J Autism Dev Disord (2015) 45:919–931
DOI 10.1007/s10803-014-2240-5

ORIGINAL PAPER

A Physiologically Informed Virtual Reality Based Social Communication System for Individuals with Autism

Uttama Lahiri · Esubalew Bekele · Elizabeth Dohrmann · Zachary Warren · Nilanjan Sarkar

Published online: 27 September 2014

Abstract Clinical applications of advanced technology may hold promise for addressing impairments associated with autism spectrum disorders (ASD). This project evaluated the application of a novel physiologically responsive virtual reality based technological system for conversation skills in a group of adolescents with ASD. The system altered components of conversation based on (1) performance alone or (2) the composite effect of performance and physiological metrics of predicted engagement (e.g., gaze pattern, pupil dilation, blink rate). Participants showed improved performance and looking pattern within the physiologically sensitive system as compared to the performance based system. This suggests that physiologically informed technologies may have the potential of being an effective tool in the hands of interventionists.

Keywords ASD · Virtual-reality · Eye-tracking · Fixation duration · Pupil diameter · Blink rate

Introduction

With an estimated prevalence of 1 in 68 in United States (CDC 2014), 1 in 250 in India (http://www.autismsocietyofindia.org/), and 1 in 64 in United Kingdom (www.nas.uk.org), effective treatment of autism spectrum disorders (ASD) is a pressing clinical and public health issue. The costs of ASD across the lifespan are thought to be enormous with recent individual incremental lifetime cost projections exceeding $3.2 million (Ganz 2007; Peacock et al. 2012). To address the powerful impairments and costs associated with ASD, a wide variety of interpersonal, biomedical, and behavioral interventions have been offered. Given recent rapid developments in technology, it has been argued that specific computer and virtual reality (VR) based applications could be harnessed to provide effective and innovative clinical treatments for individuals with ASD (Goodwin 2008). A growing number of studies are investigating applications of advanced interactive technologies (e.g., computer technology, robotic systems, and virtual reality environments) to social and communication related intervention (Park et al. 2011; Rus-Calafell et al. 2014; Blocher and Picard 2002; Kozima et al. 2005; Parsons et al. 2004). However, few studies have measured the engagement of people with ASD with these tasks, and how physiologically informed engagement (as expressed through gaze pattern, pupil dilation, and blink rate) can be utilized. The research reported in this paper combined a VR-based conversational learning system with physiological measures of predicted engagement to determine if there are possibilities for improvement in learning

U. Lahiri (✉)
Electrical Engineering, Indian Institute of Technology, Gandhinagar, India
e-mail: uttamalahiri@iitgn.ac.in

E. Bekele · N. Sarkar
Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA

E. Dohrmann
University of Tennessee Health Science Center, Memphis, TN, USA

Z. Warren
Pediatrics and Psychiatry, Vanderbilt University, Nashville, TN, USA

N. Sarkar
Mechanical Engineering, Vanderbilt University, Nashville, TN, USA

trajectories of people with ASD when the VR system responds to the participants considering their engagement levels in addition to their performance on the task.

Rationale for VR Technology

Virtual Reality technology possesses several strengths in terms of potential application to ASD intervention, including: malleability, controllability, replicability, modifiable sensory stimulation, and the capacity to implement individualized intervention approaches and reinforcement strategies. The main sensory output of VR is auditory and visual, which may represent a reduction of information from a real-world setting but also represents a full description of a setting without the need for imagined components (Sherman and Craig 2003; Strickland 1997). Individuals with ASD can improve their learning skills related to a situation if the proposed setting can be manifested in a physical or visual manner (Kerr and Durkin 2004). Virtual environments can easily change the attributes of, add, or remove objects in ways that may not be possible in a real-world setting but could be valuable to teach abstract concepts. Therefore, VR can offer the benefit of representing abstract concepts through visual means (e.g., thought bubbles with text descriptions of a virtual character's thoughts) and seamlessly allows for changes to the environment (e.g., changing the color of a ball or making a table disappear) that may be difficult or even impossible to accomplish in a real-world setting (Sherman and Craig 2003; Strickland 1997). VR can also depict various scenarios that may not be feasible in a "real world" therapeutic setting given naturalistic social constraints and resource challenges (Parsons and Mitchell 2002). Thus, VR is well-suited for creating interactive skill training paradigms in core areas of impairment for individuals with ASD.

Increasingly, researchers have attempted to develop VR applications that respond not only to explicit human–computer interactions (e.g., utilization of keyboards, joysticks, etc.), but to dynamic interactions such as by using gaze patterns. This includes utilizing gaze patterns to understand and alter how individuals process salient social and emotional cues in faces, in addition to technologies that respond to human gaze position during computer game interactions (Wilms et al. 2010).

A Case for Physiology Sensitive VR-Based Systems

Despite the promise of VR-based technological applications for ASD intervention, current VR environments applied to assistive intervention for children with ASD are primarily designed to chain learning via aspects of performance alone (i.e., correct, or incorrect), thereby limiting individualization of application. Specifically, though these VR systems may automatically detect one's eye-gaze and respond based on one's looking pattern, they do not adaptively respond to other physiological markers of engagement (such as pupil dilation or blink rate) that may further optimize learning (Anderson et al. 2006; Jensen et al. 2009). Physiological measures and one's looking pattern can act as important indices of affective experience, which can provide feedback regarding participant's engagement, learning, and intervention (Ernsperger 2003). For example studies have shown pupillary constriction (Anderson et al. 2006) and decreased blink rate (Jensen et al. 2009) for individuals with ASD with increased engagement to a task. Also one's looking pattern quantified by fixation duration while looking to the face region of a communicator is an important indicator of one's engagement (Jones et al. 2008). Emerging research also suggests that the effect of face and gaze processing in children with autism on other physiological indices (Kylliäinen et al. 2006; Kylliäinen and Hietanen 2006; Jansen et al. 2006) may be different than the responses in typically developing children. Specifically, varied experimental social paradigms have yielded significant differences in electroencephalogram (EEG) (Kylliäinen et al. 2006; Kylliäinen and Hietanen 2006), skin conductance (Kylliäinen et al. 2006); and heart rate response (Jansen et al. 2006) for individuals with ASD. If such physiological signals provide information about engagement within social learning scenarios, paradigms capable of meaningfully detecting and responding to such variability may hold advantages over static performance systems.

Thus, the development of technological systems that can both predict performance and capture the intricacies of participant's engagement with a social task could increase the effectiveness of VR-based technological intervention tools. Responding to physiological cues may be useful to effectively intervening with individuals with ASD with communicative challenges. These children often experience states of emotional or cognitive stress that can be measured as autonomic nervous system activation without external expression (Picard 2009). However, physiological signals, which are continuously available and are not necessarily directly impacted by the communicative impairments core to the disorder (Ben Shalom et al. 2006; Jones et al. 2008; Palomba et al. 2000) and which can be used to infer their psychological states, can be effectively utilized to design VR-based system that can monitor one's engagement while recording task performance.

