International Conference on Virtual Rehabilitation 2011
Rehab Week Zurich, ETH Zurich Science City, Switzerland, June 27 - 29, 2011

Dynamic Gaze Measurement with Adaptive Response Technology in Virtual Reality based Social Communication for Autism

Uttama Lahiri
Mechanical Engineering Department
Vanderbilt University
Nashville, Tennessee, USA

Zachary Warren
Pediatrics and Psychiatry
Vanderbilt University
Nashville, Tennessee, USA

Nilanjan Sarkar
Mechanical Engineering Department
Vanderbilt University
Nashville, Tennessee, USA

Abstract— **Impairments in skills related to social communication are thought to be core deficits in children with autism spectrum disorders (ASD). Specifically, these children demonstrate atypical viewing patterns in part characterized by greater fixation towards non-social objects than faces of individuals during social communication. Additionally, several assistive technologies, particularly Virtual Reality (VR), have been investigated to promote social interactions in this population. Thus given the promise of VR-based social interaction and atypicalities surrounding eye-gaze and social information processing characterizing ASD, in the current study, a novel technology was developed. This study combined social tasks presented in VR environment with a computationally-enhanced eye-tracker to provide individualized feedback. The developed system is capable of delivering individualized feedback based on a child's dynamic gaze patterns during VR-based social communication task. Result from a usability study with six adolescents with ASD demonstrate the technological capacity of such a system to adaptively respond and potentially modify aspects of behavioral viewing patterns (e.g., fixation counts, fixation duration, etc.) during VR-based social task.**

Keywords- ASD; virtual-reality; eye-tracking; fixation counts; fixation duration

I. INTRODUCTION

Autism is a neurodevelopmental disorder characterized by core deficits in social interaction and communication accompanied by restricted patterns of interest and behavior [1]. Because children with Autism Spectrum Disorder (ASD) show difficulties in social judgment (e.g., deciding on appropriate social behaviors), [2] a main focus of autism research has been to understand how people with autism process salient social cues, notably from faces. For normal reciprocal social interactions and interpersonal communication, the ability to derive socially relevant information from faces is a fundamental requirement. However, children with ASD often demonstrate difficulty in using face as a channel for social communication [3]. Thus social attention and its measurement is considered a major aspect in ASD intervention [4], which could involve acquiring and analyzing eye-gaze data [5] [6] while an individual with ASD is involved in a social communication task.

While at present there is no single accepted intervention for ASD, there is growing consensus that intensive behavioral intervention programs can significantly improve long term outcomes for these individuals [7]. Moreover, given the limitation on the availability of trained professional resources in ASD intervention, it is likely that emerging technology will play an important role in providing more accessible intensive individualized intervention [8]. In response to this need, a growing number of studies have been investigating the application of advanced interactive technologies to address core deficits related to ASD, namely computer technology [9], virtual reality environments [10] [11] and robotic systems [12].

Virtual reality (VR) technology possesses several strengths in terms of potential application for children with ASD, namely, malleability, controllability, replicability, modifiable sensory stimulation, and an ability to pragmatically individualize intervention approaches and reinforcement strategies [13]. While VR does not necessarily include direct human-to-human interaction, having controllable complexity of a virtual world with minimized distractions or distresses may allow for simplified but embodied social interaction that could be less intimidating or confusing for some children with ASD [14]. However, VR should not be considered an isolating agent, because dyadic communication accomplished between a child and a VR environment can lead into triadic communication including a clinician, caregiver, or peer and in due course potentially accomplish the intervention goals of developing social communication skills between the child with ASD and another person [15]. VR can also illustrate scenarios which can be changed to accommodate various situations that may not be feasible in a given therapeutic setting with space limitations, resource deficits, etc. [16]. Also, 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 need for imagined components [13] [17]. 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 [18]. Since VR mimics real environments in terms of imagery and contexts, it may allow for efficient generalization of skills from the VR environment to the real world [19]. Furthermore, the spectrum nature of autism means an individual approach is appropriate, and computers can accommodate individualized treatment [13]. Thus VR represents a medium well-suited for creating interactive intervention paradigms for skill training in core areas of impairment for these children. In our usability study, we present VR-based task on a computer screen as VR is often effectively experienced on a desktop system using standard computer input devices [16] for ASD intervention.

Despite potential advantages, current VR environments as applied to assistive intervention for children with ASD are designed to chain learning via aspects of performance; however, they are not capable of a high degree of individualization [10] [11]. A recent work using VR has demonstrated the feasibility of linking the gaze behavior of a virtual character with a human observer's gaze position during joint-attention tasks [20]. Specifically, these systems though may automatically detect one's eye-gaze and respond based on one's viewing pattern, they cannot objectively identify and predict social engagement, understand viewing patterns, and psychophysiological effect of the specific child based on attentive indices. Thus, development of systems that are responsive to the dynamic gaze patterns of these children to address some of their core deficits in communication and social domains is still at its infancy.

