Computers in Human Behavior 104 (2020) 106163

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Human-Computer Interaction based Joint Attention cues: Implications on functional and physiological measures for children with autism spectrum disorder

Vishav Jyoti [1,*], Uttama Lahiri [2]

[1] Centre for Cognitive and Brain Sciences, Indian Institute of Technology Gandhinagar, Gandhinagar, India
[2] Department of Electrical Engineering, Indian Institute of Technology Gandhinagar, Gandhinagar, India

ARTICLE INFO

Keywords:
Autism spectrum disorder
Cue type
Human-computer interaction
Joint Attention
Physiology
Virtual reality

ABSTRACT

One of the important facets of effective social communication is Joint Attention (JA). However, children with Autism Spectrum Disorder (ASD) are often characterized by JA-related deficits, adversely affecting their social communication. In conventional interventions, therapists use different types of JA cues depending on one's capability to pick up the delivered cue. Though effective, conventional approaches suffer from restricted healthcare resources, cost, etc. With an increase in computational power, investigators are exploring alternative robot-based and computer-based techniques for JA skill training while delivering different types of JA cues. However, robot-assisted techniques are powerful but suffer from limitations such as high cost, restricted flexibility, etc. Thus, researchers are exploring the use of computer-based techniques for JA skill training since it can be controllable, flexible, cost-effective, more accessible, etc. With the advent of rich graphics, researchers are augmenting computer-based interfaces with Virtual Reality (VR) while designing Human-Computer Interaction (HCI)-based JA tasks. Given the importance of VR-enabled HCI-based JA training platform, studying the comparative potential of different types of JA cues (having varying information content) implemented using a VR-enabled HCI-based task platform is important. In this research work, we presented a VR-enabled HCI-based JA task platform that can deliver avatar-mediated and environment-triggered JA cues of varying information content. Results of a preliminary study with twenty typically developing and twenty age-matched children with ASD indicate differentiated implications of JA cues of varying information content on one's functional and physiological measures.

1. Introduction

Children with Autism Spectrum Disorder (ASD *henceforth*) are often characterized by deficits in social communication (Wetherby, Watt, Morgan, & Shumway, 2007) that disturbs their social functioning and their potential to earn their own livelihood. The problem becomes prominent given the high prevalence rate of ASD (with 1 in 59 children in USA (Data and Statistics, 2014), 1 in 100 in UK (The National Autistic Society, 2017, p. 40) and 1 in 68 in India (Dalwai et al., 2017)). Difficulties in social communication can be attributed to this target group having deficits in Joint Attention (JA) skill (Charman, 2003). The JA skill can be considered as one in which an individual can triangulate gaze with a social partner towards a target of interest (as cued by the partner) resulting in a triadic relationship (Bakeman & Adamson, 1984, pp. 1278–1289). This skill is important since joint focus (by an individual with his/her partner) towards the cued target immersed in a visual space or context can provide one with an insight to the experiences of the social partner (Shin, 2012). Additionally, the JA skill plays an important role in the development of cognitive processes related to social motivation (Mundy & Crowson, 1997) along with language development during later stages in life (Dawson et al., 2004). The JA skill can be broadly categorized as Response to Joint Attention (RJA) and Initiating Joint Attention (IJA) (Meindl & Cannella-Malone, 2011). The RJA refers to the type of JA skill in which an individual follows the cue offered by his/her social partner while attending to a stimulus (Bruinsma, Koegel, & Koegel, 2004). For example, a father might be cueing his child using finger-pointing to look towards a balloon in the visual space. In this case, if the child follows his father's pointed finger and in turn

* Corresponding author.
E-mail addresses: vishav.jyoti@iitgn.ac.in (V. Jyoti), uttamalahiri@iitgn.ac.in (U. Lahiri).

https://doi.org/10.1016/j.chb.2019.106163
Received 24 June 2019; Received in revised form 10 September 2019; Accepted 8 October 2019
Available online 12 October 2019
0747-5632/

looks towards the balloon, then the RJA is considered to be established. The Initiating Joint Attention (IJA) occurs when an individual initiates an intentional cue to get the attention of his/her social partner towards a stimulus in the shared environment (Bruinsma et al., 2004). Children with ASD face milestones, both with respect to RJA and IJA skills (Jones, Carr, & Feeley, 2006). Though IJA is also important, in our present work, we have focused on RJA (referred to as JA *henceforth*) since this is related with one's ability to reciprocate during social interaction (Jones et al., 2006) .

Given the importance of JA, intervention services incorporate JA skill training as an important component. A conventional JA skill training setup comprises of targets of interest and a therapist who uses different types of cueing for administering the JA tasks (Taylor & Hoch, 2008). The purpose of using different types of JA cueing is to attract the attention of an individual towards an object of interest, thereby facilitating the administration of the JA tasks. Again, a therapist's decision on the type of cue to be delivered often depends on the individual's ability to pick up a cue. For example, often, the therapist starts with the gaze-based cue in which the therapist changes the orientation of his/her gaze to point towards the target of interest (Leekam, Hunnisett, & Moore, 1998). A child with ASD might be unable to pick up the gaze-based JA cue since these children often exhibit atypical looking pattern (Frazier, Strauss, Klingender, Zetzer, Hardan, & Eng, 2017) in which they tend to demonstrate gaze avoidance while interacting with a social partner. In such a scenario, the therapist might increase the information content in his/her prompt since adding more information to a cueing prompt might make it easier for the child to track the cued target in the shared visual space. Thus, the therapist might use head turn (along with gaze orientation) to point towards the target of interest. In case the child is unable to pick up this cue, the therapist might decide to add more information to his/her cueing prompt. For example, the therapist might add a pointing finger (gesture) along with the head being turned and gaze being oriented towards the target of interest (Leekam et al., 1998). Though administration of JA tasks using conventional techniques is powerful, yet such techniques require long hours of repetitive one-on-one sittings with the therapist that is often costly (Chasson, Harris, & Neely, 2007) given the restricted availability of adequately trained interventionists particularly in developing countries like India. Additionally, bringing in variations in the training environment through the use of different objects (that can be used as targets of interest) physically in the clinical setup might not always be feasible.

Given the importance of JA skill training and the limitations faced by conventional intervention settings, experimenters have been exploring alternative techniques such as robot-assisted and computer-based techniques. For example, researchers have implemented some of the aspects of JA-based cueing (similar to that used in conventional techniques) that were delivered through the robotic agent. Specifically, researchers (such as Esubalew et al., 2012) have implemented a robot-assisted JA task paradigm in which the robot was programmed to demonstrate JA-based cueing prompts with a gradual increase in the cueing information. Subsequently, an individual's (child with ASD undergoing therapy) JA skill-related functional outcomes such as task performance was studied. Though powerful, robot-assisted techniques often suffer from some limitations such as safety issues, high cost, power constraints, etc. Thus, experimenters have been exploring the use of other alternatives e.g., computer-based techniques (Billeci et al., 2016) for administering JA tasks. With an increase in computational power coupled with rich graphics, researchers are augmenting computer-based systems with Virtual Reality (VR *henceforth*) platforms (Bernardini, Porayska-Pomsta, & Smith, 2014). Specifically, the research study (Caruana, McArthur, Woolgar, & Brock, 2017) has reported the power of using VR-based interfaces while designing Human-Computer Interaction (HCI *henceforth*) based tasks. Also, computer-based platforms are becoming cost-effective, more accessible, etc. Additionally, VR provides a space for training where individuals can afford to make mistakes in the process of learning a skill without facing the negative consequences (Bozgeyikli, Raij, Katkoori, & Alqasemi, 2017). Also, VR can feature human-like facilitators (important for administering skill training (Amaral, Simões, Mouga, Andrade, & Castelo-Branco, 2017)) as 3-D humanoid characters (Avatars) embedded in the VR space making the training sessions realistic. For example, use of VR-based social environments such as a gaming room, shopping mall, classroom, etc. coupled with avatars (Schroeder, 2012) used for imparting social communication skills has been reported. In spite of the several advantages offered by VR-enabled HCI-based platforms, only a limited number of studies (Pfeiffer-Leßmann et al., 2012; Bernardini et al., 2014; Courgeon, Rautureau, Martin, & Grynszpan, 2014; Little, Bonnar, Kelly, Lohan, & Rajendran, 2016; Amaral et al., 2017; Andrist, Gleicher, & Mutlu, 2017; Caruana et al., 2017; Caruana et al., 2018; Mei et al., 2018) have used HCI-based platform to address deficits in one's JA skill while exploring the use of various cueing strategies. Mostly, these studies have used either gaze-based cueing (Caruana et al., 2018), or head turn cue along with finger-pointing (Bernardini et al., 2014) to direct one's attention towards an object of interest followed by studying some of the functional outcomes. Given the importance of using different types of cues (having varying information content) and VR-enabled HCI-based platforms as far as JA skill training is concerned, further exploration to study the comparative potential of the different types of JA cues (delivered through the VR-enabled HCI-based JA task platform) to have implications on one's JA skill-related functional outcomes is warranted.

Motivated by this, in present research work, we have designed a VR-enabled HCI-based JA task platform for administering JA tasks for children with ASD. Our system was equipped with a virtual agent (a 3-D humanoid character referred to as avatar). The JA-based cueing used was of two types, namely avatar-mediated and environment-triggered. The avatar-mediated cueing prompts were (i) gaze-based cueing (that is gaze orientation), (ii) head turn (with gaze orientation) and (iii) pointing finger (along with head turn and gaze orientation) towards the objects of interest. The environment-triggered cueing involved sparkling of the targeted objects (cued by the avatar). We added the environment-triggered cueing since literature indicates a preference for sparkling/shiny objects among individuals with ASD (Coulter, 2009). Here, we wanted to understand the potential of each JA cue (in isolation rather than delivering these sequentially) to have implications on one's JA skill-related functional outcomes irrespective of the order in which these cues are delivered. Thus, instead of offering the JA cues sequentially with gradually increasing prompt information, we offered the JA cues in a randomized manner to the participants. The randomization of the order of the JA cues was controlled in the manner (referred as randomized controlled manner) so that same type of JA cue should not offer consecutively and also any two consecutive delivery of JA cues should not satisfy hierarchical increasing prompt protocol (Jyoti & Lahiri, 2019). Additionally, we also wanted to understand the ease with which an individual can pick up a JA cue. Thus, added to the functional measures, we wanted to study physiology while an individual interacted with our VR-enabled HCI-based JA task platform. This is because children with ASD often possess communication deficits and physiology (e.g., cardiac activity and electrodermal activity) can offer an undiluted measure (independent of the communication vulnerabilities) of the ease (Kushki et al., 2013) with which a JA cue is picked up. In our present research, we have chosen one's cardiac activity and electrodermal activity that was recorded using non-intrusive methods e.g., Pulse Plethysmogram (PPG) (Selvaraj, Jaryal, Santhosh, Deepak, & Anand, 2008) and Electrodermal Activity (EDA) (Critchley, 2002), respectively.

Thus, the objectives of our research were three-fold, namely to (i) design a VR-enabled HCI-based interactive JA task platform offering avatar-mediated and environment-triggered JA cues (ii) carry out a study with both typically developing (TD) children and children with ASD and (iii) study the comparative potential of different types of JA cues to have implications on one's JA skill-related functional and physiological measures when delivered in randomized controlled manner.

The rest of the paper is organized as: Section 2 presented materials and methods, section 3 discussed results of our study and section 4 covers detailed discussion and finally section 5 concludes our study with future scope.

2. Material and methods

2.1. Participants

A total of forty age-matched participants, twenty typically developing (TD *henceforth*) (TD1–TD20) ($Group_{TD}$ *henceforth*) and rest children with ASD (ASD1-ASD20) ($Group_{ASD}$ *henceforth*) volunteered for the study. The participant characteristics of the $Group_{TD}$ and $Group_{ASD}$ are shown in Table 1. The participants belonging to $Group_{TD}$ and $Group_{ASD}$ were recruited from a neighboring regular school and a mental health institute, respectively. All the enrolled participants were comfortable with the usage of cellular phone-based touch screens. They were assessed on their autism measures and JA ability using standard scales.