Our Physiology Sensitive VR-Based System

In this paper, we present a novel virtual reality-based system and evaluate its usability in a group of adolescents with ASD and a group of age-matched typically developing (TD) children. This system was designed to administer and alter social interaction in VR involving computer-based bi-directional conversation and provide feedback based on two criterion, (1) objective task performance and (2) engagement as measured by dynamic measures of eye gaze. Specifically, the system measured physiological metrics of gaze, such as, pupil diameter and blink rate, and looking pattern quantified by fixation duration to make predictions regarding engagement during communication tasks. To evaluate task success based on a combination of performance and predicted engagement, we developed an adaptive response technology. This technology uses a rule-governed strategy generator to intelligently merge predicted engagement with performance during the VR-based social task, which results in an individualized task-modification strategy. For example, if a participant's performance was unsatisfactory (i.e., inadequate) and his predicted engagement was low (i.e., not good enough), then the system would automatically adjust conversational prompts and task presentation in an attempt to recapture the participant's attention. Note that our interpretation of 'engagement' partly matches with the views of Schilbach et al. (2013). This conceptualized second-person approach is based on interaction and emotional engagements between people, rather than mere observation for acquiring social knowledge. In our study, the participants do not merely observe their virtual friends presenting their thoughts, but also interact with them during bidirectional conversation. However, our within-system operationalization of "engagement" was not tied to specific concrete emotional constructs, such as participant self-ratings of task enjoyment. Rather, "engagement" was assessed via a dynamically fused mathematical construction of relative and individualized changes in specific physiological signals and looking pattern. Thus, this system uniquely combined VR with measurements of eye physiology and looking pattern to develop an individualized adaptive capability in VR-based autism intervention technology.

In the current manuscript we present a brief overview of the VR system and the results of a usability study that compared the Performance-sensitive System (PS) and the Engagement-sensitive System (ES). Specifically the PS was sensitive to one's performance alone. The ES on the other hand was sensitive to the composite effect of one's predicted engagement and performance in the VR-based social task (Lahiri 2011a). Although not an intervention study, here we present the results of a usability study as a proof-of-concept of our designed system. We hypothesized that compared to the PS individuals interacting with the ES System would demonstrate improved task performance and differences in looking pattern. We also examined differences in eye physiological indices in order to determine if the system would result in variation in a similar form to those noted in non-VR based studies.

Methods

Participants

In our study presented here, we designed a VR-based Adaptive Response Technology based system for adolescents with ASD. However, before we tested the system with the target population we pilot tested the system for usability with four TD adolescents (M = 15.75 years, SD = 2.15 years) recruited from schools neighboring the university. This was done for two reasons, namely, it gave us an opportunity to (1) fine-tune and refine our system before starting our study with the target population, and (2) understand the underlying similarity and dissimilarity in the implication of the interaction with our system on TD and ASD participants. Subsequently eight adolescents with ASD (M = 15.88 years, SD = 2.18 years) were recruited to participate in brief trial of the system (Table 1). Participants with ASD were recruited through existing clinical and research programs in a university-affiliated hospital system. All participants had clinically confirmed ASD diagnoses from clinical psychologists. All the participants with ASD had their Autism Diagnostic Observation Scheduled—Generic (ADOS-G; Lord et al. 2000) and Autism Diagnostic Interview—Revised (ADI-R; Rutter et al. 2003b) scores except one participant (ASD5). To gather additional information on current functioning and ASD symptom profiles, the caregivers of the participants with ASD also completed the Social Responsiveness Scale (SRS; Constantino 2002) and the Social Communication Questionnaire (SCQ; Rutter et al. 2003a). Due to the language-based nature of the tasks, participants were required to have a Receptive Language standard score of 80 or above on the Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn 1997). All research procedures were approved by the University Institutional Review Board.

Inclusion Measures

For our study presented here, we used ADOS-G, ADI-R, SRS, SCQ and PPVT-III for enrolling our participants with ASD. The Autism Diagnostic Observation Schedule-Generic (ADOS-G) is a 45-min semi-structured standardized observational assessment of play, social interaction, and communicative skills that was designed as a diagnostic tool for identifying the presence of autism. It is organized into four modules with each module providing a set of behavioral ratings in five domains: Language and Communication, Reciprocal Social Interaction, Play or Imagination/

Table 1 Participant characteristics

	Age (years)	PPVT[a] standard score	SRS[b] total *T*-score	SCQ[c] total score	ADOS-G[d] total score (cutoff = 7)	ADI-R[e] total score (cutoff = 22)
ASD1	17.583	134	80	12	13	49
ASD2	16.917	110	73	13	7	33
ASD3	14.250	130	89	16	15	34
ASD4	13.833	170	92	14	13	53
ASD5	16.500	92	87	20	–	–
ASD6	18.250	97	63	17	9	49
ASD7	13.000	133	90	10	7	25
ASD8	18.250	97	63	17	9	49

[a] Peabody Picture Vocabulary Test—3rd Edition (PPVT-III; Dunn and Dunn 1997)

[b] Social Responsiveness Scale (SRS; Constantino 2002)

[c] Social Communication Questionnaire (SCQ; Rutter et al. 2003a)

[d] Autism Diagnostic Observation Scheduled-Generic (ADOS-G) (Lord et al. 2000)

[e] Autism Diagnostic Interview-Revised (ADI-R) (Rutter et al. 2003b)

Creativity, Stereotyped Behaviors and Restricted Interests, and Other Abnormal Behaviors.

The Autism Diagnostic Interview-Revised (ADI-R) is a semi-structured, investigator-based interview for parents/caregivers that was developed for the purpose of diagnostic classification of individuals who may have autism or other pervasive developmental disorders (Rutter et al. 2003b). The ADI-R provides explicit scoring criteria that yield cutoff scores in the domains of social reciprocity, language and communication, and restricted and repetitive activities.

The Peabody Picture Vocabulary Test (PPVT) to assess cognitive function (Dunn and Dunn 1997) is a measure of single-word receptive vocabulary.

The Social Responsiveness Scale (SRS) is a 65-item, 15-min parent-report questionnaire designed to quantitatively measure the severity of autism-related symptoms (Constantino 2002). This measure provides an index of ASD-related social competence with questions related to social awareness, social information processing, capacity for reciprocal social communication, social anxiety/avoidance, and autistic preoccupations and traits.

The Social Communication Questionnaire (SCQ) is a brief instrument for the valid screening or verification of ASD symptoms in children that has been developed from the critical items of the Autism Diagnostic Interview (ADI) and compiled into a parent report questionnaire (Rutter et al. 2003a). These questions tap the three critical autism diagnostic domains of qualitative impairments in reciprocal social interaction, communication, and repetitive and stereotyped patterns of behavior.

The instruments such as, ADOS-G, ADI-R, SRS, and SCQ demonstrate fairly robust psychometric properties of ASD classification (see McClintock and Fraser 2011).

VR Apparatus and System Design

The VR-based adaptive response technology system includes three modules: (a) a VR-based social communication task module, (b) a real-time eye-gaze monitoring module, and (c) an individualized adaptive response module that uses a physiologically informed rule-governed intelligent engagement prediction mechanism.