In this work, we focus on the applicability of VR environment to design social communication scenarios to understand eye-gaze information, because atypical gaze in face-to-face social communication is one of the core impairments of children with ASD. While not an intervention study, our present work demonstrates the development of a new system that can integrate VR-based interaction with the participant's real-time viewing pattern and investigates on how eye-gaze sensitive real-time adaptive response has the potential to modify certain aspects of social communication, e.g., behavioral viewing pattern of individuals with ASD. This paper is organized as follows: In Section II, we present the system design. Section III discusses the methods used for this study. Section IV presents the results obtained in our usability study. Section V summarizes the contributions of this work and indicates the direction of future work.

II. SYSTEM DESIGN

The VR-based gaze-sensitive system with adaptive response technology has three main subsystems: 1) a VR-platform that can present social tasks; 2) a real-time eye-gaze monitoring mechanism; and 3) an integration module that establishes the bidirectional interface between the VR-based task presentation module and the real-time gaze monitoring module.

A. VR-based Task Presentation

VR-based tasks are created using Vizard VR design package from Worldviz (www.worldviz.com) as the primary design platform. This software comes with a limited number of avatars, virtual objects, and scenes that can be used to create a story in VR. However, a number of enhancements have been made on the VR-platform. In order to perform social communication tasks with participants with ASD, we develop more extensive social situations with appropriate contexts, and avatars whose age and appearance resemble those of the participants' peers without trying to achieve exact similarities.

Thus new avatar heads are created from 2D photographs of teenagers, which are then converted to 3D heads by '3DmeNow' software for compatibility with Vizard. Facial expressions (e.g., 'neutral', 'happy', and 'angry') are morphed by 'PeopleMaker' software (Fig. 1).

Figure 1. Screenshots of avatars demonstrating neutral (*top*) (a), happy (*middle*) (b) and angry (*bottom*) (c) facial expression.

The avatar's eyes are made to blink randomly with an interval between 1 and 2 s to render automatic animation of a virtual face similar to the work of Itti et al. [21]. The avatar stands at 3 ft from the origin of main scene of virtual world. Different cultures have varying rules for social distance. For example, the overcrowded nature of Asian countries causes people to be accustomed with very close distances, whereas western culture considers very close distances as uncomfortable [22]. Though our system is capable of simulating variations in personal distance, in our presented work, the avatar stands at 3 ft from the origin of main scene of virtual world to simulate the social distance suitable for western culture [22] [23]. One can view the avatars within the system from first-person perspective while the avatars narrate personal stories, which is comparable to research on social anxiety and social conventions [24]. In the present study, the first-person stories shared by avatars are adapted from

Dynamic Indicators of Basic Early Literacy Skills [25] reading assessments and includes content thought to be related to potential topics of school presentations (e.g., reports on experiences, trips, favorite activities, etc.). In order for the avatars to narrate the story, the recorded audio files are lip-synched using a vizard-based speak module. Additionally, where a participant is looking at the VR-based visual stimuli (e.g., avatar's face, objects of interest, etc.) is characterized by a set of Regions of Interest (ROIs). These ROIs have been programmed such that the dynamic eye-tracking algorithm we develop would keep track of the eye-gaze of the participants as they communicate with the VR-based tasks. In the present study, we segmented the VR-based visual stimulus into 3 ROIs: avatar's face (Face_ROI), a context-relevant object (Object_ROI), and rest of the VR environment (Others_ROI) (Fig. 2). Face_ROI captures the forehead, eye brows, eyes and surrounding muscles, nose, cheeks, mouth and surrounding muscles. Object_ROI captures a context-relevant object (e.g., for a story on a zoo visit, the context-relevant object is a picture displaying collage of snapshots of the animals seen at the zoo).

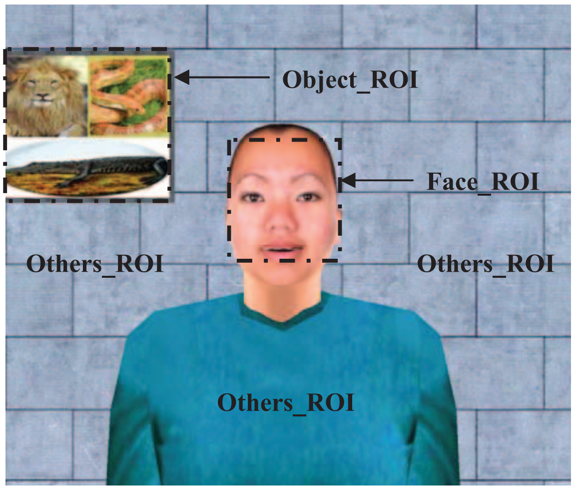

Figure 2. Allocation of Regions of Interest (ROIs) within the visual stimulus (Face_ROI, Object_ROI, and Others_ROI).