2.2. Assessment scales used

We used two assessment scales [Social Responsiveness Scale (SRS) (Coon et al., 2010) and Social Communication Questionnaire (SCQ) (Chandler et al., 2007)] to check whether a participant was in the autistic range. Since the present study was regarding JA skill, we also used the Childhood Joint Attention Rating Scale (C-JARS) (Mundy et al., 2017) to check the participants' JA skill level before taking part in our research study. Three of these assessment scales are in the form of parent reports.

2.2.1. Social Responsiveness Scale (SRS)

The Social Responsiveness Scale (SRS) consisted of 65 questions, and this scale is used to assess the severity of autism. Scores between 60 and 65, 66 and 75, and ≥76 suggest mild, moderate, and severe autism, respectively. The cutoff score is 59 points. This scale evaluates one in social communication, social awareness, and anxiety in social situations. The answers are given on a 4-point rating scale which ranges from "not true" (i.e., 1) to "almost always true" (i.e., 4). From Table 1, we can see that ASD3 was with severe autism, ASD1, ASD2, ASD4, ASD6, ASD9, ASD12, ASD13, ASD14, ASD16, ASD17, ASD19 were with moderate autism and rest were in the mild autism range. In contrast, TD participants ($Group_{TD}$) were well below the cutoff score of SRS.

2.2.2. Social Communication Questionnaire (SCQ)

The Social Communication Questionnaire (SCQ) is a questionnaire-based scale consisted of 40 questions that are used to assess the presence of ASD symptoms in an individual. The cutoff score is 15 points that evaluate one's social communication, repetitive behavior, and reciprocal interaction skills. The answers can be given using "yes" or "no" options. It can be seen from Table 1 that all the participants from $Group_{TD}$ were below the cutoff, unlike the participants belonging to the $Group_{ASD}$.

2.2.3. Childhood Joint Attention rating scale (C-JARS)

We used C-JARS to get an estimate of one's JA skill when the individual came in for the study. It consists of 55 questions with the final scores lying between −4 and +4. A positive score suggests the presence of JA skill (depends on score; high score represents improved JA skill) and vice-versa for the negative score. As can be seen from Table 1, 16 out of 20 participants from $Group_{ASD}$ had negative C-JARS score unlike that of the $Group_{TD}$ for which the score was positive for all the participants (see Table 1).

As our data was not normally distributed for SRS, SCQ, and C-JARS scores (based on the test of normality, i.e., Shapiro-Wilk test (Shapiro & Wilk, 1965)). Thus we conducted an independent sample non-parametric test, i.e., Wilcoxon rank-sum test (Wilcoxon, 1946) on the age, SRS, SCQ, and C-JARS scores for between groups comparison. This statistical test showed that the $Group_{TD}$ and $Group_{ASD}$ were statistically different on SRS [$z = -5.42$, $p = 0.000 < 0.01$, $r = 0.857$], SCQ [$z = -5.458$, $p = 0.000 < 0.01$, $r = 0.863$] and C-JARS [$z = -5.425$, $p = 0.000 < 0.01$, $r = 0.858$] scores with large effect sizes (r) (Field, 2009) but not on age [$z = -0.924$, $p = 0.355 > 0.05$, $r = 0.146$].

2.3. System design

Our VR-enabled HCI-based JA task system comprised of three modules, namely (1) JA task presentation (2) Real-time physiological data acquisition and storage (3) Handshake modules. Fig. 1 shows a block

Table 1
Participant's characteristics.

ID(G)	Age(Yrs.)	SRS[a]	SCQ[b]	C-JARS[c]	ID(G)	Age(Yrs.)	SRS[a]	SCQ[b]	C-JARS[c]
$Group_{TD}$					GroupASD				
TD1 (M)	7	45	3	3.1	ASD1 (M)	5	70	18	−1.8
TD2 (M)	8	48	3	3.2	ASD2 (F)	8.5	69	16	−1
TD3 (F)	6	42	2	3.5	ASD3 (M)	6	78	19	−1.8
TD4 (M)	7	48	4	3.3	ASD4 (F)	5.5	71	20	−1.2
TD5 (F)	7	41	1	3.4	ASD5 (M)	8.5	65	17	−0.2
TD6 (M)	8	42	2	3.3	ASD6 (M)	5	74	21	−2
TD7 (M)	8	41	4	3.2	ASD7 (F)	6	63	16	−0.5
TD8 (M)	8	43	2	3.5	ASD8 (M)	7	65	16	−0.3
TD9 (M)	6	43	2	3.3	ASD9 (M)	7	67	18	−0.8
TD10 (M)	6	49	2	3.5	ASD10 (F)	6	62	17	0.5
TD11 (M)	6	43	1	3.3	ASD11 (M)	7	60	17	0.6
TD12 (M)	7	45	3	3.6	ASD12 (F)	4.5	67	18	−0.4
TD13 (F)	7	43	2	3.4	ASD13 (M)	6.5	68	20	−0.9
TD14 (M)	6	45	3	3.1	ASD14 (M)	7	70	18	−0.9
TD15(M)	6	42	3	3.2	ASD15 (M)	6	61	16	0.7
TD16(F)	6	44	2	3.1	ASD16 (M)	4.5	68	18	−0.8
TD17(F)	6	40	2	3.2	ASD17 (M)	8	72	19	−1.5
TD18(M)	7	39	1	3.5	ASD18 (F)	8	65	17	−0.2
TD19(F)	7	39	1	3.5	ASD19 (F)	6	71	19	−1.8
TD20(F)	5	41	3	3.1	ASD20 (M)	6	65	17	0.1
Mean (SD)	6.7 (0.865)	43.15 (2.85)	2.3 (0.92)	3.32 (0.163)	Mean (SD)	6.4 (1.23)	67.55 (4.5)	17.85 (1.46)	−0.71 (0.82)

Note: SD= Standard Deviation, M = Male, F=Female.
[a] Cutoff = 59.
[b] Cutoff = 15.
[c] Range = (−4 to+4), ID= Participant ID, G = Gender, Yrs. = in Years.

schematic diagram of our system.

While the participants interacted with our system offering the JA cues (using the JA task presentation module), their real-time physiological data was acquired using Biopac MP150 (Biopac Systems Inc.) operated in the wireless mode. The physiological data was stored in a database of the data logger computer. Corresponding to each JA task trial, our system sent event-related markers to the data logger computer used for physiological data acquisition and storage using a handshake module so as to synchronize the physiological data acquisition with the JA task progression. In every JA task trial, the participant was asked to respond by using a touch-sensitive monitor. In turn, our system recorded performance score, response time, and distribution of screen taps (i.e., finger contacts) on the touch-sensitive screen over different regions of interest along with physiological signals for each JA task trial.

2.3.1. JA task presentation module

In our present research, we used desktop-based HCI to project eight JA task trials. Each JA task trial is offering a different type of JA cue (discussed below in Section 2.3.1.5 (Step3)) with varied information content to the participant. This module consisted of various components namely (i) Virtual JA task environment (ii) VR-based target and non-target objects (VR_{OBJ} *henceforth*) and (iii) Virtual humanoids (Avatars *henceforth*) as shown in Fig. 2. Each component was designed in python-based Vizard platform (from Worldviz Inc.) (For details, see Sections 2.3.1.1-2.3.1.3 discussed below).

2.3.1.1. Virtual JA task environment. Our virtual JA task environment was projected as a virtual room (VR_{Room}) consisting of two shelves (center-left and center-right) hanging from the walls (left and right), as shown in Fig. 2. An avatar stands in the center of the VR_{Room} with virtual objects placed on the shelves.

2.3.1.2. VR-based target and non-target objects (VR_{OBJ}). The Vizard software that we used for our study comes with a very limited set of virtual objects; we needed to design a database of VR_{OBJ} for our study. These were designed using Google SketchUp [http://www.sketchup.com] and were exported to the Vizard environment. The VR_{Room} displayed VR_{OBJ}, and these objects were categorized as target and non-target objects based on whether the VR_{OBJ} being cued or not. The VR_{OBJ} were presented as 3-D objects projected on the 2-D desktop touch-sensitive monitor. Examples of VR_{OBJ} used in our study were lock, bottle, clock, etc. That we commonly used in daily living. The VR_{OBJ} were neutral in nature and chosen from The Bank of Standardized Stimuli (BOSS) dataset (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010). Before using the VR_{OBJ} in our study, we ensured the familiarity and neutrality of the VR_{OBJ} through a survey conducted among TD children [n = 32; age [Mean (SD) = 9.15(1.83) years)] recruited from the neighborhood. From the designed database, we used those VR_{OBJ} in our study that was marked as familiar and neutral ≥95% of the time.

2.3.1.3. Virtual humanoids (avatars). In our present study, avatars were used as facilitators who administered the VR-enabled HCI-based JA tasks. The idea of using humanoid characters as facilitators was to bring in the element of embodiment and real-life experience in the JA tasks. This is important since literature indicates that the inclusion of such characters is often helpful in teaching social skills to individuals with ASD (Kuriakose & Lahiri, 2015). Again, since our avatars were meant to be interacting with children from India, we wanted the avatars to look like Indians. The faces of the avatars were modeled using 2-D photographs taken from an online database of Asian faces [http://wiki.cnbc.cmu.edu/Face_Place] each of them demonstrating a neutral expression. For mounting these faces on 3-D heads, we used facegen modeller software [https://facegen.com/-modeller.htm]. Using this software, we modified the texture and the color of the skin of the face. Also, we edited the eyeball and hair color for giving the avatar an Indian look. For this study, we designed eight avatars (4 Males and 4 Females). Out of this, we chose four avatars (2 Males and 2 Females) based on a survey conducted among TD children [n = 41; age (Mean (SD) = 7.6(1.6) years)] recruited from the neighborhood schools. Also, the avatars were designed to be capable of blinking, performing animations, and speaking. As far as animations were concerned, the avatars could change the direction of their eye gaze, make gestures by turning head and/or raising hands, pointing fingers necessary for delivering avatar-mediated JA cues. Also, the avatars could speak in a lip-synched manner with audio files that were pre-recorded in three languages (English, Hindi, and Gujarati (regional language)).

2.3.1.4. Regions of interest (ROIs). The virtual environment was segmented into six regions of interest (ROIs) namely Target (ROI_1), Hand (ROI_2), Face (ROI_3), Torso (ROI_4), Non-Target (ROI_5) and

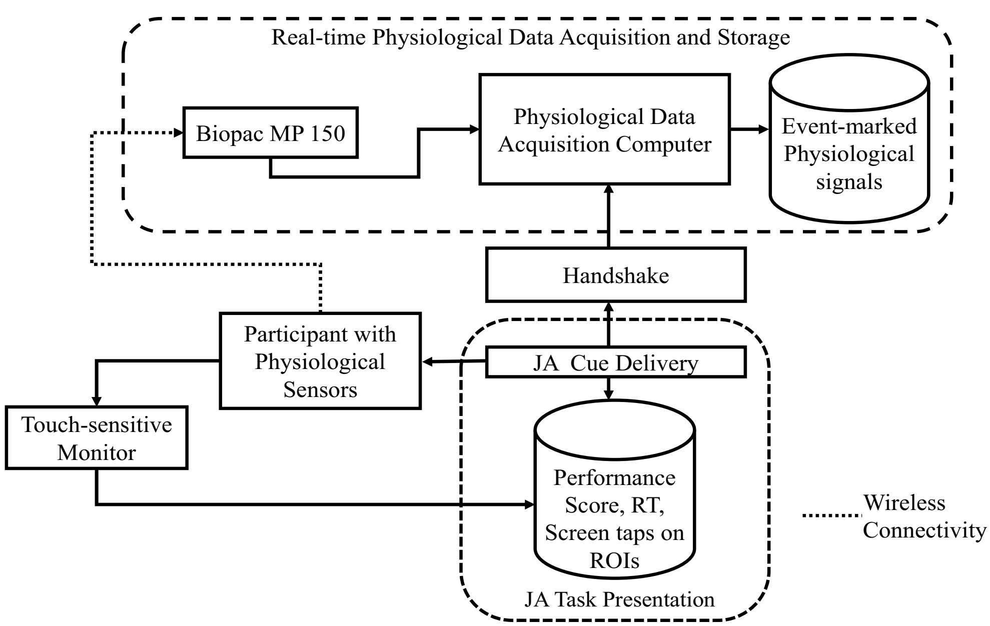


Fig. 1. Block schematic diagram of VR-enabled HCI-based JA task system. *Note*: PPG=Pulse plethysmogram; EDA = Electrodermal activity, RT = Response time, JA = Joint attention, ROIs = Regions of interest.