VR-Based Social Communication Task Module

Task Environment We used desktop computer VR applications (Cobb et al. 1999) instead of immersive VR-based systems to increase the accessibility and affordability of the technology and reduce sensory burden and disorientation. A commercially available VR design package (Vizard from Worldviz Llc.) was used to develop the virtual environments. Regarding the display, we developed social situations with context-relevant backgrounds (e.g., with images potentially relevant to the conversations, such as a beach for a discussion on a trip to a sea side), and avatars whose age and appearance resembled those of the participants' peers. We developed 24 social tasks with avatars narrating personal stories to the participants. These personal stories were based on diverse topics of interest to teenagers (e.g., favorite sport, experience with a film, field trip) extracted from an online popular database (www.allfreeessays.com/) written by teenagers. The voices for the avatars were recorded by teenagers from the regional area. Each avatar could make pointing gestures and move dynamically in a context-relevant virtual environment (Fig. 1). For example, when an avatar narrated her experience of driving her family car for a trip to a beach, the VR

Fig. 1 Snapshot of avatar narrating her tour experience to a sea beach within VR environment

environment reflected the view of the vehicle she used for her trip (Fig. 1a). When the avatar narrated the view of the beach, the VR world displayed such a view (Fig. 1b). Subsequently, when the avatar narrated her favorite activities on the beach such as, surfing, the VR scene changed smoothly to display such a situation (Fig. 1c). Thus, realistic situations, relevant to the topic being narrated by the avatar, were presented to the participants.

Twelve avatars (6 males and 6 females) were distributed randomly over the 24 task modules. Faces were acquired by taking the 12 most neutral faces from a sample of 26 heads based on a survey of 20 undergraduate students (Welch et al. 2010).

Task Difficulty Participants watched and listened to an avatar narrate a personal story. The participant was then asked to extract a piece (or multiple pieces) of information from the avatar using a bidirectional conversation module with varying levels of interaction difficulty (e.g., 'Easy', 'Medium', and 'High'). The bidirectional conversation module followed a menu-driven structure often utilized in interactive fiction community (Roberts 2001), in which the participant conversed with the avatar by choosing statements/questions in a particular sequence from a menu presentation and the avatar responded to the participant by speaking out his/her response. The level of interaction difficulty was determined by the number and sequence of questions/statements the participants needed to ask or make to obtain necessary information from the avatar. For example, in an Easy difficulty level task, one was required to ask/make 3 correct questions/comments in a correct sequence to get a single piece of information from the avatar. The Medium difficulty level task required 5 correct questions/comments to be asked in the correct sequence to obtain multiple pieces of information from the avatar. The High difficulty level task required 7 correct questions/comments to be asked in the correct sequence to obtain multiple pieces of information of which one was about a sensitive or personal piece of information. Thus, at higher levels of difficulty, if the relevant information to be obtained was personal, emotional, or sensitive in nature, the participants were required to engage in multiple levels of conversational interaction before they were successfully able to question this sensitive information. To ensure consistency between tasks, tasks in each level of difficulty were carefully designed in consultation with experienced clinicians such that the structure of conversation remained similar regardless of the topics.

Participant–Avatar Interactions There were two kinds of interaction between the avatar and the participant across session types. In the Performance-based session (PS), the avatar only answered the questions asked by the participant. If the participant asked an inappropriate question, the system was programmed to make the avatar guide him/her towards the correct question as a part of the response, thereby serving the role of a facilitator. In the Engagement-based session (ES), on the other hand, the avatar not only served the role of a facilitator but the system was also made aware of the participant's predicted engagement level mapped from the looking pattern and gaze related physiological signals. While using the ES, the system provided individualized feedback based on the participant's gaze related data (discussed in 'Individualized Adaptive Response Module').

Real-Time Eye-Gaze Monitoring

The VR-based system captures gaze data of a participant interacting with an avatar using eye-tracker goggles (from Arrington Research Inc.) with a refresh rate of 30 Hz (in high precision mode) using our custom designed VR—Eye tracker interface platform. The raw gaze data was acquired every 33 ms and stored in a temporary buffer location over the duration of each trial. Subsequently, the gaze data was processed to extract 3 features: mean pupil diameter, mean blink rate, and average fixation duration for each region of interest (e.g., where on the screen the individual was looking) from each segment of the signals, monitored at a refresh rate of 30 Hz in a time synchronized manner (Lahiri et al. 2011 for system specifications).

Individualized Adaptive Response Module

We developed two adaptive VR-based systems (PS and ES) to provide individualized responses in a conversational module. The PS only measured the participant's performance measures (e.g., adequate/inadequate retrieval of a targeted piece of information during conversation). The ES assessed participant's task engagement using both performance measures and also a rule-governed composite effect of objective metrics, such as dynamic viewing patterns [e.g., Fixation Duration (FD)], and eye physiological indices [e.g., Blink Rate (BR), Pupil Diameter (PD)], all of which have been shown to indicate markers of attention and engagement in relevant psychophysiological work with individuals with ASD (Anderson et al. 2006; Jensen et al. 2009).

For the PS, a task-switching mechanism adjusted the interaction difficulty by switching between task difficulty levels based on conversation task performance. Performance was defined by assigning a score of 6 points for each correct question (with the maximum possible score of 18, 30 and 42 for Easy, Medium and High level of difficulty, respectively) asked by the participant and a penalty of 3 points for each incorrect question. If a participant scored ≥70 % of the maximum score possible, then the performance was considered as 'Adequate' (<70 % was 'Inadequate'). Thus, a participant was allowed to make up to 2, 3, and 4 irrelevant or incorrect choices for the 'Easy', 'Medium', and 'High' difficulty level tasks, respectively for achieving 'Adequate' performance. A participant's score was automatically evaluated by the system. If a participant's performance in a task trial was 'Adequate', then the system switched to a task of higher difficulty level, otherwise for 'Inadequate' performance our system switched to a task of lower difficulty level. Note that these strategies to classify performance were chosen as a first approximation to quantify social interaction. With more data and further studies, this classification could be based on either established conventional markers or tied to relative performance change.

For the ES, however, our goal was to switch tasks based not only on performance but also on an individual's predicted engagement as measured by FD, BR, and PD. From literature review we know that increased engagement to a social task is accompanied with increased FD (Jones et al. 2008) and reduction in PD (Anderson et al. 2006) and BR (Jensen et al. 2009). However, there is no evidence in literature where FD, PD and BR are integrated and mapped to one's predicted engagement level. Thus, as a first approximation, we have considered that if two of the three indices indicate increased engagement to a task then we categorize that the individual's engagement is 'Good Enough' else, it is categorized as 'Not Good Enough'. A rule-governed integration module fuses the information on engagement (i.e., 'Good Enough' or, 'Not Good Enough') and task performance (i.e., 'Adequate', or, 'Inadequate') to dynamically switch tasks of different difficulty levels using an individualized task modification strategy (Table 2).