B. Real-time Gaze Monitoring Mechanism

The system captures eye data of a participant interacting with a virtual peer (i.e., an avatar) using eye-tracker goggles from Arrington Research Inc. The eye-tracker that we use comes with a Video Capture Module with a refresh rate of 60 Hz for low precision and 30 Hz for high precision to acquire basic features (e.g., raw fixation durations, gaze coordinates) using a software called Viewpoint for offline analysis. We have designed Viewpoint-Vizard Interface module (Fig. 3) and a new gaze database that captures the task-related event markers (e.g., trial start/stop, amount of one's viewing of different ROIs, etc.), raw behavioral viewing data (e.g., Fixation Duration (FD), and 2D gaze coordinates) and performance measures (e.g., a participant's response to question asked by the system) with a refresh rate of 30 Hz in a time-synchronized manner. Signal processing techniques such as windowing, and thresholding are used to filter these data to eliminate noise and subsequently extract the relevant features. In the present study, we compute the Fixation Duration by using a thresholding window of 200 ms as the lower limit to eliminate the blinking effects [26] and 450 ms as the upper threshold (to eliminate noise due to glare effects of the cameras of the eye-tracker that we use). Subsequently, the Sum of Fixation Counts (SFC) and Total Fixation Duration (FD_{Total}) are computed for each ROI. We chose these primary indices because they are important indicators of one's behavioral viewing patterns. The higher the fixation frequency on a region as measured by Sum of Fixation Counts [27] [28], the greater the attention and interest [29] in the target. Also the Total Fixation Duration is important as literature indicates that children with ASD exhibit lower Fixation Duration while viewing human faces than the non-human face stimuli [30] during social communication.

C. Integration module that establishes bidirectional interface between the VR-based task presentation module and the real-time gaze monitoring module

Real-time gaze coordinates of a participant are acquired and converted to VR (Vizard) compatible format using a Vizard-Viewpoint Interface module (Fig. 3).

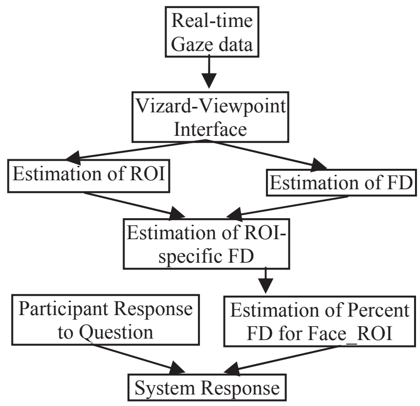

Figure 3. Dynamic Gaze-based Feedback

A task computer (C1), (Fig. 4) where the VR-based tasks are presented, runs Viewpoint Software at the background and Vizard software at the foreground. The integration module triggers a 33 ms timer to acquire the gaze coordinates. Based on the participant's 2D gaze-coordinates, the integration module then computes the specific ROI looked at by the participant. Times spent by the participant looking at different ROIs are stored in respective buffers that are added up at each instant during participant-avatar interaction. This determines the participant's looking time towards the different ROIs of the presented stimuli. Various measures can be computed from these data based on the requirement of a specific intervention protocol. For example, for the usability study presented here, these times are summed up to get the $Total_{Time}$ from which the percentage of time spent by a participant in looking at Face_ROI is computed.

Such computation is necessary, since in dyadic communication, gaze information underlying one's expressive behavior (i.e., amount of time a speaker and a listener look at each other) plays a vital role in regulating conversation flow, providing feedback, and communicating emotional

Identify applicable sponsor/s here. *(sponsors)* information. Past research has shown that the looking behavior is a function of the cultural upbringing of an individual [31] [32]. Specifically, some cultures may give a negative connotation to looking directly at others' eyes, while others may consider such looking pattern as comfortable. Our system has the flexibility of changing the gaze parameter to suit different cultural requirements. In the present work we use gaze definition suitable for western culture where, a listener looking at the speaker 70% of the time during an interaction has been identified as *'normal while listening'* [33] [34]. Further we used this '*normal while listening*' criterion, since our participants act as listeners while viewing and listening to the avatars giving presentations.

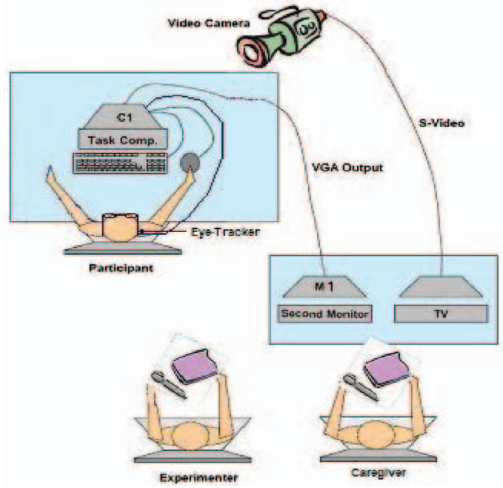

Figure 4. Experimental setup.

Our system is capable of providing a participant with individualized feedback based on the behavioral viewing patterns so as to capture his/her attention to a task. The presented system can generate complex individualized feedback based on one's response to several questions, any performance measures defined for a given task, and his/her actual viewing pattern. However, for the usability study presented here, a simple rule-based system is designed to demonstrate the potential of the developed system. Here a question regarding how the avatar is feeling is asked (Q1), and based on whether or not the participant correctly recognizes the avatar's feeling and how much he/she looked at the avatar's face, an individualized feedback is generated as shown in Table I.