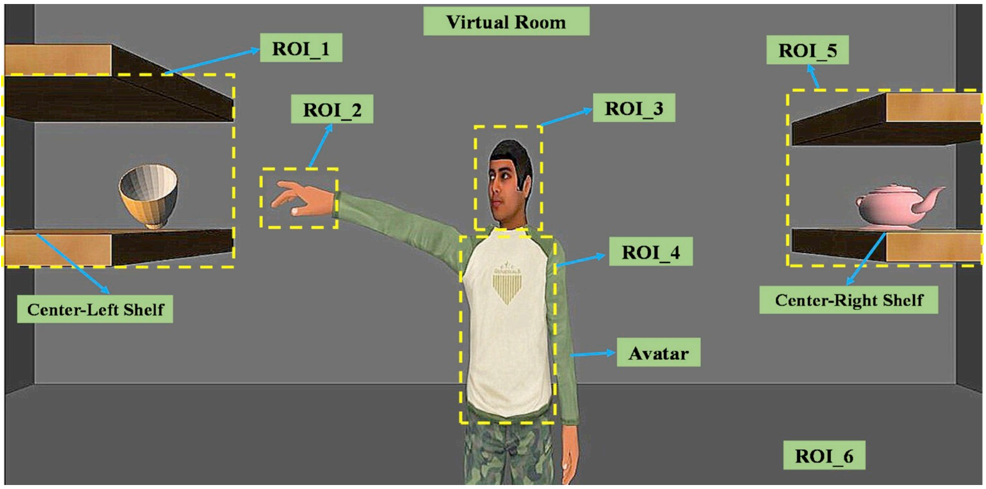


Fig. 2. Virtual Environment with segmented ROIs when Target VR_{OBJ} was placed on Center-Left Shelf.

Outside_ROIs (ROI_6) as shown in Fig. 2. Thus ROI_1 and ROI_5 were used to represent the cued VR_{OBJ} and the non-cued VR_{OBJ}, respectively (see Fig. 2). The ROI_2 to ROI_4 belonged to the avatar. The idea behind this segmentation was to record the distribution of one's finger contacts on the touch-sensitive monitor (screen taps *henceforth*) in response to JA cues. A screen tap was considered as "*relevant*"' if one touched the Target ROI; else, it was considered as "*irrelevant*" screen tap.

2.3.1.5. Steps for administration of the VR-enabled HCI-based JA tasks. The VR-enabled HCI-based JA task administration involved five steps (i) Step1: Participant information (ii) Step2: Introduction (iii) Step3: JA cue delivery (iv) Step4: Participant's response and (v) Step5: System-generated feedback**.** These steps are discussed below.

Step1: Participant Information.

This step was used to record participant-specific information, such as name, age, gender, school, etc. Also, this was used to record the participant's preferred language out of the available three options (English/Hindi/Gujarati). Our system administers JA tasks in the preferred language chosen by the participant in this step.

Step2: Introduction.

Our VR-enabled HCI-based JA task platform projected different JA task trials. The Step2 were of two types (Step2_1 applicable for the first trial and Step2_2 for the rest). Both the Step2_1 and Step2_2 (with the preferred language) included the avatar greeting the participant and introducing himself or herself. Additionally, Step2_1 included the avatar narrating the context and purpose behind the JA tasks to the participant. Given different objects (VR_{OBJ}) in the VR_{Room}, the purpose of the JA tasks were to find an object in the VR_{Room} that the avatar will prompt using different types of JA cues.

Step3: JA Cue Delivery.

This step began with the avatar standing in the middle of the VR_{Room} and looking down followed by looking towards the participant to establish eye contact (similar to one of the studies (Neufeld, Ioannou, Korb, Schilbach, & Chakrabarti, 2016)) followed by delivery of JA cue. The types of JA cues (Cue_{Type1} to Cue_{Type4}; differ by the information content delivered by the avatars increasing from Cue_{Type1} to Cue_{Type4} (see Table 2)). The delivery of JA cues was chosen in a randomized controlled manner from the JA cues database (stored in the backend) with each JA cue delivered twice (in the eight JA task trials) and by different avatars. Avatar changes from trial to trial (as we used a total of four avatars (2 males and 2 females)). The visualization of the four types of JA cues delivered is shown in Fig. 3. The duration of each cue was chosen as 5 s (similar to use in conventional methods used for JA skill training (Taylor & Hoch, 2008)).

Step4: Participant's Response.

Table 2
Types of JA Cues.

JA Cues	Information Content in JA Cue delivered by an Avatar
Cue_{Type1}	Gaze-based cue
Cue_{Type2}	Gaze-based cue + Head turn
Cue_{Type3}	Gaze-based cue + Head turn + Finger-pointing
Cue_{Type4}	Gaze-base cue + Head turn + Finger-pointing + Sparkling target VR_{OBJ}

After a JA cue has been delivered by an avatar, the participant was expected to select cued VR_{OBJ} by touching the target VR_{OBJ} (i.e., ROI_1 in example shown in Fig. 2, i.e., *relevant* screen tap) using the touch-sensitive monitor. Though our system was programmed to be capable of recording all the screen taps, it did not react to all screen taps except for taps on ROI_1 (Target) and ROI_5 (Non-Target) in our example (see Fig. 2) (For details on feedback, see Step5). The maximum duration of each task trial was 10 s (chosen as an initial approximation). In case the participant did not respond within 10 s, then our system presented the next task trial. Corresponding to each screen tap by the participant, our system stored the various task execution-related functional measures, i.e., performance score, response time, and ROIs tapped by the participant (for details, see Section 2.5.1) in a time-synchronized manner for offline analysis.

Step5: System-generated Feedback.

After the participant's response, our system delivered feedback. Specifically, only, if a participant selected either of the VR_{OBJ} (i.e., Target (ROI_1) or Non-Target (ROI_5) VR_{OBJ}), our system generated an audio feedback (monotone of duration 0.5 s) and displayed a dialog box "Do you want to continue with the current selection?" along with "Yes" and "No" options to make sure that the participant had intentionally selected the VR_{OBJ}. If a participant selected the "Yes" our VR-enabled HCI-based JA system offered the next task trial. However, if a participant had selected the "No" option, then our system allowed the participant to make another selection until the task trial duration of 10 s was over. Once all the task trials were over, a message of thanks ("Thank you for your time") appeared on the desktop monitor screen.

2.3.2. Real-time physiological data acquisition and storage module

While a participant took part in the JA task trials, his/her Pulse Plethysmogram (PPG) and Electrodermal Activity (EDA) were recorded using the Biopac MP150 in the wireless mode and with a sampling frequency of 1000 Hz. The data was stored in a data logger computer synchronized with the progression of JA task trials. The PPG and EDA data acquisition setup (with light-weight sensors (see Fig. 4 (a)) along with a snapshot of a typical PPG and EDA waveforms are shown in Fig. 4 (b).

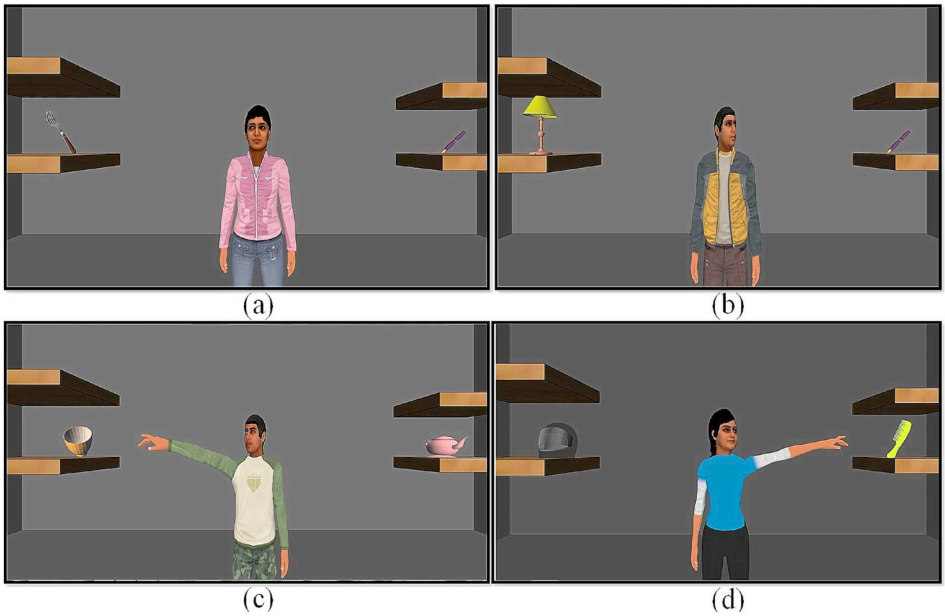


Fig. 3. Types of JA Cues (a) Cue_{Type1}, (b) Cue_{Type2}, (c) Cue_{Type3} and (d) Cue_{Type4} delivered by different avatars.

2.3.2.1. Physiological measures used. We used two physiological measures, i.e., PPG and EDA (described below). Our aim was to analyze the physiological measures to get an estimate of the variations in one's Autonomic Nervous System (ANS) activation that can help to quantify one's ease of understanding (Billeci et al., 2018) of a JA cue, important for social functioning while attending to VR-enabled HCI-based JA task. This is because easiness in understanding a task can be a critical component of effective skill training.

2.3.2.1.1. Pulse Plethysmogram (PPG). The PPG is a non-invasive technique to record the blood volume pulse waveform and can indicate one's cardiac activity. The literature review indicates that a reduction in cardiac activity represents ease in approach to a task along with adaptive social behavior (Billeci et al., 2018). Again, the PPG can be used to measure the changes in ANS activity (Gil et al., 2010) that in turn, is linked to the social functioning of individuals with ASD. In our study, the PPG signal was acquired by placing the sensor on the distal phalange of the middle finger of one's non-dominant hand (see Fig. 4(a)). Subsequently, we extracted the mean pulse rate (PR_{MEAN}) from the PPG signal (for details, see Section 2.3.2.2.1) as one of the physiological indices.

2.3.2.1.2. Electrodermal activity (EDA). The EDA is a non-invasive measure of skin conductivity and this is also used as a psychophysiological measure of ANS activity (Critchley, 2002). Reduced EDA has been reported to be linked with promoting adaptation to social needs and improved communication (Billeci et al., 2018). The EDA sensors were attached to the distal phalange of the ring and index fingers of one's non-dominant hand (see Fig. 4(a)). Subsequently, we extracted mean tonic activity ($Tonic_{MEAN}$) from the EDA signal (for details, see Section 2.3.2.2.2) as one of the physiological indices.

2.3.2.2. Extraction of physiological indices. The raw PPG and EDA signals were processed to extract PR_{MEAN} and $Tonic_{MEAN}$, respectively.

2.3.2.2.1. Extraction of PR_{MEAN}. Before extraction of the PR_{MEAN}, the PPG signal was pre-processed. For this, thresholding was applied to the raw PPG signal to remove noise. Then the Pulse Rate (PR) was extracted from the segments of the pre-processed PPG signal by measuring the inter-beat interval between the corresponding peaks of the PPG waveform. Subsequently, we computed the mean Pulse Rate (PR_{MEAN}) for each type of JA cue for each participant.

2.3.2.2.2. Extraction of $Tonic_{MEAN}$. The EDA signals were pre-processed through thresholding to remove noise spikes. Finally, we extracted the Tonic Mean ($Tonic_{MEAN}$) for each type of JA cue for each participant.