Table 2 Task modification strategy based on composite effect of behavioral viewing, eye physiology, and performance [engagement-sensitive system (ES)]

Case No.	Engagement	Task performance	Action taken by the system
Case1	Good enough	Adequate	Move up
Case2	Good enough	Inadequate	Move down
Case3a/b	Not good enough	Adequate	Move same/ move down
Case4	Not good enough	Inadequate	Move down

Table 3 Rationale behind feedback based on one's fixation pattern

Fixation duration	Feedback
T ≥ 90 %	Your classmate noticed that you were continuously staring at her, and it made her feel awkward. You might try looking somewhere else sometimes to make her feel comfortable
90 % > T ≥ 70 %	Your classmate really enjoyed talking with you. You paid attention to her and made her feel comfortable. Keep it up!
30 % < T < 70 %	Your classmate felt pretty comfortable talking with you, but sometimes she noticed you weren't paying attention. Try to let your classmate know that you're engaged in the conversation
T ≤ 30 %	Your classmate didn't think you were interested in your conversation with her. If you pay more attention to her, she will feel more comfortable

T: Percent FD (Fixation Duration) towards Face_ROI (during conversation) out of total FD

The ES was also made aware of the looking pattern of the participant, and based on this information the system provided additional feedback as given in Table 3. In the present work, we chose certain rules for the feedback given by the system based on one's fixation pattern as a first approximation. For example, a listener looking at the speaker ≥70 % of the time during an interaction has been identified as 'normal while listening' (Argyle and Cook 1976; Colburn et al. 2000).

Procedure

Each participant completed two sessions on separate days (totaling approximately 2.5 h). A visual schedule showing the steps of the study was provided at each session. Each participant sat on a height-adjustable chair and put on eye-tracker goggles. The participant was then asked to rest for

3 min to acclimate to the experimental set-up. At the beginning of each session the eye-tracker was calibrated, which took approximately 15 s. Following calibration, an initial instruction screen appeared. Participants were asked to imagine that the avatars were their classmates at school giving presentations on several different topics. They were informed that after the presentations they would be required to interact with their classmate to find out some information from them. The instructions were followed by an interaction with the avatar, who narrated a personal story. Each storytelling trial was approximately 1½ min long. The first three VR-based social communication task trials (one of each difficulty level) were used to establish a baseline for each participant.

On one day, the VR-based social task modification strategy was based only on one's task performance metric (PS). On another day, the VR-based social task modification strategy was based on the engagement level predicted from the composite effect of one's behavioral viewing, eye physiological indices and the task performance metric (ES; see Table 2). The order of presentation of the PS and ES based tasks was randomized among the participants.

Results

Before analyzing the data gathered from our usability study, we wanted to see whether our VR-based interactive adaptive response technology was acceptable to our participants. All participants (TD and ASD) completed the sessions despite being given the option of withdrawing from the experiment at any time. An exit interview conducted by the experimenter at the end of the experiment revealed that all the participants liked interacting with the system, particularly while using the bidirectional conversation module. No problems were reported with wearing the eye-tracker goggles or understanding the narrated stories. When asked about any take-home lesson that they had from the conversation between them and their virtual classmates, most participants (10 out of 12) said that they learned that they should introduce themselves first when speaking to a new friend for the first time and should look at the faces of friends during conversation. These findings suggest that our system has the potential to be accepted by the target population of adolescents with ASD.

Effect of Interaction with ES System on Performance

We analyzed our findings to see whether there was any improvement in the performance score achieved by the participants while interacting with the ES than that with the PS system. We quantified the participants' performance in the form of a score normalized on a 0–1 scale (for details on the rationale for quantitative estimation, please see Lahiri 2011a). The formulae that we have used to compute the normalized scores are as follows:

Let us consider that the VR-based social task trials of 'Easy', 'Medium', and 'High' difficulty levels have weights designated by 'x', 'y', and 'z', respectively. Also, let a participant acquire an average performance score of '$XAvg$', '$YAvg$', and '$ZAvg$', out of the maximum possible scores of '$XMax$' (i.e., 18), '$YMax$' (i.e., 30), and '$ZMax$' (i.e., 42) for trials of 'Easy', 'Medium', and 'High' difficulty levels, respectively. Thus, if a participant interacted with VR-based social task trials of 'Easy', 'Medium', and 'High' difficulty levels, the normalized performance score achieved was:

$$\text{PERF.SCORE (Normalized)} = \frac{\left(\frac{x}{x+y+z}XAvg\right) + \left(\frac{y}{x+y+z}YAvg\right) + \left(\frac{z}{x+y+z}ZAvg\right)}{\left(\frac{x}{x+y+z}XMax\right) + \left(\frac{y}{x+y+z}YMax\right) + \left(\frac{z}{x+y+z}ZMax\right)} \quad (1)$$

Thus, if one achieved maximum possible scores in tasks of each level of difficulty, then his normalized performance score was 1. In this way, one is not additionally penalized for not having tasks of a particular difficulty level. Likewise, the normalized performance scores were computed for other combinations of VR-based social task trials of varying difficulty levels. Results (Fig. 2a) indicate that there was a non-significant positive trend (Cohen's d effect size = 0.5696) in performance scores for all TD participants, with the range of improvement being 0.55–15.25 %.

When examining ASD participants, statistically significant improvement (Fig. 2b) was noted (effect size = 0.4614, $p = 0.0102$) in the performance score for all the participants (except ASD8) from PS to ES, with the range of improvement being approximately 1.5–12.16 %. The statistical tests applied above were the dependent sample t test and the Cohen's d computing the effect size. In the t test, we compute the t value and subsequently calculate the p value, which is the probability of the null hypothesis being correct. In other words, p gives the probability of seeing what we can see in our data by chance alone. This probability goes down as the size of the effect goes up and as the size of the sample goes up. So what is needed is not just a system of null hypothesis testing but also a system for telling us precisely how large the effects we see in our data really are. This is where effect-size measures come in. Since our sample size was small, we computed the most frequently used Cohen's d representation of effect size. Cohen suggested that d = 0.2 be considered a 'small' effect size, 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size (http://staff.bath.ac.uk/pssiw/stats2/page2/page14/page14.html).

