–1002.

[34] M. Hirschmanner, C. Tsiourti, T. Patten, and M. Vincze, "Virtual reality teleoperation of a humanoid robot using markerless human upper body pose imitation," in *IEEE International Conference on Humanoid Robots*, 2019, pp. 259–265.

[35] J. Burgoon, L. Guerrero, and K. Floyd, *Nonverbal Communication*. Allyn & Bacon, 2010.

[36] J. Kennedy, P. Baxter, and T. Belpaeme, “The Impact of Robot Tutor Nonverbal Social Behavior on Child Learning,” *Frontiers in ICT*, vol. 4, no. 6, pp. 1–15, 2017.

[37] M. K. Lee and L. Takayama, "“Now, I Have a Body”: Uses and Social Norms for Mobile Remote Presence in the Workplace,” in *ACM CHI Conference on Human Factors in Computing Systems*, 2011, pp. 33–42.

[38] J. Jung, T. Kanda, and M. S. Kim, “Guidelines for Contextual Motion Design of a Humanoid Robot,” *International Journal of Social Robotics*, vol. 5, no. 2, pp. 153–169, 2013.

[39] L. Wang, P.-L. P. Rau, V. Evers, B. K. Robinson, and P. Hinds, “When in Rome: The Role of Culture & Context in Adherence to Robot Recommendations,” in *ACM/IEEE International Conference on Human-Robot Interaction*, 2010, pp. 359–366.

[40] SoftBank Robotics. (2020, October) Naoqi developer guide. [Online]. Available: https://developer.softbankrobotics.com/nao6/naoqi-developer-guide

[41] Valve Corporation. (2020, October) OpenVR. [Online]. Available: https://github.com/ValveSoftware/openvr/

[42] Valve Corporation. (2020, October) SteamVR. [Online]. Available: https://www.steamvr.com/en/

[43] J. Häkkinen, M. Pölönen, J. Takatalo, and G. Nyman, “Simulator sickness in virtual display gaming: A comparison of stereoscopic and non-stereoscopic situations,” in *Proceedings of the Conference on Human-Computer Interaction with Mobile Devices and Services*, 2006, pp. 227–229.

[44] M. H. Draper, E. S. Viirre, T. A. Furness, and V. J. Gawron, “Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments,” *Human Factors*, vol. 43, no. 1, pp. 129–146, 2001.

[45] E. D. Visser and C. Tossell, “Improving the Visual Comfort of Virtual Reality Telepresence for Robotics,” in *International Conference on Social Robotics*, 2019, pp. 697–706.

[46] SoftBank Robotics. (2021, March) Developer center. [Online]. Available: https://developer.softbankrobotics.com/

[47] M. R. Nosik and L. L. Grow, “Prominent Women in Behavior Analysis: An Introduction,” *Behavior Analyst*, vol. 38, no. 2, pp. 225–227, 2015.

[48] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, 2013.

[49] S. Kuusikko, *et al.*, “Emotion recognition in children and adolescents with autism spectrum disorders,” *Journal of Autism and Developmental Disorders*, vol. 39, no. 6, pp. 938–945, 2009.

[50] C. C. Peterson, V. Slaughter, and C. Brownell, “Children with autism spectrum disorder are skilled at reading emotion body language,” *Journal of Experimental Child Psychology*, vol. 139, pp. 35–50, 2015.

[51] J. O. Cooper, T. E. Heron, and W. L. Heward, *Applied Behavior Analysis*. Hoboken, NJ: Pearson Education, 2019.

[52] J. Brooke, “System usability scale (sus): a quick-and-dirty method of system evaluation user information,” in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland, Eds. London, UK: Taylor and Francis, 1986, pp. 184–194.

[53] S. G. Hart and L. E. Staveland, “Development of nasa-tlx (task load index): results of empirical and theoretical research,” *Advances in Psychology*, vol. 52, pp. 139–183, 1988.

[54] R. A. Grier, “How high is high? a meta-analysis of nasa-tlx global workload scores,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1, pp. 1727–1731, 2015.