While not utilized in the usability study presented in this paper, the developed system has the capability of communicating with the VR-based platform using our Vizard-Viewpoint interface (Fig. 3) to modify the avatar's states to create an interactive interaction. We have developed three capabilities to date to enhance interaction with the avatar. These are: i) the avatar can change his/her facial expressions during communication with the participant; for example, if the story demands a mood change, the avatar may start with a neutral face and then based on the content of the story gradually change to happy, or angry facial expression. Actually, the system triggers a token that activates the vizard-based 'morph' module to change the percentage of smiling, or frowning, etc. to create such a change. ii) The avatar can walk towards an object of interest, point to an object, and turn and look at an object either as a part of interaction or to bring back the participant's attention based on his/her current looking pattern. The system initiates the "walk to" or "turn and look" or "point" modules that we have developed and modified in vizard. iii) The avatar can also provide verbal feedback during the trial interrupting its original conversation/storytelling if the protocol demands such a feedback. The system initiates a separate conversation thread within vizard for this purpose. These capabilities can be accessed every 33 ms if required. However, in the present usability study we have not fully exploited these extended capabilities of the system since they were not needed for the basic study. It is conceivable that for a sophisticated intervention study, there will be intervention rules as to when and how to change the states of the avatar and such capabilities would be beneficial.

TABLE I. RATIONALE BEHIND INDIVIDUALIZED FEEDBACK

Response to Q1	$t \geq$ **70%**	**VIGART Response [Label]**
Right	Yes	Your classmate really enjoyed having you in the audience. You have paid attention to her and made her feel comfortable. Keep it up!
Right	No	Your classmate did not know if you were interested in the presentation. If you pay more attention to her, she will feel more comfortable.
Wrong	Yes	Your classmate felt comfortable in having you in the audience. However, you may try to pay some more attention to her as she makes the presentation so that you can correctly understand how she is feeling.
Wrong	No	Your classmate would have felt more comfortable if you had paid more attention to her. If you pay more attention, then you will correctly know how she is feeling and make her feel comfortable.

Q1 : Question asked by the system; t : Duration of participant's looking towards the Face_ROI of visual stimulus.

III. METHODS

A. *Participants*

Six adolescents (Male: n = 5, Female: n = 1) with high-functioning ASD (ASD1-ASD6), ages 13-17y (M=15.60y, SD=1.27y) participated in this study. All participants were recruited through existing clinical research programs at Vanderbilt University and had established clinical diagnoses of ASD. Participants were also required to score ≥ 80 on Peabody Picture Vocabulary Test-3rd Edition (PPVT-III) [35] to ensure that language skills were adequate to participate in the current protocol. Data on core ASD related symptoms and functioning was obtained through parent report on Social Responsiveness Scale (SRS) [36] profile sheet and Social Communication Questionnaire (SCQ) [37] with all participants falling above the clinical thresholds. Table II provides a summary of the participant characteristics. The SRS (cutoff T-score ≥ 60T) generates a total score reflecting severity of social deficits in the autism spectrum. The SRS generates a total score reflecting severity of social deficits in the autism spectrum, as well as 5 treatment subscales: Receptive, Cognitive, Expressive, and Motivational aspects of social behavior, and Autistic Preoccupations. The SRS T-score categorizes measurements in the Normal Range (≤ 59T), Mild to Moderate ASD Range (60T-75T), or Severe Range (≥ 76T) [36]. Three of the six participants were in Mild to Moderate ASD range, with the remaining three in the Severe Range of ASD. For the SCQ scores, a cutoff score of 13 is recommended to maximize valid ascertainment of cases of ASD [37]. The SCQ is a parent report questionnaire aimed at evaluating 3 critical autism diagnostic domains of qualitative impairments, namely, reciprocal social interaction, communication, and repetitive and stereotyped patterns of behavior. Our participants were above the clinical threshold on the SRS measure (Table II). The Autism Diagnostic Observation Schedule-Generic (ADOS-G) (cutoff score of 7) is a 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 [38]. On the ADOS-G measure, our participants were in the clinical range with one of them falling marginally. All research procedures were approved by the Vanderbilt University Institutional Review Board.

TABLE II. SUMMARY OF PARTICIPANT CHARACTERISTICS

	Age (years)	**PPVT[a] Standard score**	**SRS[b] Total *T*-score (cutoff=60)**	**SCQ[c] Total score (cutoff=13)**	**ADOS-G Module[d] Total score (cutoff=7)**
Mean (SD)	15.60 (1.27)	102 (15)	78 (11)	22 (7)	11 (5)

[a]Peabody Picture Vocabulary Test-3rd edition Standard score [35]
[b]Social Responsiveness Scale Total T-score [36]
[c]Social Communication Questionnaire Total score [37]
[d]Autism Diagnostic Observation Schedule-3rd Edition [38]
SD – Standard Deviation.