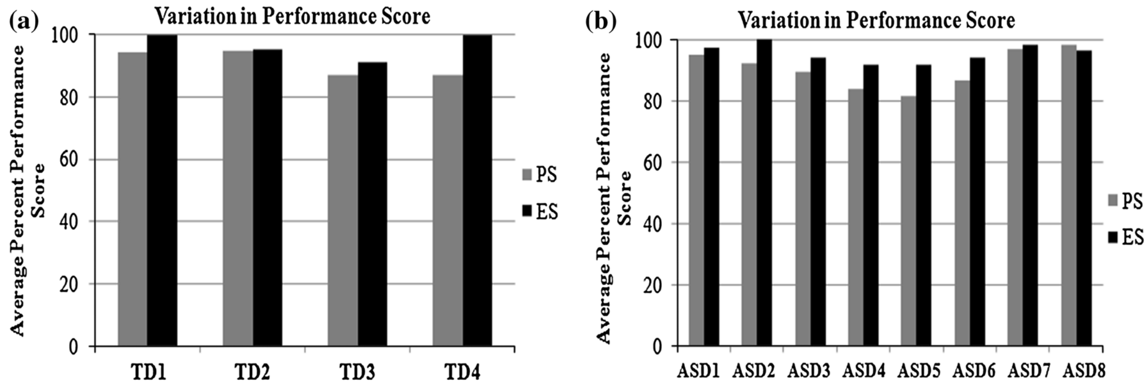

Fig. 2 Variation in Percent Performance Score of **a** TD participants and **b** participants with ASD while interacting with performance-sensitive system (PS) and engagement-sensitive system (ES)

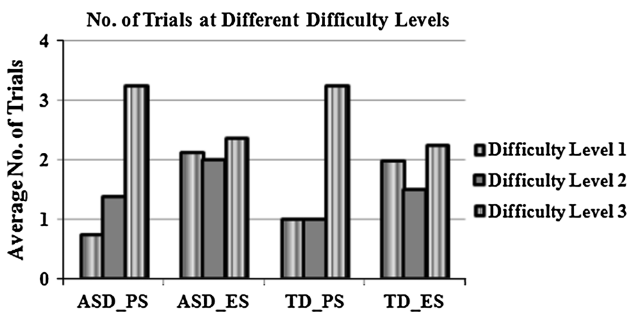

Fig. 3 VR-based social communication tasks of different difficulty levels while interacting with a PS and an ES system for ASD and TD participants

The results of this study indicate that the ES VR-based system contributed to improved performance of the participants as compared to the PS. In fact, there was an improvement in the quantitative measure of performance (Fig. 2a, b) during ES compared to PS for majority of the participants. In order to understand the cause of such improvement, we noticed that there was an increase in number of easy trials in ES as compared to PS and this improvement in performance score might seem as a result of increase in the number of trials played in 'Easy' difficulty level tasks (Fig. 3). However, we also found that the number of trials in 'Medium' difficulty level (Fig. 3) were higher for ES than for the PS. Additionally for both the ASD and TD groups, the number of trials in the 'High' difficulty level were higher than the 'Easy' and 'Medium' difficulty levels for ES. Thus it is entirely not clear what caused the improved performance and a future more in-depth study with a bigger sample size may provide us with more insight.

Effect of Interaction with ES System on Looking Pattern

For the ES System, the system was made aware of the looking pattern of the participant (i.e., where and for how long the participant was viewing), and based on this information the system provided feedback (Table 3). This feedback was designed to be indirect in nature in the sense that it did not ask the participant explicitly to look at the avatar's face. Note that if the participant looked towards the face of the avatar for greater than 90 %, in order to encourage normal looking pattern, rather than looking so intently towards the communicator during social conversation, our system prompted the participant to modify his looking pattern. We examined whether our ES system has the potential to improve one's looking pattern. This is important since in order to achieve improved social communication skills, one must not only improve task performance, but also be able to carry out communication in a socially appropriate way (e.g., paying attention to the face of the communicator). One's viewing patterns can be quantified by Fixation Duration (FD) on different regions of the presented visual stimulus. Results (Fig. 4a) indicate that there was improvement ($p = 0.008$, effect size = 0.9076) in the viewing pattern of the TD participants, with increased FD while looking towards the face region during interaction with the ES compared to the PS. Participants with ASD also looked more towards the face region of the avatar during interaction with the ES compared to the PS ($p = 0.002$, effect size = 0.4824; see Fig. 4b).

Thus, the ES was able to achieve improvement in one's looking pattern during social communication for both the TD and the ASD groups of participants.

Effect of Interaction with ES System on Eye Physiological Indices

One's eye physiological indices e.g., Pupil Diameter (PD) and Blink Rate (BR) were used to measure task engagement. Previous non-VR based studies indicate that individuals demonstrate reduction in PD (Rutherford and Towns 2008)

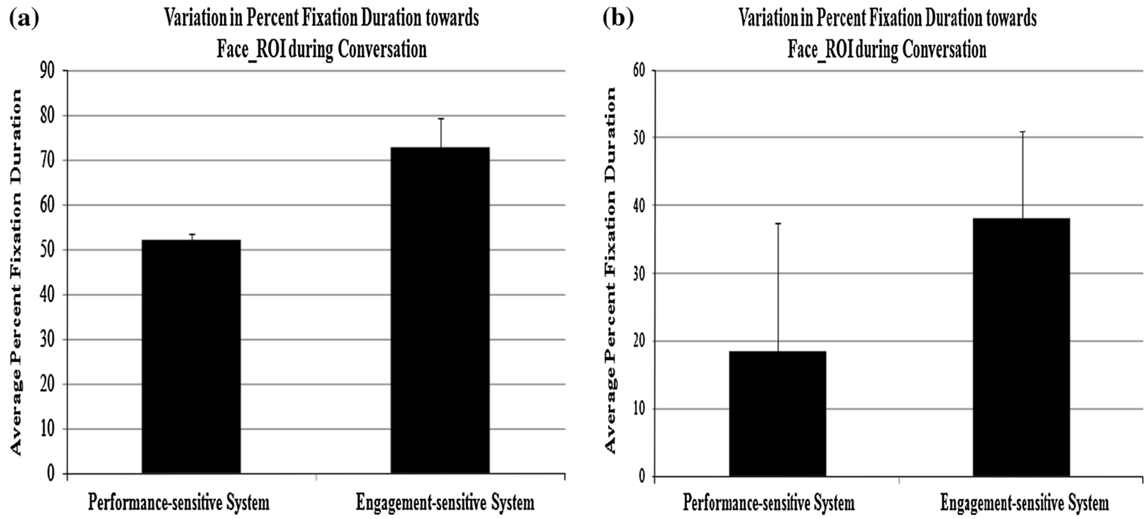

Fig. 4 Variation in percent fixation duration of **a** TD participants and **b** participants with ASD while looking towards the Face Region (Face_ROI) of the avatars during VR-based social conversation