2.3.3. Handshake module

This module consisted of an Arduino Due-based hardware unit to integrate the JA task computer (meant for JA task presentation) with the data logger computer (used for storing the physiological signals (PPG and EDA) through the Biopac MP150). The purpose of using this module was to route event-related markers triggered by the JA task computer to the data logger computer for synchronizing the physiological data with the progression of the JA task trials.

2.4. Experimental setup and procedure

We designed the VR-enabled HCI-based JA task platform for

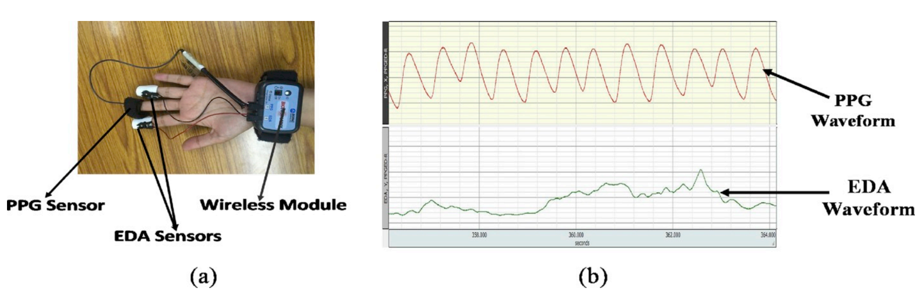


Fig. 4. (a) Sensor setup for physiological data acquisition (b) Typical PPG and EDA waveforms.

delivering different types of JA cues in a randomized controlled manner, as discussed in Section 2.3.1.5 (Step5).

2.4.1. Experiment setup

The experimental setup is shown in Fig. 5. We used a touch-sensitive monitor (C1) for administering the JA tasks. Also, a physiological data acquisition module (Biopac MP150 from Biopac Inc. and operated in the wireless mode) was used to acquire one's real-time physiological signals (PPG and EDA). A data logger **computer (C2) was** used to store one's PPG and EDA signals acquired via the Biopac MP150 for subsequent offline analysis using AcqKnowledge 4.3 software. For analysis, we needed to synchronize the physiological signals with the VR-enabled HCI-based JA task trials progression. For this, we used event-markers triggered using an Arduino Due-based interface (see Section 2.3.3). The study room was uniformly lit. The study followed institute ethics. The brief and understandable overview of protocol followed for this study is shown in Fig. 6.

2.4.2. Procedure

Our study required a commitment of approximately 20 min from each participant. Once a participant entered the study room, the introduction session began in which the experimenter introduced herself to the participant and asked both the participant and his/her caregiver to sit down on chairs. Then the experimenter shared her thoughts with the participant to make him/her comfortable. Following this, the experimenter explained the task to the participant and showed the experimental setup. While describing the nature of the study, the experimenter used a visual schedule along with verbal narration to explain the tasks to the participant. Thereafter, the experimenter told the participant that there would be avatars who will appear on the computer screen and invite him/her to find an object of interest inside a room presented on the screen. The participants were asked to think the avatars as their virtual peers. Additionally, the experimenter showed a demo of the task on the touch screen monitor (of the JA Task computer) and asked the participant to choose an object by using his/her finger. Following this, the experimenter handed over the light-weight electrodes (PPG and EDA) to the participant and helped the participant in attaching the sensors on his/her fingers. If the participant was comfortable with the sensor for approximately 5 min, then only the experimenter considered the participant's participation in the study. This introduction and acclimatization session took ~10 min. All the participants were told that they could leave the experiment in-between if they are not feeling comfortable. Once the participant is ready, written consent was taken from the caregiver for the participation. After this, the caregiver was asked to leave the study room. The experimenter explained to the participant that he/she needs to respond to the avatar's question by tapping the object of interest on the touch-sensitive monitor. Upon getting a "yes" nod from the participant indicating that he/she was ready to start to interact with the system, the experimenter attached the PPG and EDA sensors on the middle finger (PPG) and index and ring fingers (for EDA) of the non-dominant hand followed by a 3 min baseline recording. The experimenter took her seat behind the participant so as not to disturb him/her during the JA tasks. Depending on the participant's preference, the experimenter chose the language (out of the options of English/Hindi/Gujarati) that will be verbally used by the avatars while delivering the JA cues. This was followed by the execution of the VR-enabled HCI-based JA task trials. The avatars used the JA cue delivery module (see Section 2.3.1.5 (Step3)) to present JA cues to the participants in a randomized controlled manner. After the participant completed the JA task trials, the experimenter administered post-study feedback survey from the participant.

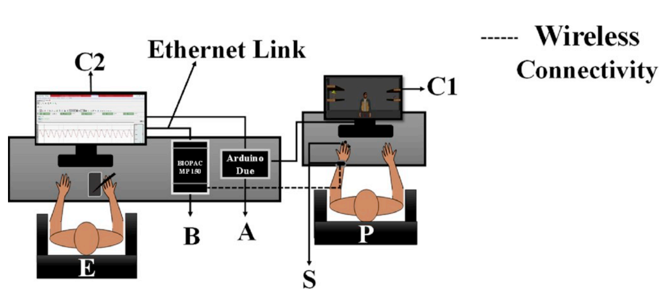


Fig. 5. Experimental Setup. *Note: C1 = Touch-sensitive JA Task Presentation Computer, C2 = Physiological Data Acquisition and Storage Computer, E = Experimenter, P = Participant, S = Physiological Sensors, A = Arduino Due, B = Biopac system.*

2.5. Computation of indices used for data analysis

While the participants interacted with the VR-enabled HCI-based JA tasks, we wanted to understand the implications of various JA cues on one's explicit functional measures such as performance, response time, etc. and implicit measures such as physiology. Thus, corresponding to each participant's response to each JA cue, we stored the task performance in terms of correct/incorrect responses, response time, screen taps along with physiological signals (PPG and EDA) in a time-synchronized manner. From this, we computed average performance score, average response time, and distribution of average % screen taps on different ROIs (see Section 2.5.1 below). Also, we computed the effectiveness index of each participant (described below in Section 2.5.2) based on his/her performance score and response time for different types of JA cues. From physiological signals, we computed the % change in PR_{MEAN} and $Tonic_{MEAN}$ w.r.t baseline condition.

2.5.1. Computation of average performance score, average response time and % screen taps on various ROIs

In our present study, we wanted to understand the implications of different JA prompting cues delivered by the avatars during the JA task trials. A correct response was awarded 100 points and 0 otherwise. One's response was considered as correct if one's choice of the VR_{OBJ} (participant's response) matched with the VR_{OBJ} cued by the avatar. As mentioned in section 2.3.1.5 (Step3), each type of JA cue was delivered twice in all eight JA task trials, i.e., trial 1 (t1) and trial 2 (t2). Subsequently, we computed the average performance score ($Perf_{AVG}$) and average response time (RT_{AVG}) over two task trials for each type of JA cue using eqs. (1) and (2) respectively.

$$Perf_{AVG} = \frac{Perf_{CueType\_i\_t1} + Perf_{CueType\_i\_t2}}{2} \tag{1}$$

$$RT_{AVG} = \frac{RT_{CueType\_i\_t1} + RT_{CueType\_i\_t2}}{2} \tag{2}$$

where i^{th} subscript means cue type, $Perf_{CueType\_i\_t1}$, $Perf_{CueType\_i\_t2}$ were performance score obtained in trial 1 and trial 2 respectively for each type of JA cue and $RT_{CueType\_i\_t1}$, $RT_{CueType\_i\_t2}$ was the time taken by a participant to respond to each type of JA cue in trial 1 and trial 2 respectively. The graphical user interface of our VR-enabled HCI-based JA system was segmented into 6 ROIs (see Section 2.3.1.4). While each participant responded to the avatar during the JA task trials by communicating his/her response through screen tapping, we computed the distribution of average % screen taps on various ROIs.

2.5.2. Computation of effectiveness index (EI)

The term Effectiveness Index (EI) has been used in literature to evaluate a user's interaction with a system (Cheng, Hu, & Heidorn, 2010). The evaluation is done based on performance score and response time while resolving the problem of speed-accuracy trade-off (Cheng et al., 2010). In our study, we wanted to understand the effectiveness with which various types of JA cues can help a participant to perform correctly and in the least possible time (response time). Also, we wanted to investigate how the effectiveness differ both within and across the

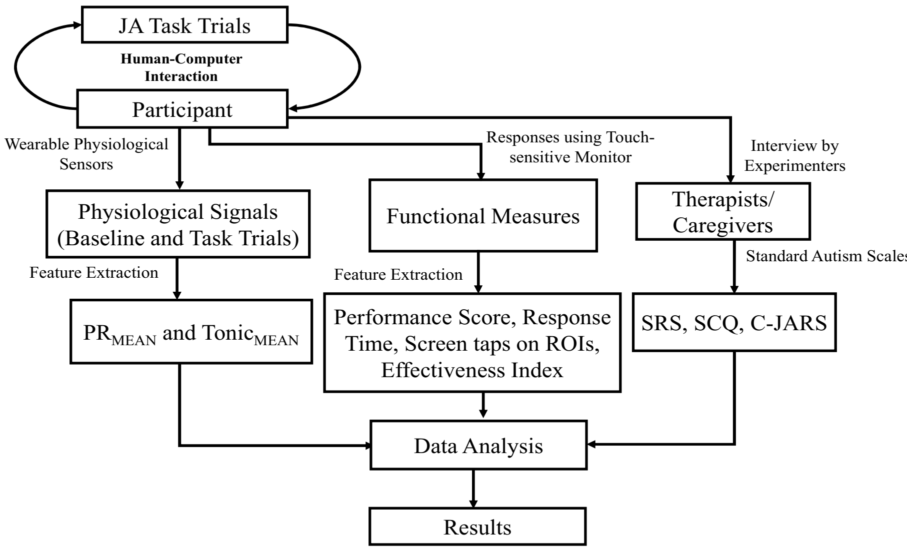


Fig. 6. Brief overview of the study protocol.

$Group_{TD}$ and $Group_{ASD}$ for different types of JA cues. The calculation of EI is based on two parameters $Perf_{AVG}$ and RT_{AVG}. For this, we normalized the performance score ($Perf_{AVG}$) and response time (RT_{AVG}) of each participant on a 0–1 scale and designated these as Performance Index (PI) and Response Time Index (RTI) using eqs. (3) and (4), respectively, for different types of JA cues. Thus the participant with the highest $Perf_{AVG}$ and RT_{AVG} will get PI = 1 and RTI = 1. We calculated the Effectiveness Index (EI) using eq. (5).

$$PI_J = (Perf_{AVG})_J \,/\, (Perf_{AVG})_{MAX} \tag{3}$$

$$RTI_J = (RT_{AVG})_J \,/\, (RT_{AVG})_{MAX} \tag{4}$$

$$EI_J = PI_J + (1 - RTI_J) \tag{5}$$

Where subscript J represents the J^{th} participant, $(Perf_{AVG})_{MAX}$, $(RT_{AVG})_{MAX}$ are the maximum values of $Perf_{AVG}$ and RT_{AVG} of our participant pool. The EI_J was calculated for each participant separately for each type of JA cue with the final EI value ranging from 0 to 2 (ideal case). Subsequently, we computed the average effectiveness index (EI_{AVG}) over two task trials, i.e., trial 1 (t1) and trial 2 (t2) for each type of JA cue using eq. (6).

$$EI_{AVG} = \frac{EI_{CueType\_i\_t1} + EI_{CueType\_i\_t2}}{2} \tag{6}$$

where i^{th} subscript means cue type, $EI_{CueType\_i\_t1}$ and $EI_{CueType\_i\_t2}$ were the effectiveness index obtained by the participant for each type of JA cue over two task trials t1 and t2, respectively.