B. Procedure

Each participant participated in an approximately 50 min laboratory visit. First a brief adaptation session was carried out. In the first phase of adaptation, before the participants and their caregivers were asked to sign the assent and the consent forms respectively, the experimenter briefed the participant about the experiment, showed the experimental setup, and the eye-tracker goggles. This phase ran for approximately 10 min. In the second phase of adaptation, the participant sat comfortably on a height-adjustable chair and was asked to wear the eye-tracker goggles. The chair was adjusted so that his/her eyes were collinear with center of task computer, C1 (Fig. 4). The experimenter told the participant that he/she could choose to withdraw anytime from the experiment for any reason, especially if he/she was not comfortable interacting with the system. The participant was then asked to rest for 3 min to acclimate him/her to the experimental set-up. This second phase of adaptation took approximately another 10 min. Then the eye-tracker was calibrated. The average calibration time was approximately 15 s in which the participant sequentially fixated on a grid of 16 points displayed randomly on the task computer. The participants viewed an initial instruction screen followed by an interaction with their virtual classmate (i.e., the avatar) narrating a personal story. Each storytelling trial was approximately 3 min long. The participants were asked to imagine that the avatars were his/her classmates at school giving presentations on several different topics. They were informed that after the presentations they would be required to answer a few questions about the presentation. They were also asked to try and make their classmate feel as comfortable as possible while listening to the presentation. However, it was not explicitly stated that in a presentation a speaker feels good when the audience pay attention to him/her (by looking towards the speaker). The idea here was to give feedback to the participants about their viewing patterns and thereby study how that affects the participants as the task proceeded. The experiment began with Trial1 with the virtual classmate exhibiting a 'neutral' facial expression and narrating a personal story. This trial was followed by 4 other trials that were similar to the Trial1 except that in these subsequent trials the virtual classmate displayed 'happy' (Fig. 1b) or 'angry' (Fig. 1c) facial expression to capture the mood inherent in the content of the story. In the present study we used 3 female, and 2 male avatar heads. We randomized the Trial2 – Trial5 among the participants while the avatars displayed context-relevant happy or angry facial expression. After each trial, the participant was asked an emotion-identification question (Q1) and a story-related question (Q2). The Q1 was about the virtual classmate's emotion which had 3 answer choices (A. Happy, B. Angry, C. Not Sure). The Q2 was about some basic facts as narrated in the story. It also had 3 answer choices. The participant responded with a keypad. Q2 was asked to encourage a participant to pay attention to the story. Depending on the participant's response to Q1 and how much attention he/she paid to the virtual classmate, as measured by the real-time computation of the percentage of time spent in looking at the classmate's face, our system encouraged the participant to either pay more or keep the same attention towards the presentation (Table I). After each trial, the observer (e.g., the caregiver) rated about what he/she thought about how engaged the participant was during the VR-based social interaction using a 1-9 scale (1 - least engagement, 9 - most engagement). Each participant was compensated in the form of gift cards for completing a session.

IV. RESULTS

A usability study was carried out with six participants with ASD (ASD1-ASD6). The implication of gaze-based individualized feedback on their engagement level, behavioral viewing patterns of the participants in terms of the set of quantitative indices (e.g., Sum of Fixation Counts (SFC), Total Fixation Duration (FD_{Total})) of their viewing pattern distributed over the different ROIs of the visual stimulus were investigated while they participated in the virtual social tasks. We used these behavioral indices to interpret the participant's behavioral viewing pattern from Trial1 (i.e., the trial preceded by No Feedback) to Trial5 (i.e., the last trial which was preceded by 4 Feedbacks during the previous trials). Also we investigated the scan paths of the participants.

A. Impact of gaze-based dynamic feedback on Participants' Engagement (based on Observers' rating)

We wanted to assess whether our system can be used in virtual social interaction to create different engagement levels among the participants so that engagement manipulation using individualized feedback could be potentially feasible in the future as a part of intervention. In our usability study with the system, the participants' caregivers rated as to what they felt regarding the participants' engagement level while participating socially with their virtual classmates. As engagement can be represented in terms of "sustained attention to an activity or person" [39], we asked the caregivers to rate the participants' engagement by observing their attention to the social task by using a 1-9 scale.

With individualized feedback during VR-based social task, the reported group engagement mean (Table III) improved during Trial5 from Trial1. For all participants (except ASD2) the engagement rating improved from Trial1 to Trial5. Further analysis revealed that ASD2 was incorrect in responding to story-related question in Trial5 which may be due to his lower engagement. The caregiver of ASD2 reported that he liked the story in Trial1 the most and the Trial5 the least. Also the range (1-9 scale) of engagement rating shows that group engagement increased during Trial5.

TABLE III. IMPACT OF GAZE-BASED INDIVIDUALIZED FEEDBACK ON PARTICIPANTS' ENGAGEMENT.