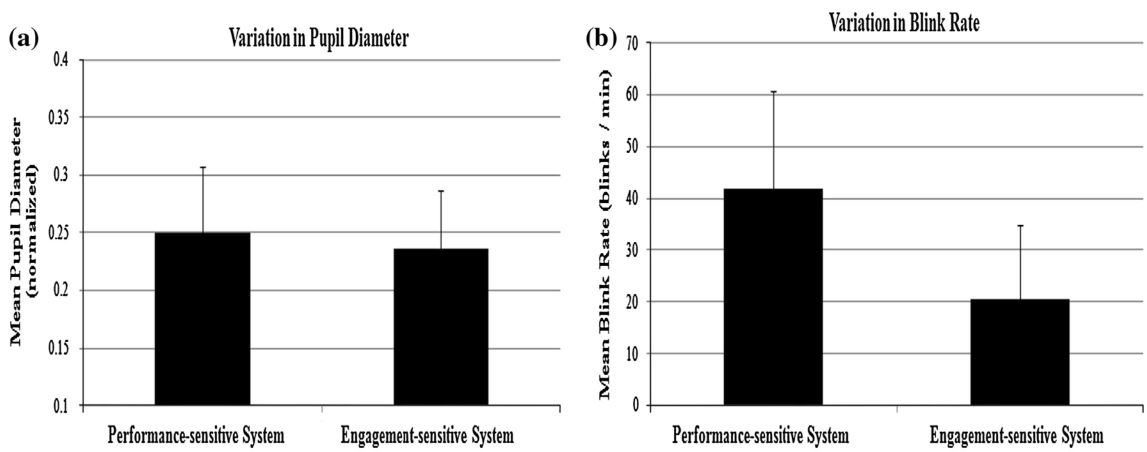

Fig. 5 Eye physiological indices **a** pupil diameter and **b** blink rate of TD participants while interacting with PS system and the ES system

and BR (Anderson et al. 2006) while being engaged in a task. We normalized PD on a 0–1 scale with respect to the eye camera window of the eye-tracker that we used. We compared participants' eye physiological indices while they interacted with PS and ES systems to determine if variation in eye physiological indices was indicative of participants' improved engagement level with ES versus PS. As seen in Fig. 5a, a non-significant decrease of 5.55 % (effect size = 0.14) in mean PD was found from PS to ES for the TD participants. A statistically significant decrease of 50.96 % was observed for the BR (effect size of 0.53, $p = 0.04$; see Fig. 5b) for the TD participants.

For the participants with ASD, no significant difference emerged for PD between the PS and ES (see Fig. 6a). Similarly, although a decreasing trend was found in BR, the difference was not significant (effect size = 0.15; see Fig. 6b).

Effect of Interaction with ES System on Performance, Looking Pattern and Eye Physiology for a Matched Participant Pair

Not only there was an effect of interaction with our ES system on the participant groups' (ASD and TD) performance, looking pattern and eye physiological indices, the implication can also be realized at an individual level. For this let us consider one participant from each group (matched on age).

As an example, first, let us consider the task progression of one of the TD (henceforth 'T1') participants while interacting with the PS and the ES systems (Fig. 7a). As can be seen from this figure, this participant progressed though six trials during PS and seven trials during ES session. During the PS session, he started at the 'Easy' difficulty level as the baseline (Trial1), continued at the 'Easy' difficulty level in Trial2, and

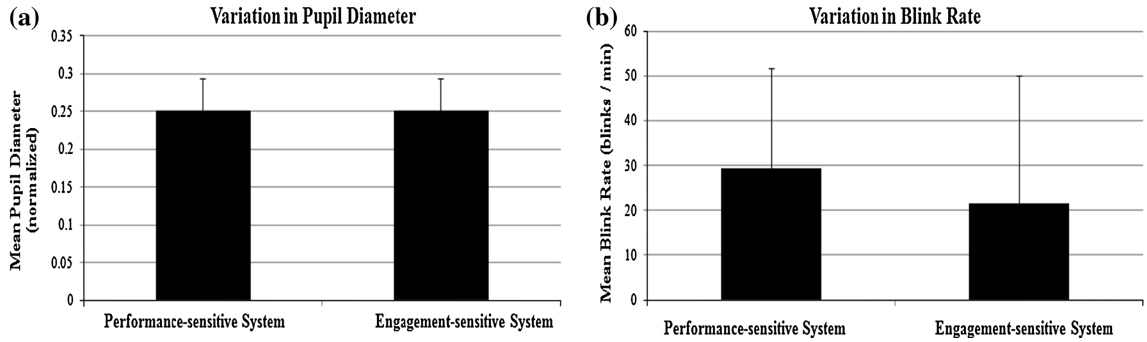

Fig. 6 Eye physiological indices **a** pupil diameter and **b** blink rate of participants with ASD while interacting with PS system and the ES system

then moved to 'Medium' difficulty level in Trial3. In Trial4, he moved to 'High' difficulty level and remained at the same difficulty level until Trial6. However, during the ES session, he started at the 'Easy' difficulty level as the baseline (Trial1), continued at the 'Easy' difficulty level in Trial2, then moved to 'Medium' difficulty level in Trial3. At the end of Trial3, the strategy generator detected a low predicted engagement level along with 'Adequate' performance, thereby causing the adaptive response technology to maintain the same difficulty level, i.e., 'Medium' with the hope of regaining the engagement level of the participant during Trial4. The strategy generator detected an improved engagement level of the participant at the end of Trial4. Thus, the adaptive response technology offered a task of 'High' difficulty level in Trial5 and this continued till Trial6. Now let us consider an example of task progression for a participant with ASD (henceforth 'A1') while interacting with the PS and the ES systems (Fig. 7b). As can be seen from this figure, this participant progressed through seven trials during each of PS and ES sessions. During PS session, he started at the 'Easy' difficulty level as the baseline (Trial1), continued at the 'Easy' difficulty level in Trial2, and then moved to 'Medium' difficulty level in Trial3. In Trial4, he moved to 'High' difficulty level and remained at the same difficulty level up to Trial6. Subsequently, he moved down to 'Medium' difficulty level in Trial7. Although the same number of trials was executed by him during ES session, yet we get a different picture for VR-based social task progression during this session. During ES session, he started with the VR-based social communication task of 'Easy' difficulty level as the baseline (Trial1). Then he continued in the 'Easy' difficulty level during Trial2. At the end of Trial2, the strategy generator predicted an increased engagement level of the participant which caused him to be shifted to the 'Medium' difficulty level in Trial3. At the end of Trial3, the strategy generator detected a low predicted engagement level along with 'Adequate' performance, thereby causing the adaptive response technology to maintain the same difficulty level, i.e., 'Medium' with the hope of regaining the engagement level of the participant during Trial4. This strategy of the strategy generator worked out well and the strategy generator then detected an improved engagement level of the participant at the end of Trial4. Thus, the adaptive response technology offered a task of 'High' difficulty level in Trial5. Thereafter, the strategy generator detected a continued high engagement level of the participant, thereby causing him to carry on with the 'High' difficulty level up to Trial7.

Additionally, on an individual level, we found that for both T1 and A1, there was an improvement in the average percentage performance score from the PS to the ES system, with the improvement being more for A1 as compared to T1. Also, such improvement in the looking pattern with respect to looking towards the face region of the virtual peer was observed for both T1 and A1 (Table 4) during ES as compared to PS. For the TD participant (T1), the average percentage of fixation duration (FD) was greater as compared to the ASD participant (A1). Also, for the eye physiological features, (pupil diameter and blink rate) we observe reduction in blink rate from PS to ES for both A1 and T1, and reduction in pupil diameter from PS to ES for A1 but not for T1. This might be used to explain the reason for smaller improvement in the percentage performance score for the TD participant (T1) as compared to the ASD participant (A1) (Table 4).