2.5.3. Computation of % change in physiological indices with respect to baseline

While a participant interacted with the VR-enabled HCI-based JA task trials, we acquired his/her physiological signals (PPG and EDA) in a time-synchronized manner using AcqKnowledge 4.3 software that comes with Biopac MP150. From the acquired PPG signals, we computed the Pulse Rate (PR). Subsequently, we computed one's mean Pulse Rate (PR_{MEAN}) for each type of JA cue. Similarly, we extracted the Tonic Mean ($Tonic_{MEAN}$) from the acquired EDA signal. Subsequently, we computed % change in physiology w.r.t baseline. This was necessary since we wanted to nullify the effect of baseline variability while carrying out a comparative analysis of the implications of the varying JA cues on one's physiology both between and within the $Group_{TD}$ and $Group_{ASD}$. Similarly, for the baseline condition, we extracted the corresponding $PR_{MEANBaseline}$ and $Tonic_{MEANBaseline}$ for each participant. Thus, we computed % change in PR_{MEAN} for JA cue w.r.t baseline ($PR_{MEANw.r.tBaseline}$) and % change in $Tonic_{MEAN}$ w.r.t baseline ($Tonic_{MEANw.r.tBaseline}$) using eqs. (7) and (8), respectively.

$$PR_{MEANw.r.tBaseline} = \frac{PR_{MEAN(JA\_Cue\_i\ (t1\&t2))} - PR_{MEANBaseline}}{PR_{MEANBaseline}} X100 \tag{7}$$

$$Tonic_{MEANw.r.tBaseline} = \frac{Tonic_{MEAN(JA\_Cue\_i(t1\&t2))} - Tonic_{MEANBaseline}}{Tonic_{MEANBaseline}} X\ 100 \tag{8}$$

where i^{th} subscript means cue type, $PR_{MEAN(JA\_Cue\_i(t1\&t2))}$, $Tonic_{MEAN(JA\_Cue\_i(t1\&t2))}$ represents the average mean pulse rate and tonic mean for each type of JA cue, respectively over two task trials.

2.6. Statistical analysis

While we extracted various functional and physiological indices corresponding to each type of JA cue, we wanted to understand the statistical relevance of each of these indices. For this, we carried out statistical analysis using SPSS (Field, 2009). First, we conducted Shapiro-Wilk test of normality that showed that our extracted data for various functional and physiological indices was not normally distributed. Thus, we carried out the independent non-parametric test, i.e., Wilcoxon rank-sum test (for between groups comparison) and dependent non-parametric Wilcoxon signed-rank test (for within-group comparison for different combinations of JA cue types) for statistical evaluation of extracted functional and physiological indices using SPSS. We reported the z-statistic (as we used the non-parametric test) values of our conducted statistical analyses. We also computed effect size (r) for obtained statistical results.

3. Results

While the participants ($Group_{TD}$ and $Group_{ASD}$) interacted with our VR-enabled HCI-based JA task platform equipped with JA cue delivery module, the avatars exhibited different types of JA cues (Cue_{Type1}, Cue_{Type2}, Cue_{Type3}, and Cue_{Type4}) (see Table 2) in a randomized controlled manner. Subsequently, their performance, response time and screen taps (on the touch-sensitive monitor) were stored at the backend for subsequent offline analysis. Similarly, the physiological data was stored in the data logger computer in a time-synched manner for offline analysis. Here, we present our observations on the functional measures, e.g., average performance score, average response time, distribution of average % screen taps across different ROIs, effectiveness index along with physiological measures, i.e., $PR_{MEANw.r.tBaseline}$ and $Tonic_{MEANw.r.tBaseline}$ corresponding to each type of JA cue. While we analyzed both the functional and physiological data, we also wanted to understand the acceptability of our system, particularly by the target population.

3.1. Acceptability of our system

After the participants completed their interaction with our VR-enabled HCI-based JA tasks, we conducted a post-study feedback survey. The idea was to understand the participants' views on our system. Thus, they were asked whether they (i) were comfortable while doing the JA tasks (ii) found wearing the PPG and EDA sensors as convenient (iii) liked interacting with our system and (iv) were interested in interacting with our system again in future. All the participants communicated that they were comfortable interacting with the JA tasks offered by our system and did not report any inconvenience while wearing the PPG and EDA sensors. All the participants expressed that they liked the interaction with their virtual peers. Again, most of them enquired about the future possibilities of interacting with our system. In spite of given the option to withdraw from the study at any point in case of discomfort, all the participants completed the study. From this, we can say that our VR-enabled HCI-based JA task platform has the potential to be accepted by the target group.

3.2. Implications of various types of JA cues on participants' functional measures

3.2.1. Implications of various types of JA cues on average performance score ($Perf_{AVG}$)

Fig. 7(a) shows the average performance score ($Perf_{AVG}$) corresponding to each of the four types of JA cues (Cue_{Type1}, Cue_{Type2}, Cue_{Type3}, and Cue_{Type4}) for both the $Group_{TD}$ and $Group_{ASD}$. From Fig. 7(a) we can see a ceiling effect with $Perf_{AVG} = 100\%$ for the $Group_{TD}$ irrespective of the type of JA cue. However, for $Group_{ASD}$, the scenario was different. Specifically, for the $Group_{ASD}$, the $Perf_{AVG}$ was minimum for the Cue_{Type1}. Such an observation can be possibly attributed to children with ASD demonstrating atypical looking pattern characterized by reduced looking toward social stimulus such as the eye region of the avatar (Papagiannopoulou, Chitty, Hermens, Hickie, & Lagopoulos, 2014). However, the $Perf_{AVG}$ was better for Cue_{Type2} compared to Cue_{Type1}. This can be possibly due to the fact that the head turn cue (along with gaze orientation) was comparatively easier for the participants to be picked up (Leekam et al., 1998). For Cue_{Type3}, the cueing prompt being a culmination of head turn and finger-pointing, the cue was still further easier to be picked up by the children with ASD. Finally, added to the avatar-mediated cueing, when the system offered environment-triggered cueing (e.g., sparkling of an inanimate virtual object), the added information in the cueing prompt made it easiest for $Group_{ASD}$ to be pick up the cue resulting in maximum $Perf_{AVG}$ for Cue_{Type4}. Though highest, the $Perf_{AVG}$ for Cue_{Type4} was nearly similar to that for Cue_{Type3}. Also, please note that augmenting the cueing information with the sparkling of the target object is possible in a simulated world, but may not always be feasible in real-life settings. From the results, we can see that even in the absence of the Cue_{Type4}, the Cue_{Type3} can equip a child with ASD with an ability to perform satisfactorily (with $Perf_{AVG} = \sim 92.5\%$) in the JA task.

We wanted to understand whether the performance score varied statistically across various types of JA cues between and within groups. For carrying out a comparative analysis between participant groups, we conducted a Wilcoxon rank-sum test on $Perf_{AVG}$. Results (see ***Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)*** in Table 3) indicated that the $Perf_{AVG}$ between $Group_{TD}$ and $Group_{ASD}$ differed statistically for Cue_{Type1} and Cue_{Type2}. However, no statistical difference was observed for $Perf_{AVG}$ as far as Cue_{Type3} and Cue_{Type4} were concerned between $Group_{TD}$ and $Group_{ASD}$. This means that $Group_{ASD}$ performed equally well for Cue_{Type3} and Cue_{Type4} as their TD counterparts. Next, we carried out within the participant group statistical analysis. As far as the $Perf_{AVG}$ of $Group_{TD}$ was concerned, since there was a ceiling effect ($Perf_{AVG} = 100\%$) irrespective of the type of JA cue, we did not conduct any within group statistical analysis for the $Group_{TD}$. In contrast, since such ceiling effect in $Perf_{AVG}$ was not seen for the $Group_{ASD}$, we carried out a Wilcoxon signed-rank test on the $Perf_{AVG}$ between various JA cues combinations within $Group_{ASD}$ (see ***Within $Group_{ASD}$ (Between different combinations of JA Cue Types)*** in Table 3). A significant difference in the $Perf_{AVG}$ was found within all the combinations of types of JA cues except that between Cue_{Type3} and Cue_{Type4}, which implies that $Group_{ASD}$ performed equally well for Cue_{Type3} and Cue_{Type4}. For details on the statistical results, please refer to Table 3.

3.2.2. Implications of various types of JA cues on average response time (RT_{AVG}) for correct responses

Reduced social awareness often leads to an abnormal social response in children with ASD (Billeci et al., 2018) possibly delaying the

Table 3
Statistical analysis of $Perf_{AVG}$.

Between/Within (Groups and Cue Types)	z-statistic	p-value	Effect Size (r)
Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)			
Cue_{Type1}	−5.19	0.000**	0.82
Cue_{Type2}	−3.8	0.000**	0.6
Cue_{Type3}	−1.78	0.075◆	0.281
Cue_{Type4}	0.00	1◆	0
Within $Group_{ASD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−3.127	0.002**	0.494
Cue_{Type1}-Cue_{Type3}	−3.729	0.000**	0.589
Cue_{Type1}-Cue_{Type4}	−3.758	0.000**	0.594
Cue_{Type2}-Cue_{Type3}	−2.76	0.006**	0.436
Cue_{Type2}-Cue_{Type4}	−3.035	0.002**	0.479
Cue_{Type3}-Cue_{Type4}	−1.73	0.083◆	0.27

Note: *significantly different ($p < 0.05$), **significantly different ($p < 0.01$), ◆not significantly different ($p > 0.05$).

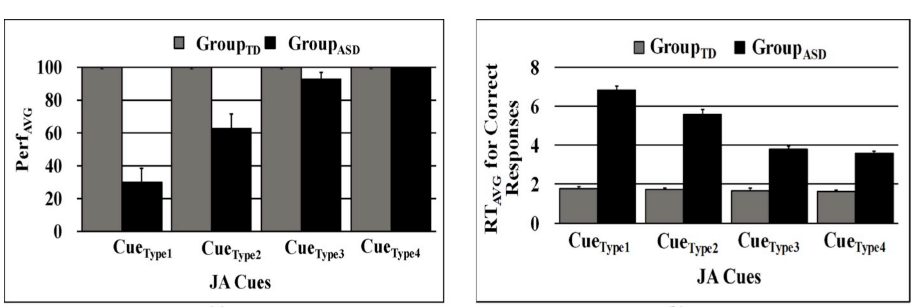


Fig. 7. (a) Comparative group analysis of average performance score ($Perf_{AVG}$) (b) Comparative group analysis of average reaction time for correct responses (RT_{AVG}).

spontaneity to pick up a cue. Thus, monitoring the response time of children with ASD while picking up JA cues is critical. Fig. 7(b) shows the RT_{AVG} for correct responses for each type of JA cue for the $Group_{TD}$ and $Group_{ASD}$. From Fig. 7(b), we find that all the TD participants demonstrated similar RT_{AVG} irrespective of the type of JA cue. In contrast, the $Group_{ASD}$ demonstrated an overall RT_{AVG} that was much higher (~3 times) than that of the $Group_{TD}$. This can be possibly attributed to the fact that children with ASD can be slower in picking up a prompt. Again, we see a reduction in the RT_{AVG} from Cue_{Type1} to Cue_{Type4} for $Group_{ASD}$, implying the positive contribution of increased information content in the cueing prompts with regard to the speed of response. Thus, reduction in response time (see Fig. 7(b)) coupled with improved performance score (see Fig. 7(a)) might possibly imply increased effectiveness in the JA task trials offering cueing prompts that are more informed. Again, similar to that in the case of $Perf_{AVG}$ (see Fig. 7(a)), we find that for the $Group_{ASD}$, the RT_{AVG} for Cue_{Type3} and Cue_{Type4} were nearly similar ($\Delta\% = 0.229\%$). In short, with Cue_{Type4} not being always feasible in real-life settings, the Cue_{Type3} can be the preferred modality for administering the JA tasks, at least during earlier stages of JA skill training.