	Reported Observer rating on Engagement (Full Range: 1-9)	
	Trial1	**Trial5**
ASD1	2	5
ASD2	7	6
ASD3	4	7
ASD4	6	7
ASD5	4	5
ASD6	4	7
Mean	4.50	6.17
Range	2 – 7	5 – 7

B. Implication of gaze-based individualized feedback on Behavioral Viewing Patterns of the participants in terms of Attention to the Faces of the Avatars

We chose to use the behavioral viewing indices (e.g., Sum of Fixation Counts, and the Total Fixation Duration) of the participants, as they viewed the Face_ROI of the avatars while attending to the avatars' presentations to infer attention toward social stimuli in the VR environment. Results indicate that the participants looked more frequently towards the face region (Face_ROI) of the avatars from the Trial1-to-Trial5 measurement. This is reflected from the improvement in the Sum of Fixation Counts for each participant from Trial1 to Trial5 measurement for Face_ROI viewing with individualized feedback (Table IV) with Sum of Fixation Counts for Face_ROI viewing during Trial1 being statistically significantly different ($t = 3.464$; $p = 0.0180$) from that during Trial5 by using a dependent sample t-test between these two groups.

Also, the Fixation Duration of the participants was analyzed while viewing Face_ROI due to its importance as an indicator of social engagement [40]. The FD_{Total} of the participants for the Face_ROI was computed during viewing and the results indicate increase in this index for all of the participants from Trial1 to Trial5 and in statistically different ways ($t = 8.068$; $p = 0.0005$) by using a dependent sample t-test between these two groups. (Table IV). Overall, the results reflect a trend for the participants to not only fixate on the Face_ROI more frequently, but also for a longer duration with dynamic feedback.

TABLE IV. IMPROVEMENT IN VIEWING PATTERN IN TERMS OF SUM OF FIXATION COUNTS, AND TOTAL FIXATION DURATION WHILE VIEWING FACE_ROI

	SFC for Face_ROI viewing			**FD_{Total} for Face_ROI viewing**		
	Trial1 (no.)	**Trial5 (no.)**	**%Impro -vement**	**Trial1 (s)**	**Trial5 (s)**	**%Improve -ment**
ASD1	314	400	27.39	98.80	135.72	37.37
ASD2	418	494	18.18	141.87	179.49	26.52
ASD3	411	564	37.23	122.35	170.39	39.27
ASD4	427	501	17.33	146.37	172.78	18.05
ASD5	214	543	153.74	58.27	121.17	107.93
ASD6	140	253	80.71	41.31	77.78	88.29
Mean (SD)	321 (121)	459 (116)	55.76 (53.40)	101.49 (43.76)	142.89 (39.38)	52.91 (36.38)

SFC: Sum of Fixation Counts; FD_{Total}: Total Fixation Duration; SD: Standard Deviation.

C. Implication of gaze-based individualized feedback on Behavioral Viewing Patterns of Participants in terms of Scanning of the total Visual Stimulus

The percent of total fixation duration towards Face_ROI, as compared to the Object_ROI and Others_ROI improved (Fig. 5) for all participants implying that each participant looked at avatar's face for a longer duration of time during the Trial5 than during Trial1. Thus, with the gaze-based individualized feedback, the participants attended to the Face_ROI of the avatars more than the non-face regions (i.e., the Object_ROI and the Others_ROI).

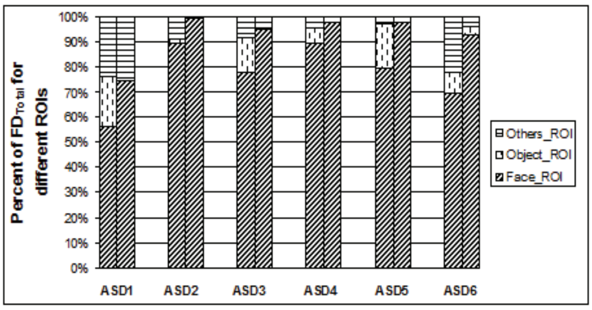

Figure 5. Comparative Analysis of Total Fixation Duration (FD_{Total}) for Face_ROI, Object_ROI, and Others_ROI viewing for each Participant (Left bars indicate Trial1 and Right bars indicate Trial5).

We also studied the effect of individualized feedback on the scan paths of the participants as children with ASD have been shown to exhibit atypical scan paths during social interaction [41]. Our investigation revealed that all participants fixated more on the Face_ROI of the avatars, with reduced attention towards the Object_ROI and the Others_ROI, during Trial5 as compared to Trial1. For example, as is evident from the scan path (Fig. 6), ASD3 fixated on different ROIs of the visual stimulus during the Trial1. However, during Trial5, ASD3 fixated mainly on the Face_ROI and much less on the Object_ROI and Others_ROI. Note that, these scan paths were analyzed in the background and they were not visible to the participant.