Discussion

Given rapid progress and developments in technology, it has been argued that specific computer and VR based applications may be effective tools in the hands of the interventionist working with children with ASD (Goodwin 2008). Although a growing number of studies have examined applications of advanced interactive technologies to social and communication related intervention (Park et al. 2011; Rus-Calafell et al. 2014; Blocher and Picard 2002; Kozima et al. 2005; Parsons et al. 2004), current

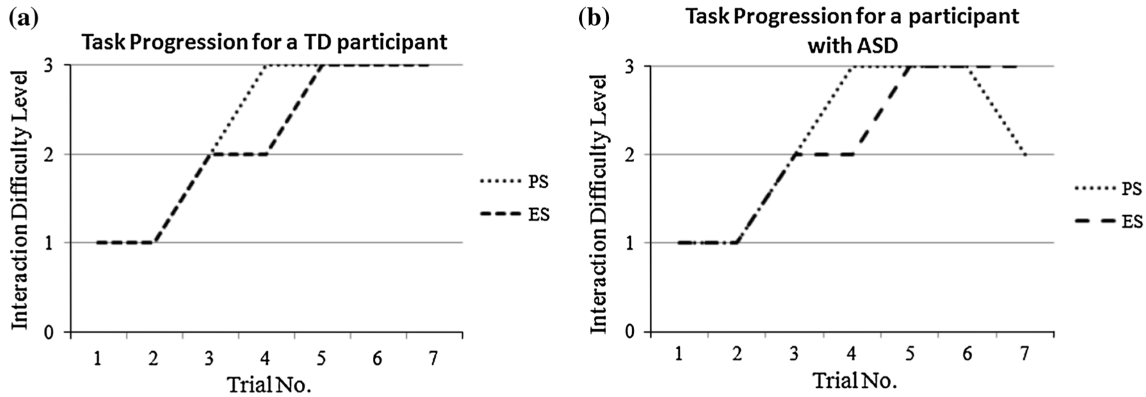

Fig. 7 Progression of VR-based social communication tasks while interacting with a PS and an ES system for **a** a TD participant and **b** a participant with ASD

Table 4 Composite effect of engagement-sensitive system (ES) and performance-sensitive system (PS) on behavioral viewing, eye physiology, and performance for participant pair (A1 and T1)

Indices (units)	For TD participant (T1)	For ASD participant (A1)
Performance score (%)	86.8 (PS)	86.8 (PS)
	91.2 (ES)	94.1 (ES)
Average FD_Face ROI (%)	53.02 (PS)	28.16 (PS)
	69.18 (ES)	36.28 (ES)
Average PD (normalised)	0.1965 (PS)	0.2011 (PS)
	0.2595 (ES)	0.1885 (ES)
Average BR (blinks/min)	53.91 (PS)	4.39 (PS)
	37.91 (ES)	0.82 (ES)

system applications tend to operate with limited abilities to alter system function based on markers other than simple performance measures within the system. Thus, current VR environments as applied to assistive intervention for children with ASD are primarily designed to chain learning via aspects of performance alone (i.e., correct, or incorrect) and thereby may limit individualization of application. The ability to incorporate and automatically adapt system interactions based on physiological markers (Kylliäinen et al. 2006; Jansen et al. 2006) of engagement may be an important avenue for the development of technological systems of more pronounced effect and meaning.

In the present work we developed and applied VR-based physiologically informed adaptive response technology to a small sample comprising of participants with ASD in addition to a pilot sample of TD participants. Results from this small usability study suggest that such affectively sensitive technology has the potential to contribute to the improvement in social communication.

The results from our small usability study reflect the potential of VR-based ES system to improve task performance measured in terms of one's ability to extract relevant information from a communicator during social communication task. Also we find that interaction with an ES social communication platform may contribute to improved task performance in terms of carrying out social communication tasks of increased degree of difficulty.

Improved social communication skills require one to achieve not only improved task performance, but also acquire the ability to carry out conversation in a socially appropriate way. In fact, to understand the social communication vulnerabilities of individuals with ASD, research has examined how they process salient social cues, specifically from faces (Rutherford, and Towns 2008; Jones et al. 2008). The ability to derive socially relevant information from faces is thought to be a fundamental skill for facilitating reciprocal social interactions (Trepagnier et al. 2002) and an early deficit may contribute in part to the developmental cascade associated with core vulnerabilities of the disorder (Dawson 2008). Thus given the importance of paying attention to the face of the communicator during social communication, we tried to analyze whether our ES VR-based system has the potential to contribute to the improvement in one's looking pattern. Our results indicate that both ASD and TD participants showed significant improvement in looking pattern in terms of fixation towards the face of their virtual peers for the ES compared to the PS.

There is evidence in literature that increased engagement in non-VR based tasks are associated with pupillary (PD) constriction with reduced blink rate (BR) (Anderson et al. 2006; Jensen et al. 2009). Our ES system was programmed to be sensitive to one's eye-physiological indices (PD and BR). Though our PS system was not sensitive to one's eye-physiological indices, yet we collected the PD and BR data for offline analysis. In the present usability study, with a limited sample size, we found significant decrease in the BR of TD participants and a non-significant decreasing trend for ASD participants when interacting with the ES compared to the PS. Both ASD and TD participants, showed non-significant decreases in PD. These overall decreasing trends in BR and PD suggest that participants with ASD, showed improved engagement while interacting with the ES system as compared to that with the PS system.

Limitations

When considering the applicability of this technology for practical intervention with adolescents with ASD, several methodological considerations limit interpretation of our findings. The present study demonstrates the potential of using real-time, gaze-based technology for social skill improvement for individuals with ASD. However, these findings are preliminary and limited in nature. The task designed for this study was employed as a first pilot step in evaluation of the benefits of such a technological system within ASD intervention. Almost all participants were able to demonstrate changes in terms of more (1) looking at faces of virtual peers (avatars) for a longer duration as well as (2) performing better while interacting with the ES system than that with the PS system. Because these findings are weakened by our limited sample size and corresponding issues of low power, these trends underscore the need for additional research with more complex systems and larger sample sizes. However, the present study was designed as a proof-of-concept study and not as an intervention study. Certainly more work is needed to understand the ultimate potential of VR platforms that can integrate physiological and engagement processes in order to target and treat core and associated features of ASD.

Another limitation of this current study was the mechanism used for bidirectional conversation between the participant and the avatar. Specifically, it needs to be mentioned that carrying out conversation while using drop-down menu may be a skill relevant within a computer-generated environment and is not related to actual real-life conversation skills. However, as a first step towards developing a proof-of-concept application we adhered to the computer-based menu-driven communication. Future research should be pointed towards developing more natural bidirectional communication platform where a participant can speak out his/her question to the avatar.

An additional methodological limitation of our present study was the somewhat limited range of children with ASD included in the study. Specifically, given the language, or more specifically reading, burden of the current platform we limited enrollment to children with tested cognitive/language abilities commensurate with the demands of the platform (i.e., average to above average intelligence). A more robust system would likely need to incorporate the ability to utilize speech recognition within a bidirectional conversation module in order to address this specific limit.