We wanted to understand the statistical significance of each type of JA cue with regard to the RT_{AVG} both between and within groups. An independent sample Wilcoxon rank-sum test revealed a between groups statistical difference in the RT_{AVG} for all four types of JA cues with large effect sizes (r) (see ***Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)*** in Table 4). Again, for within group analysis, we carried out Wilcoxon signed-rank test, which showed no statistical difference in RT_{AVG} of $Group_{TD}$ for all the combinations of types of JA cues (see ***Within $Group_{TD}$ (Between different combinations of JA Cue Types)*** in Table 4) which implies that $Group_{TD}$ took equal time to respond to all types of JA cues. However, we found statistical differences while comparing the RT_{AVG} for all the combinations of types of JA cues for $Group_{ASD}$ (see ***Within $Group_{ASD}$ (Between different combinations of JA Cue Types)*** in Table 4) with large effect sizes (r) except that between Cue_{Type3} and Cue_{Type4} which implies that $Group_{ASD}$ took equal time to respond for both Cue_{Type3} and Cue_{Type4}. For details on the statistical results, please refer to Table 4.

3.2.3. *Implications of various types of JA cues on the participants' average effectiveness index (EI_{AVG})*

Fig. 8 shows the group comparison of average effectiveness index (EI_{AVG}) of participants belonging to both $Group_{TD}$ and $Group_{ASD}$ for each type of JA cue. It can be seen from this Fig. 8 that all TD participants had $EI_{AVG} > 1.5$ with their overall $Perf_{AVG}$ being high (100% for each task trial), and their overall RT_{AVG} was low (1.69 s for each task trial on an average) compared with those for the $Group_{ASD}$ (with an overall $Perf_{AVG} = 71.25\%$ and $RT_{AVG} = 4.94$ s). In contrast, children with ASD demonstrated varying EI_{AVG} across the different types of JA cues. For example, the EI_{AVG} for the $Group_{ASD}$ corresponding to Cue_{Type3} and Cue_{Type4} was greater than that for Cue_{Type2}, which in turn was greater than that for Cue_{Type1}. However, the difference in the EI_{AVG} between Cue_{Type3} and Cue_{Type4} was less ($\%\Delta = 0.096\%$) that might again indicate the optimality of using Cue_{Type3} for JA skill training. An independent sample Wilcoxon rank-sum test conducted on the EI_{AVG} of the $Group_{TD}$ and $Group_{ASD}$ for group comparison indicated that the EI_{AVG} differed statistically for all types of JA cues between groups with large effect sizes (r) (see ***Between $Group_{TD}$ and $Group_{ASD}$(Within Four Types of JA Cues)*** in Table 5). Again, a dependent sample Wilcoxon signed-rank test carried out on the EI_{AVG} of the $Group_{TD}$ indicated no statistical difference between various combinations of types of JA cues (see ***Within $Group_{TD}$(Between different combinations of JA Cue Types)*** in Table 5) which implies that $Group_{TD}$ performed with same effectiveness for all types of JA cues. In contrast, statistical difference in the EI_{AVG} with large effect size (r) was observed in the case of the $Group_{ASD}$ across various combinations of types of JA cues (see ***Within $Group_{ASD}$ (Between different combinations of JA Cue Types)*** in Table 5) except that between Cue_{Type3} and Cue_{Type4} which shows that both Cue_{Type3} and Cue_{Type4} are equally effective for JA skill training. For details on the statistical results, please refer to Table 5.

Table 4
Statistical analysis of RT_{AVG}.

Between/Within (Groups and Cue Types)	z-statistic	p-value	Effect size (r)
Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)			
Cue_{Type1}	−5.4	0.000**	0.855
Cue_{Type2}	−5.4	0.000**	0.855
Cue_{Type3}	−5.4	0.000**	0.855
Cue_{Type4}	−5.4	0.000**	0.855
Within $Group_{TD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−0.22	0.823◆	0.035
Cue_{Type1}-Cue_{Type3}	−0.34	0.737◆	0.05
Cue_{Type1}-Cue_{Type4}	−0.971	0.332◆	0.15
Cue_{Type2}-Cue_{Type3}	−0.64	0.53◆	0.1
Cue_{Type2}-Cue_{Type4}	−0.97	0.332◆	0.15
Cue_{Type3}-Cue_{Type4}	−0.485	0.627◆	0.077
Within $Group_{ASD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−3.36	0.001**	0.53
Cue_{Type1}-Cue_{Type3}	−3.92	0.000**	0.62
Cue_{Type1}-Cue_{Type4}	−3.92	0.000**	0.62
Cue_{Type2}-Cue_{Type3}	−3.92	0.000**	0.62
Cue_{Type2}-Cue_{Type4}	−3.92	0.000**	0.62
Cue_{Type3}-Cue_{Type4}	−1.42	0.156◆	0.22

Note: *significantly different($p < 0.05$), **significantly different($p < 0.01$), ◆not significantly different ($p > 0.05$).

3.2.4. *Implications of various types of JA cues on the distribution of average % screen taps on different ROIs*

While the participants interacted with our VR-enabled HCI-based JA task platform using a touch-sensitive interface, our system recorded the screen taps of the participants. As the graphical user interface of our system was segmented into various ROIs (see Section 2.3.1.4), we wanted to analyze the *relevant* and *irrelevant* screen taps on different ROIs. Fig. 9 shows the distribution of screen taps by both the $Group_{TD}$ and $Group_{ASD}$ in terms of average % screen taps over different ROIs corresponding to each type of JA cue. From Fig. 9, we find that for the $Group_{TD}$, 100% of the screen taps were *relevant* (see Section 2.3.1.4) irrespective of the type of JA cue. In contrast, for the $Group_{ASD}$, the number of *relevant* screen taps increased accompanied by a reduction in the *irrelevant* screen taps as the JA cue became more informative, i.e., moving from Cue_{Type1} to Cue_{Type4}. Again, the distribution of average % screen taps for the $Group_{ASD}$ corresponding to Cue_{Type3} and Cue_{Type4} on Target ROI were nearly similar ($\Delta\% = 1.67\%$) that echoes with the observation made in case of $Perf_{AVG}$, RT_{AVG}, and EI_{AVG} values.

3.3. *Implications of various types of JA cues on participants' physiological measures*

While the participants interacted with our VR-enabled HCI-based JA task platform, we acquired their physiological signals such as Pulseplethysmogram (PPG) and Electrodermal Activity (EDA) signals synched

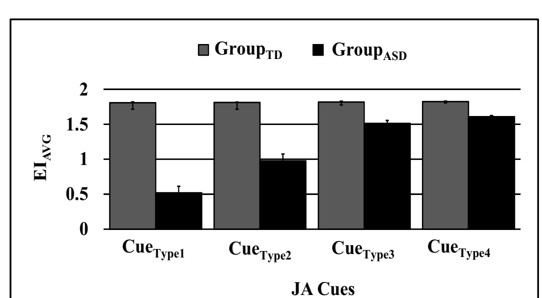


Fig. 8. Comparative group analysis of average effectiveness index (EI_{AVG}).

Table 5
Statistical analysis of EI_{AVG}.

Between/Within (Groups and Cue Types)	z-statistic	p-value	Effect size (r)
Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)			
Cue_{Type1}	−5.4	0.000**	0.855
Cue_{Type2}	−5.4	0.000**	0.855
Cue_{Type3}	−5.4	0.000**	0.855
Cue_{Type4}	−5.4	0.000**	0.855
Within $Group_{TD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−0.261	0.794◆	0.04
Cue_{Type1}-Cue_{Type3}	−0.336	0.737◆	0.05
Cue_{Type1}-Cue_{Type4}	−0.97	0.332◆	0.15
Cue_{Type2}-Cue_{Type3}	−0.635	0.526◆	0.1
Cue_{Type2}-Cue_{Type4}	−0.952	0.341◆	0.15
Cue_{Type3}-Cue_{Type4}	−0.485	0.627◆	0.076
Within $Group_{ASD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−3.92	0.000**	0.62
Cue_{Type1}-Cue_{Type3}	−3.92	0.000**	0.62
Cue_{Type1}-Cue_{Type4}	−3.92	0.000**	0.62
Cue_{Type2}-Cue_{Type3}	−3.92	0.000**	0.62
Cue_{Type2}-Cue_{Type4}	−3.92	0.000**	0.62
Cue_{Type3}-Cue_{Type4}	−1.9	0.057◆	0.3

Note: *significantly different($p < 0.05$), **significantly different($p < 0.01$), ◆not significantly different ($p > 0.05$).

with the JA task progression. Subsequently, we extracted physiological indices such as mean pulse rate (PR_{MEAN}) and tonic activity ($Tonic_{MEAN}$) from the PPG and EDA signals, respectively (see Section 2.3.2.2). Our aim was to analyze the physiological indices to get an estimate of one's ease of understanding of a JA cue while attending to VR-enabled HCI-based JA task. This is because easiness in understanding a task can be a critical component of effective skill training. Again, this becomes important from an interventionist's perspective since it helps to inform the clinician on the type of JA cue that can be easily understood by an individual.

3.3.1. Study of the feasibility of using $PR_{MEANw.r.tBaseline}$ to estimate the ease in understanding of different types of JA cues

We wanted to study the feasibility of using PR_{MEAN} (an indicator to one's cardiac activity) to estimate the ease in the understanding of a JA cue delivered by our VR-enabled HCI-based task platform. This is because the literature review shows that a reduction in one's cardiac activity indicates the ease in understanding a task (Billeci et al., 2018). Again, for the group comparison of PR_{MEAN} for each type of JA cue, we used the % change in PR_{MEAN} w.r.t. baseline ($PR_{MEANw.r.tBaseline}$) (see Section 2.5.3) to nullify the effect of baseline variability.

As can be seen from Fig. 10 (a), there was minimal variation in the $PR_{MEANw.r.tBaseline}$ irrespective of the type of JA cue for the $Group_{TD}$ as compared with the $Group_{ASD}$. The minimal variations in the $PR_{MEANw.r.tBaseline}$ for $Group_{TD}$ across the types of JA cues can be possibly attributed to the fact that all the JA cues were easily understood by the TD participants. However, for the $Group_{ASD}$, we see a different picture. Specifically, for the $Group_{ASD}$, we can see that the $PR_{MEANw.r.tBaseline}$ was highest for Cue_{Type1}, where the non-verbal component of the JA-related information was delivered only through the use of the eye gaze. Such an observation might infer that the $Group_{ASD}$ found it difficult to understand the Cue_{Type1} resulting in the least $Perf_{AVG}$ for this type of cue (see Fig. 7(a)). This difficulty in understanding the gaze-based cueing might be possibly due to their atypical looking pattern in which they tend to avoid looking towards the eye region of a social communicator (Papagiannopoulou et al., 2014). Further, with the JA cue becoming more informed, the $PR_{MEANw.r.tBaseline}$ showed a reducing trend corresponding to Cue_{Type1} to Cue_{Type3}. However, we see an increase ($\%\Delta = 1.86\%$) in $PR_{MEANw.r.tBaseline}$ from Cue_{Type3} to Cue_{Type4} (see Fig. 10(a)). This is unlike the observations on $Perf_{AVG}$, RT_{AVG}, and EI_{AVG} for the $Group_{ASD}$. We carried out an in-depth review of the observations recorded by the experimenter while conducting the study so as to find a possible reason behind this increase. The experimenter reported that when the Cue_{Type4} was presented to the participants, most of the children from $Group_{ASD}$ showed excitement along with smiling. While they could easily pick up this Cue_{Type4} and be successful in making the correct choice of the cued VR_{OBJ}, their excitement might have contributed to an increase (Piira, Huikuri, & Tulppo, 2011) in the $PR_{MEANw.r.tBaseline}$ while attending Cue_{Type4} as compared to Cue_{Type3}. While excitement (for Cue_{Type4}) might be beneficial for effective learning (Bonwell & Eison, 1991, pp. 20036–21183) resulting in improved performance (see Fig. 7(a)) with reduced response time (see Fig. 7(b)). We conducted an independent sample Wilcoxon rank-sum test to find whether there exist group differences with regard to the physiological measures. A significant group difference in $PR_{MEANw.r.tBaseline}$ was observed for all types of JA cues between both groups (see ***Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)*** in Table 6). With regard to the within group differences for different types of JA cues combinations, we conducted a dependent sample Wilcoxon signed-rank test on both groups separately. The $Group_{TD}$ did not exhibit a statistically significant difference in $PR_{MEANw.r.tBaseline}$ for all the combinations of types of JA cues (see ***Within $Group_{TD}$ (Between different combinations of JA Cue Types)*** in Table 6). This means that $Group_{TD}$ perceives all types of JA cues with the same ease. However, for $Group_{ASD}$, a significant difference was observed for $PR_{MEANw.r.tBaseline}$ between Cue_{Type3} and remaining three types of JA cues which means that $Group_{ASD}$ perceives Cue_{Type3} differently from other three types of JA cues. There might be the reason that $Group_{ASD}$ perceives Cue_{Type3} with more ease compared to other types of JA cues. However $PR_{MEANw.r.tBaseline}$ for Cue_{Type4} does not differ significantly from Cue_{Type1} and Cue_{Type2} (see ***Within $Group_{ASD}$ (Between different combinations of JA Cue Types)*** in Table 6). The reason might be the