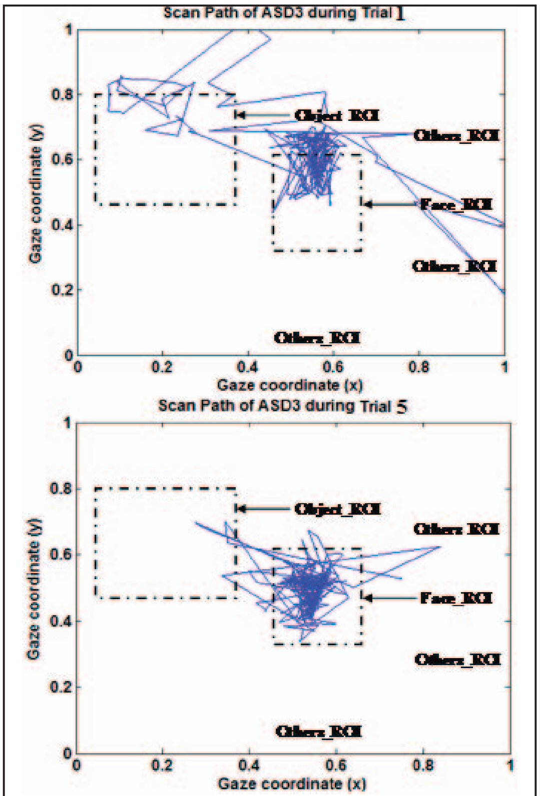

Figure 6. Improvement in Scan Path of ASD3 distributed over different ROIs during Trial1 and Trial5.

V. CONCLUSION

In the present work we developed a technology-based system that seamlessly integrates VR system with real-time eye-gaze measurement to provide individualized feedback during one's social communication tasks presented in a VR environment. In addition, we designed a usability study with a small sample of six adolescents with ASD to examine the applicability of such a system and also to investigate how they respond to such a system in terms of their behavioral viewing patterns, e.g., Sum of Fixation Counts, Fixation Duration, and Scan Path. The preliminary results of the usability study are promising. At this current configuration, our system uses a wearable eye-tracker, which may not be suitable for children with low-functioning ASD. However, in the future, we plan to use a non-contact desktop-based eye-tracker to mitigate this problem. It is worth mentioning that the presented system is capable of providing extensive real-time social interaction. For example, it can provide continuous information to the avatars about whether and for how long the participant is looking at them during interaction, and based on this information the avatars can change their mode of talking to bring his/her attention back and so on. Thus we plan to make the system prompt the avatar to change its behavior and respond adaptively and in real-time to the participant while he/she enters into bidirectional social conversation with the avatar in the virtual environment featuring a higher level of interactivity between the participant and the system. Such sophisticated interaction will be performed in the future. The present usability study shows, in principle, that such a dynamic gaze-based virtual social system has the potential to be used as a supplement to real-life social skills training tasks in an individualized and intensive manner. However, a much larger study must be conducted before such findings can be generalized.

ACKNOWLEDGMENT

The authors would like to thank the participants and their families for making this study possible. We gratefully acknowledge the support of Autism Speaks Pilot Study Grant 1992, National Science Foundation Grant 0967170 and National Institute of Health Grant 1R01MH091102-01A1 to conduct this study.

REFERENCES

[1] APA, "Diagnostic and statistical manual of mental disorders: DSM-IV-TR," American Psychiatric Association, Washington, DC, 2000.
[2] M. Carpenter, B.F. Pennington, and S.J. Rogers, "Interrelations among social-cognitive skills in young children with autism," Journal of Autism and Developmental Disorders, vol. 32, no. 2, pp. 91–106, 2002.
[3] L. Kanner, "Autistic disturbances of affective contact," Nervous Child, vol. 2, pp. 217–250, 1943.
[4] S. Wieder, and S. Greenspan, "Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective?," The Journal of Developmental and Learning Disorders, vol. 9, 2005.
[5] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, D. "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," Archives of General Psychiatry, vol. 59, no. 9, pp. 809-816, 2002
[6] C.Y. Trepagnier, M.M. Sebrechts, A. Finkelmeyer, W. Stewart, J. Woodford, and M. Coleman, "Simulating social interaction to address deficits of autistic spectrum disorder in children," Cyberpsych Behav, vol. 9, no. 2, pp. 213-217, 2006.
[7] H. Cohen, M. Amerine-Dickens, and T. Smith, "Early intensive behavioral treatment: replication of the UCLA model in a community setting," J Dev Behav Pediatr, vol. 27, no. 2 Suppl, pp. S145-155, 2006.
[8] M.S. Goodwin, "Enhancing and accelerating the pace of Autism Research and Treatment: The promise of developing Innovative Technology," Focus on autism and other developmental disabilities, vol. 23, pp.125-128, 2008.
[9] K. Blocher, and R.W. Picard, "Affective social quest: emotion recognition therapy for autistic children," In K. Dautenhahn, A. H. Bond, L. Canamero, & B. Edmonds (Eds.), Socially Intelligent Agents: Creating Relationships with Computers and Robots. Kluwer Academic Publishers, 2002.
[10] S. Parsons, P. Mitchell, and A. Leonard, "The use and understanding of virtual environments by adolescents with autistic spectrum disorders," J Autism Dev Disord, vol. 34, no. 4, pp. 449-466, 2004.
[11] A. Tartaro, and J. Cassell, "Using Virtual Peer Technology as an Intervention for Children with Autism," In J. Lazar (Ed.), Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations. UK: John Wiley and Sons, Chichester, 2007.
[12] H. Kozima, C. Nakagawa, and Y. Yasuda, "Interactive robots for communication-care: A case-study in autism therapy," IEEE International Workshop on Robot and Human Interactive Communication, pp. 341-346, 2005.
[13] D. Strickland, "Virtual reality for the treatment of autism," In G. Riva (Ed.), Virtual reality in neuropsycho-physiology (pp81-86). Amsterdam: IOS Press, 1997.
[14] D.J. Moore, P. McGrath, and J. Thorpe, "Computer aided learning for people with autism - A framework for research and development," Innovations in Education and Training International, vol. 37, no. 3, pp. 218-228, 2000.
[15] V. Bernard-Opitz, N. Sriram, and S. Nakhoda-Sapuan, "Enhancing social problem solving in children with autism and normal children through computer-assisted instruction," Journal of Autism and Developmental Disorders, vol. 31, no. 4, pp. 377-384, 2001.