Conclusion

This work demonstrates proof-of-concept of the technology and improvement in performance and viewing pattern while interacting with a VR-based gaze-sensitive social communication task. Even though through our preliminary usability study we have seen some improvement within the VR environment, the generalization of skill improvement in real-life remains an open question. Thus, questions about the practicality, efficacy, and ultimate benefit of the use of this and other technological tools for demonstrating clinically significant improvements in terms of ASD impairment remain. Although some initial improvements were seen within the system, there was no evidence that this technology realized change for participants outside of the limited environment of the experiment itself or over time.

Acknowledgments The authors would like to thank the participants and their families for making this study possible. We also gratefully acknowledge National Science Foundation Grant (Grant Number 0967170) and National Institute of Health Grant (Grant Number R01 MH091102) that partially supported the research presented here.

References

Anderson, C. J., Colombo, J., & Shaddy, D. J. (2006). Visual scanning and pupillary responses in young children with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology, 28*, 1238–1256.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, MA: Cambridge University Press.

Ben Shalom, D., Mostofsky, S. H., Hazlett, R. L., Goldberg, M. C., Landa, R. J., Faran, Y., et al. (2006). Normal physiological emotions but differences in expression of conscious feelings in children with high functioning autism. *Journal of Autism and Developmental Disorders, 36*(3), 395–400.

Blocher, K., & Picard, R. W. (2002). Affective social quest: emotion recognition therapy for autistic children. In L. Canamero, B. Edmonds, K. Dautenhahn, & A. H. Bond (Eds.), *Socially intelligent agents: Creating relationships with computers and robots*. Dordrecht: Kluwer.

CDC. (2014). Prevalence of autism spectrum disorders—ADDM Network. *MMWR Surveill Summ, 63*(2), 1–22.

Cobb, S. V. G., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality induced symptoms and effects. *Presence, 8*(2), 169–186.

Colburn, A., Drucker, S., & Cohen, M. (2000). *The role of eye-gaze in avatar-mediated conversational interfaces*. New Orleans, LA: In SIGGRAPH Sketches and Applications.

Constantino, J. N. (2002). *The social responsiveness scale*. Los Angeles, CA: Western Psych. Services.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, *20*(3), 775–803.

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Ernsperger L (2003) Keys to success for teaching students with autism, *Future Horizons*.

Ganz, M. L. (2007). The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics and Adolescent Medicine, 161*(4), 343–349.

Goodwin, M. S. (2008). Enhancing and accelerating the pace of autism research and treatment: The promise of developing Innovative Technology. *Focus on Autism and Other Developmental Disabilities, 23*, 125–128.

Jansen, L. M. C., Gispen-de Wied, C. C., Wiegant, V. M., Westenberg, H. G. M., Lahuis, B., & van Engeland, H. (2006). Autonimic and neuroendocrine responses to a psychosocial stressor in adults with Autistic Spectrum Disorder. *Journal of Autism and Developmental Disorders, 36*, 891–899.

Jensen, B., Keehn, B., Brenner, L., Marshall, S. P., Lincoln, A. J., & Müller, R. A. (2009) Increased eye-blink rate in autism spectrum disorder may reflect dopaminergic abnormalities. In *International Society for Autism Research*, Poster Presentation.

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry, 65*(8), 946–954.

Kerr, S., & Durkin, K. (2004). Understanding of thought bubbles as mental representations in children with autism: Implications for theory of mind. *Journal of Autism and Developmental Disorders, 34*(6), 637–648.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2005). Children-robot interaction: A pilot study in autism therapy. *Progress in Brain Research, 164*, 385–400.

Kylliäinen, A., Braeutigam, S., Hietanen, J. K., Swithenby, S. J., & Bailey, A. J. (2006). Face- and gaze-sensitive neural responses in children with autism: A magnetoencephalographic study. *European Journal of Neuroscience, 24*(9), 2679–2690.

Kylliäinen, A., & Hietanen, J. K. (2006). Skin conductance responses to another person's gaze in children with autism. *Journal of Autism and Developmental Disorders, 36*(4), 517–525.

Lahiri U. (2011a) Virtual-reality based gaze-sensitive adaptive response technology for children with autism spectrum disorder. *Dissertation.*

Lahiri, U., Warren, Z., & Sarkar, N. (2011). Design of a gaze-sensitive virtual social interactive system for children with autism. *Transactions on Neural Systems and Rehabilitation Engineering, 19*(4), 443–452.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H, Jr, Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

McClintock, J. M., & Fraser, J. (2011) Diagnostic instruments for autism spectrum disorder. *Ministries of Health and Education,* 1–30.

Palomba, D., Sarlo, M., Angrilli, A., Mini, A., & Stegagno, L. (2000). Cardiac responses associated with affective processing of unpleasant film stimuli. *International Journal of Psychophysiology, 36*, 45–57.

Park, K. M., Ku, J., Choi, S. H., Jang, H. J., Park, J. Y., Kim, S. I., et al. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: A randomized, controlled trial. *Psychiatry Research, 189*(2), 166–172.

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disability Research, 46*(Pt 5), 430–443.

Parsons, S., Mitchell, P., & Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 34*(4), 449–466.

Peacock, G., Amendah, D., Ouyang, L., & Grosse, S. D. (2012). Autism spectrum disorders andhealth care expenditures: The effects of co-occurring conditions. *Journal of Developmental and Behavioral Pediatrics, 33*(1), 2–8.

Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B., 364*(1535), 3575–3584.

Roberts MJ (2001) *Choosing a conversation system*. http://www.tads.org/howto/convbkg.htm.

Rus-Calafell, M., Gutiérrez-Maldonado, J., & Ribas-Sabaté, J. (2014). A virtual reality-integrated program for improving social skills in patients with schizophrenia: a pilot study. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 81–89.

Rutherford, M. D., & Towns, M. T. (2008). Scan path differences and similarities during emotion perception in those with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 1371–1381.

Rutter, M., Bailey, A., Berument, S., Lord, C., & Pickles, A. (2003a). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Rutter, M., Le Couteur, A., & Lord, C. (2003b). *Autism diagnostic interview revised WPS edition manual*. Los Angeles, CA: Western Psychological Services.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*, 393–462.

Sherman, W. R., & Craig, A. B. (2003) Understanding virtual reality: Interface, application, and design. Boston: Morgan Kaufmann Publishers.

Strickland, D. (1997) Virtual reality for the treatment of autism. In G. Riva, (Ed.), *Virtual reality in neuropsycho-physiology* (pp. 81–86). Amsterdam: IOS Press.

Trepagnier, C., Sebrechts, M. M., & Peterson, R. (2002). Atypical face gaze in autism. *Cyberpsychology and Behavior*, *5*(3), 213–217.

Welch, K., Lahiri, U., Sarkar, N., & Warren, Z. (2010). An approach to the design of socially acceptable robots for children with autism spectrum disorders. *International Journal of Social Robotics, 2*(4), 391–403.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience., 5*(1), 98–107.