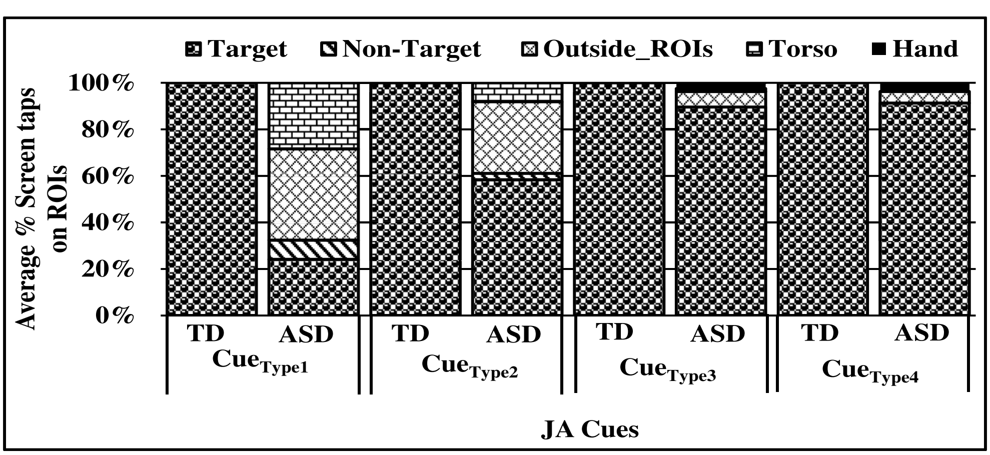
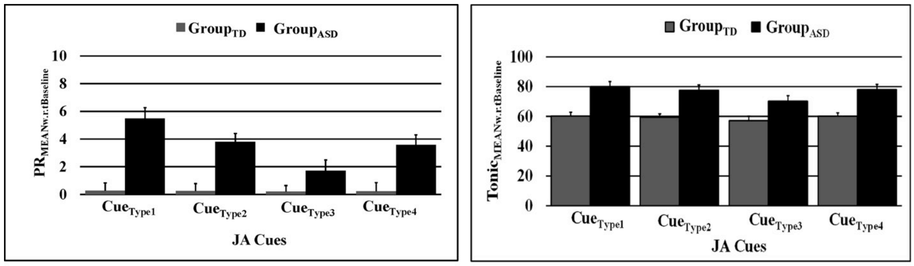


Fig. 9. Comparative group analysis of average % screen taps on ROIs. (*Note:* TD = $Group_{TD}$ and ASD = $Group_{ASD}$).



(a) (b)

Fig. 10. (a) Group analysis of $PR_{MEANw.r.tBaseline}$ (b) Group analysis of $Tonic_{MEANw.r.tBaseline}$.

Table 6
Statistical analysis of $PR_{MEANw.r.tBaseline}$.

Between/Within (Groups and Cue Types)	z-statistic	p-value	Effect size (r)
Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)			
Cue_{Type1}	−4.1	0.000**	0.64
Cue_{Type2}	−3.62	0.000**	0.57
Cue_{Type3}	−2.1	0.04*	0.33
Cue_{Type4}	−3.16	0.002**	0.5
Within $Group_{TD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−0.299	0.77◆	0.04
Cue_{Type1}-Cue_{Type3}	−0.56	0.58◆	0.088
Cue_{Type1}-Cue_{Type4}	−0.373	0.71◆	0.059
Cue_{Type2}-Cue_{Type3}	−0.71	0.478◆	0.11
Cue_{Type2}-Cue_{Type4}	0	1◆	0
Cue_{Type3}-Cue_{Type4}	−0.672	0.5◆	0.11
Within $Group_{ASD}$ (Between different combinations of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−2.0	0.04*	0.32
Cue_{Type1}-Cue_{Type3}	−3.3	0.001**	0.53
Cue_{Type1}-Cue_{Type4}	−1.94	0.052◆	0.31
Cue_{Type2}-Cue_{Type3}	−2.698	0.007**	0.425
Cue_{Type2}-Cue_{Type4}	−0.523	0.6◆	0.08
Cue_{Type3}-Cue_{Type4}	−2.13	0.033*	0.34

Note: **significantly different($p < 0.01$), *significantly different($p < 0.05$), ◆not significantly different ($p > 0.05$).

Table 7
Statistical analysis of $Tonic_{MEANw.r.tBaseline}$.

Between/Within (Groups and Cue Types)	z-statistic	p-value	Effect size (r)
Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)			
Cue_{Type1}	−3.59	0.000**	0.57
Cue_{Type2}	−3.22	0.001**	0.51
Cue_{Type3}	−2.14	0.033*	0.34
Cue_{Type4}	−3.43	0.001**	0.54
Within $Group_{TD}$ (Between different combination of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−1.045	0.296◆	0.16
Cue_{Type1}-Cue_{Type3}	−1.57	0.117◆	0.25
Cue_{Type1}-Cue_{Type4}	−1.08	0.279◆	0.17
Cue_{Type2}-Cue_{Type3}	−0.149	0.881◆	0.023
Cue_{Type2}-Cue_{Type4}	−0.523	0.601◆	0.082
Cue_{Type3}-Cue_{Type4}	−0.784	0.433◆	0.124
Within $Group_{ASD}$ (Between different combination of JA Cue Types)			
Cue_{Type1}-Cue_{Type2}	−1.31	0.191◆	0.21
Cue_{Type1}-Cue_{Type3}	−3.3	0.001**	0.52
Cue_{Type1}-Cue_{Type4}	−1.08	0.279◆	0.17
Cue_{Type2}-Cue_{Type3}	−2.86	0.004**	0.454
Cue_{Type2}-Cue_{Type4}	−0.037	0.97◆	0.005
Cue_{Type3}-Cue_{Type4}	−2.76	0.006**	0.44

Note: **significantly different($p < 0.01$), *significantly different($p < 0.05$), ◆not significantly different ($p > 0.05$).

excitement due to the presence of sparkling stimulus in Cue_{Type4} resulting higher $PR_{MEANw.r.tBaseline}$ as compared to Cue_{Type3}. For details on the statistical results, please refer to Table 6.

3.3.2. *Study of the feasibility of using $Tonic_{MEANw.r.tBaseline}$ to estimate the ease in understanding of different types of JA cues*

We also analyzed the $Tonic_{MEANw.r.tBaseline}$ extracted from the EDA signals. Fig. 10(b) shows the variation of $Tonic_{MEANw.r.tBaseline}$ for $Group_{TD}$ and $Group_{ASD}$ for each type of JA cue. From this Fig. 10(b), we see a minimal variation in the $Tonic_{MEANw.r.tBaseline}$ for $Group_{TD}$ across the different types of JA cues similar to that was observed for $PR_{MEANw.r.tBaseline}$ (see Fig. 10(a)). However, for the $Group_{ASD}$, we observe variations in the $Tonic_{MEANw.r.tBaseline}$ for different types of JA cues with the trend being similar to that in the case of the $PR_{MEANw.r.tBaseline}$.

An independent sample Wilcoxon rank-sum test on $Tonic_{MEANw.r.tBaseline}$ was carried out between the participant groups. The results indicate a significant difference in this physiological index for all four types of JA cues between both groups (see ***Between $Group_{TD}$ and $Group_{ASD}$ (Within Four Types of JA Cues)*** in Table 7). Again, a dependent sample Wilcoxon signed-rank test was carried out to find out the within group statistical difference for the $Tonic_{MEANw.r.tBaseline}$ for different combinations of types of JA cues. No significant within group difference for this physiological index was found for the $Group_{TD}$ corresponding to the various combinations of types of JA cues (see ***Within $Group_{TD}$ (Between different combinations of JA Cue Types)*** in Table 7). In contrast, for the $Group_{ASD}$, a significant difference was observed for $Tonic_{MEANw.r.tBaseline}$ between Cue_{Type3} and remaining three types of JA cues (see ***Within $Group_{ASD}$ (Between different combinations of JA Cue Types)*** in Table 7). This implies that $Group_{ASD}$ perceived Cue_{Type3} differently compared to the other three types of JA cues. Again Cue_{Type4} does not differ significantly from Cue_{Type1} and Cue_{Type2} similar to the findings for $PR_{MEANw.r.tBaseline}$. For details on the statistical results, please refer to Table 7. However, given the spectrum nature of autism, the implications of the different types of JA cues being differentiated, we do not intend to generalize our findings.

4. Discussion

One of the major objectives of our present research was to use a VR-enabled HCI-based interactive JA task platform offering avatar-mediated (i.e. Cue_{Type1} to Cue_{Type3} with gradually increasing JA-related information content) and environment-triggered (Cue_{Type4} having the maximum JA-related information content) JA cueing. The aim was to carry out a study with both TD children and children with ASD to understand the comparative potential of different types of JA cues to have implications on one's JA skill-related functional and physiological measures. In the conventional intervention, generally, the therapist begins with the cue (such as gaze-based cue) that might be most difficult for a child with ASD (due to atypical looking pattern (Frazier et al., 2017; Papagiannopoulou et al., 2014) followed by observing the ability of a child to pick up the cue, gradually increases the cue-related information content, if the child is unsuccessful in picking up the cue. Unlike the conventional approach, in our study, we offered presentation of the various types of JA cues (each with varying information content) in a randomized controlled manner. The idea was to see whether the amount of information content in the JA-related cue plays a vital role as far as the ability of a child with ASD to pick up the cue. The results of our study provide substantial insights into this aspect of JA-related cueing. As far as the functional measures are concerned, the JA task performance scores of participants with ASD showed improvement with an increase in information content in the JA cue when different JA cues were presented in a randomized controlled manner. In fact, the performance of the $Group_{ASD}$ was almost at par with that of the $Group_{TD}$ for Cue_{Type3} and same for the Cue_{Type4}. In contrast, the $Group_{TD}$ showed a ceiling effect across all the types of JA cues. Like the JA task performance score, similar was our findings for the response time, effectiveness index and the distribution of screen taps by the $Group_{ASD}$. However, though Cue_{Type4} (having a combination of object sparkling along with the other avatar-mediated cueing) is possible to be simulated in the VR-enabled HCI-based platform, this may not be feasible in real-world settings. Thus, as far as the performance, response time and distribution of screen taps were concerned, the Cue_{Type3} can be considered as the most powerful JA cue for children with ASD.

Again, for a better understanding of the underlying mechanisms that make the Cue_{Type3} as the most powerful JA cue for children with ASD, we used implicit measures such as one's physiological indices namely, mean pulse rate (PR_{MEAN}) and tonic mean ($Tonic_{MEAN}$), extracted from PPG and EDA signals. The idea was to analyze these indices to get an estimate of the variations in one's ANS activation that can help to quantify one's ease of understanding (Billeci et al., 2018) of a JA cue that can be important for effective JA skill training. Both the physiological indices showed that the $Group_{TD}$ were at ease for all the types of JA cues irrespective of the information content that is expected. However, for $Group_{ASD}$, our results showed a completely different picture. Among all the types of JA cues, the Cue_{Type3} was the one that was the easiest to decipher for the $Group_{ASD}$ as reflected from the least variation in mean pulse rate and tonic mean from their respective baseline condition.

Here we present the behavioral manifestations of the participants as observed by the experimenter during the study. For $Group_{TD}$, the experimenter did not report any observable behavioral changes while participating were interacting with VR-enabled HCI-based JA task platform. However, this was not the case with the participants belonging to the $Group_{ASD}$. As far as the Cue_{Type1} was concerned, the majority of participants in $Group_{ASD}$ missed attending to the cue, possibly due to the fact that they did not look towards the avatar's eyes. Also, some of them seemed confused when Cue_{Type1} is delivered and resulted in *irrelevant* screen taps. For example, faced with the Cue_{Type1}, ASD3 looked confused and made random screen taps and then looked at the experimenter seemingly appearing to ask experimenter for the right selection. As far as the Cue_{Type2} was concerned, ~45% of the $Group_{ASD}$ were able to pick up Cue_{Type2} in both trials (t1 and t2) unlike that in the case of Cue_{Type1} (only ~15% of the $Group_{ASD}$ were able to pick up Cue_{Type1} in both trials (t1 and t2)). For Cue_{Type3}, the participants with ASD seemed quite active in responding with considerably less *irrelevant* screen taps, as observed by the experimenter for $Cue_{Type3.}$ For example, ASD10 did not make any *irrelevant* screen tap while she was making a lot of *irrelevant* screen taps for Cue_{Type1} and Cue_{Type2}. As far the Cue_{Type4} was concerned, the experimenter could observe expressions of excitement in the $Group_{ASD}$. In fact, some of them were seen as smiling or clapping or both and sometimes looked towards the experimenter, possibly to share the excitement. Also, the experimenter reported that they seemed very active to select the cued target when Cue_{Type4} is delivered. Thus, considering one's ability to pick up a JA cue, spontaneity of response to a JA bid, possibly the easiness with which a JA cue can be picked up by $Group_{ASD}$ and last but not the least, the feasibility of generating the JA cue, the Cue_{Type3} might seem the recommended JA cue for administering VR-enabled HCI-based JA tasks.

5. Conclusion and future scope

In our present research, we have designed a VR-enabled HCI-based Joint Attention (JA) task platform. The JA task trials were administered by 3-D humanoid characters (Avatars) using various types of JA cues, namely avatar-mediated and environment-triggered cueing. The avatar-mediated cueing was of three types comprising of gaze-based cueing, a combination of head turn and gaze-based cueing and combination of head turn, gaze and finger-pointing-based cueing. The environment-triggered cueing offered sparkling from cued objects added to the head turn, gaze and pointing gesture-based cueing. Each cueing prompt was delivered audio-visually and in a randomized controlled manner to the participants. Both TD participants and participants with ASD took part. Results of our study indicate that considering aspects of (i) performance, (ii) speed of response, (iii) possible ease of understanding of the JA cue and (iv) feasibility in generating the cue in real-life situations, the combination of avatar-mediated gaze, head turn, and finger-pointing coupled with verbal cue is the most promising as far as JA task administration for children with ASD is concerned.

The results of our study are promising, yet our study was not devoid of limitations. The first limitation was the limited sample size. In the future, we plan to recruit more participants, both TD and those having ASD. Another limitation was the limited exposure to the system. Having understood the potential of VR-enabled HCI-based JA task platform to offer JA skill training, in future, we plan to add JA task modules that can offer JA skill training over a prolonged duration. Still, another limitation was that our present JA task platform was not integrated with an eye tracker. This is necessary, since tracking of the participant's gaze while doing the VR-enabled HCI-based JA tasks might give clues of why $Group_{ASD}$ tend to miss picking up the gaze-based cue. In the future, we plan to integrate an eye tracker with our VR-enabled HCI-based JA task platform to record gaze-based indices for different types of JA cues.

Although the preliminary results of our study are promising, yet, questions remain on the transferability of the JA skills learned from the simulated world to real-world settings. However, such a VR-enabled HCI-based JA task platform can be potentially used as a complementary tool in the hands of the interventionist and contribute to JA skill training for children with autism spectrum disorder.

Acknowledgments

We express our gratitude to participants and their families for their cooperation in this research study. The authors are thankful to the local regular government school, Gandhinagar, Pearl special needs school, Ahmedabad, and BM institute of mental health, Ahmedabad, for helping us to enroll the participants for our study. Also, we thank the TCS Ph.D. research fellowship and IIT Gandhinagar for supporting this research.

References

Amaral, C. P., Simões, M. A., Mouga, S., Andrade, J., & Castelo-Branco, M. (2017). A novel brain computer interface for classification of social joint attention in autism and comparison of 3 experimental setups: A feasibility study. *Journal of Neuroscience Methods, 290*, 105–115.

Andrist, S., Gleicher, M., & Mutlu, B. (2017, May). Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM.

Bakeman, R., & Adamson, L. B. (1984). *Coordinating attention to people and objects in mother-infant and peer-infant interaction* (pp. 1278–1289). Child development.

Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). Echoes: An intelligent serious game for fostering social communication in children with autism. *Information Sciences, 264*, 41–60.

Billeci, L., Narzisi, A., Campatelli, G., Crifaci, G., Calderoni, S., Gagliano, A., & Raso, R. (2016). Disentangling the initiation from the response in joint attention: An eye-tracking study in toddlers with autism spectrum disorders. *Translational Psychiatry, 6* (5), e808.

Billeci, L., Tonacci, A., Narzisi, A., Manigrasso, Z., Varanini, M., Fulceri, F., & Muratori, F. (2018). Heart rate variability during a joint attention task in toddlers with autism spectrum disorders. *Frontiers in Physiology, 9*.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom. 1991 ASHE-ERIC higher education reports. ERIC clearinghouse on higher education*. Washington, DC: The George Washington University, One Dupont Circle, Suite 630.

Bozgeyikli, L., Raij, A., Katkoori, S., & Alqasemi, R. (2017). A survey on virtual reality for individuals with autism spectrum disorder: Design considerations. *IEEE Trans Learn Technol, 11*(2), 133–151.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One, 5*(5), e10773.
Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews, 10*(3), 169–175.
Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Detecting communicative intent in a computerised test of joint attention. *PeerJ, 5*, e2899.
Caruana, N., Stieglitz Ham, H., Brock, J., Woolgar, A., Kloth, N., Palermo, R., et al. (2018). Joint attention difficulties in autistic adults: An interactive eye-tracking study. *Autism, 22*(4), 502–512.
Chandler, S., Charman, T., Baird, G., Simonoff, E., Loucas, T. O. M., Meldrum, D., & Pickles, A. (2007). Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(10), 1324–1332.
Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 315–324.
Chasson, G. S., Harris, G. E., & Neely, W. J. (2007). Cost comparison of early intensive behavioral intervention and special education for children with autism. *Journal of Child and Family Studies, 16*(3), 401–413.
Cheng, J., Hu, X., & Heidorn, P. B. (2010, October). New measures for the evaluation of interactive information retrieval systems: Normalized task completion time and normalized user effectiveness. In *Proceedings of the 73rd ASIS&T annual meeting on navigating streams in an information ecosystem* (Vol. 47, p. 72). American Society for Information Science.
Coon, H., Villalobos, M. E., Robison, R. J., Camp, N. J., Cannon, D. S., Allen-Brady, K., & McMahon, W. M. (2010). Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. *Molecular Autism, 1*(1), 8.
Coulter, R. A. (2009). Understanding the visual symptoms of individuals with autism spectrum disorder (ASD). *Optometry & Vision Development, 40*(3).
Courgeon, M., Rautureau, G., Martin, J. C., & Grynszpan, O. (2014). Joint attention simulation using eye-tracking and virtual humans. *IEEE Transactions on Affective Computing, 5*(3), 238–250.
Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. *The Neuroscientist, 8*(2), 132–142.
Dalwai, S., Ahmed, S., Udani, V., Mundkur, N., Kamath, S. S., & Nair, M. K. C. (2017). Consensus statement of the Indian academy of pediatrics on evaluation and management of autism spectrum disorder. *Indian Pediatrics, 54*(5), 385–393.
Data and Statistics. Autism spectrum disorder (ASD) |NCBDDD| CDC [Online]. Available: https://www.cdc.gov/ncbddd/autism/data.html.
Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., et al. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*(2), 271.
Esubalew, T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2012). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 21*(2), 289–299.
Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.
Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., et al. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *Journal of the American Academy of Child & Adolescent Psychiatry, 56*(7), 546–555.
Gil, E., Orini, M., Bailón, R., Vergara, J. M., Mainardi, L., & Laguna, P. (2010). Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. *Physiological Measurement, 31*(9), 1271.
Jones, E. A., Carr, E. G., & Feeley, K. M. (2006). Multiple effects of joint attention intervention for children with autism. *Behavior Modification, 30*(6), 782–834.
Jyoti, V., & Lahiri, U. (2019). *Virtual reality based joint attention task platform for children with autism*. IEEE Transactions on Learning Technologies.
Kuriakose, S., & Lahiri, U. (2015). Understanding the psycho-physiological implications of interaction with a virtual reality-based system in adolescents with autism: A feasibility study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23*(4), 665–675.
Kushki, A., Drumm, E., Mobarak, M. P., Tanel, N., Dupuis, A., Chau, T., et al. (2013). Investigating the autonomic nervous system response to anxiety in children with autism spectrum disorders. *PLoS One, 8*(4), e59730.
Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(7), 951–962.
Little, G. E., Bonnar, L., Kelly, S. W., Lohan, K. S., & Rajendran, G. (2016, September). Gaze contingent joint attention with an avatar in children with and without ASD. In *2016 joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob)* (pp. 15–20). IEEE.
Meindl, J. N., & Cannella-Malone, H. I. (2011). Initiating and responding to joint attention bids in children with autism: A review of the literature. *Research in Developmental Disabilities, 32*(5), 1441–1454.
Mei, C., Zahed, B. T., Mason, L., & Ouarles, J. (2018, March). Towards joint attention training for children with ASD-a VR game approach and eye gaze exploration. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 289–296). IEEE.
Mundy, P., & Crowson, M. (1997). Joint attention and early social communication: Implications for research on intervention with autism. *Journal of Autism and Developmental Disorders, 27*(6), 653–676.
Mundy, P., Novotny, S., Swain-Lerro, L., McIntyre, N., Zajic, M., & Oswald, T. (2017). Joint-attention and the social phenotype of school-aged children with ASD. *Journal of Autism and Developmental Disorders, 47*(5), 1423–1435.
Neufeld, J., Ioannou, C., Korb, S., Schilbach, L., & Chakrabarti, B. (2016). Spontaneous facial mimicry is modulated by joint attention and autistic traits. *Autism Research, 9*(7), 781–789.
Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience, 9*(6), 610–632.
Pfeiffer-Leßmann, N., Pfeiffer, T., & Wachsmuth, I. (2012). An operational model of joint attention-timing of gaze patterns in interactions between humans and a virtual human. In *Proceedings of the 34th annual conference of the Cognitive*. Science Society.
Piira, O. P., Huikuri, H. V., & Tulppo, M. P. (2011). Effects of emotional excitement on heart rate and blood pressure dynamics in patients with coronary artery disease. *Autonomic Neuroscience, 160*(1–2), 107–114.
Schroeder, R. (Ed.). (2012). *The social life of avatars: Presence and interaction in shared virtual environments*. Springer Science & Business Media.
Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., & Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering & Technology, 32*(6), 479–484.
Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika, 52*(3), 591–611.
Shin, M. (2012). The role of joint attention in social communication and play among infants. *Journal of Early Childhood Research, 10*(3), 309–317.
Taylor, B. A., & Hoch, H. (2008). Teaching children with autism to respond to and initiate bids for joint attention. *Journal of Applied Behavior Analysis, 41*(3), 377–391.
The National Autistic Society. (2017). *Autism and education in england*.
Wetherby, A. M., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. *Journal of Autism and Developmental Disorders, 37*(5), 960–975.
Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. *Journal of Economic Entomology, 39*(2), 269–270.