[16] S. Parsons, and P. Mitchell, "The potential of virtual reality in social skills training for people with autistic spectrum disorders," J Intellect Disabil Res, vol. 46(Pt 5), pp. 430-443, 2002.
[17] W.R. Sherman, and A.B. Craig, "Understanding virtual reality: interface, application, and design," Boston: Morgan Kaufmann Publishers, 2003.
[18] S. Kerr, and K. Durkin, "Understanding of thought bubbles as mental representations in children with autism: Implications for theory of mind," Journal of Autism and Developmental Disorders, vol. 34, no. 6, 2004.
[19] J.J. Cromby, P.J. Standen, and D. J. Brown, "The potentials of virtual environments in the education and training of people with learning disabilities," Journal Intellectual Disability Research, vol. 40, pp. 489-501, 1996.
[20] M. Wilms, L. Schilbach, U. Pfeiffer, G. Bente, G.R. Fink, and K. Vogeley, "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. Vol. 5, no. 1, pp. 98-107, 2010.
[21] L. Itti, N. Dhavale, and F. Pighin, "Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention," Proc. Intl. Sym. on Opt. Science & Tech., vol. 5200, pp. 64-78, 2003.
[22] E.T. Hall, "The Hidden Dimension," New York: Doubleday & Company, Inc., 1966.
[23] M. Argyle, & J. Dean, "Eye-contact, distance and Affiliation," Sociometry, vol. 28, no. 3, pp. 289-304, 1965.
[24] A.F. Pereira, C. Yu, L.B. Smith, and H. Shen, "A First-Person Perspective on a Parent-Child Social Interaction During Object Play," Proc. 31st Annual Meeting of Cog. Sc. Society, Amsterdam, 2009.
[25] Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (dibels.uoregon.edu/measures/), 2007.
[26] R.J.K. Jacob, "Eye Tracking in Advanced Interface Design", in Advanced Interface Design and Virtual Environments, ed. W. Barfield and T. Furness, Oxford University Press, Oxford, 1994.
[27] J.W. Denver, "The Social Engagement System : Functional Differences in Individuals with Autism," Dissertation. www.lib.umd.edu/drum/bitstream/1903/1351/1/umi-umd-1486.pdf, 2004.
[28] A. Poole, and L.J. Ball, "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects," In C. Ghaoui (Ed.), Encyclopedia of Human Computer Interaction, Idea Group, 2005.
[29] M.A. Just, and P.A. Carpenter, "Eye fixations and cognitive processes," Cognitive Psychology, vol. 8, pp. 441-480, 1976.
[30] C.J. Anderson, J. Colombo, and D.J. Shaddy, "Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder," Journal of Clinical and Experimental Neuropsychology, vol. 28, pp. 1238–1256, 2006.
[31] H. Z. Li, "Culture and gaze direction in conversation," RASK, vol. 20, pp. 3-26, 2004.
[32] C.L. Kleinke, "Gaze and eye contact: A research review," Psychological Bulletin, vol. 100, pp. 78-100, 1986.
[33] M. Argyle, and M. Cook, "Gaze and Mutual Gaze," Cambridge, MA: Cambridge Univ. Press, 1976.
[34] A. Colburn, S. Drucker, and M. Cohen, "The role of eye-gaze in avatar-mediated conversational interfaces," In: SIGGRAPH Sketches and Applications, New Orleans, Louisiana, USA, 2000.
[35] L.M. Dunn, and L.M. Dunn, "PPVT-III: Peabody Picture Vocabulary Test," 3rd ed. Circle Pines, Minnesota: American Guidance Service, 1997.
[36] J.N. Constantino, "The Social Responsiveness Scale," California: Western Psych. Serv., 2002.
[37] M. Rutter, A. Bailey, S. Berument, C. Lord, and A. Pickles, "Social Communication Questionnaire," Los Angeles, California: Western Psychological Services, 2003.
[38] C. Lord, S. Risi, L. Lambrecht, E.H. Cook, B.L. Leventhal, P.C. DiLavore, A. Pickles, and M. Rutter, "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, vol. 30, no. 3, pp. 205-223, 2000.
[39] NRC, "Educating children with autism," National Academy Press, Washington, DC, 2001.
[40] W. Jones, K. Carr, and A. Klin, "Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder," Arch. Gen. Psychiatry, vol. 65, no. 8, pp. 946-954, 2008.
[41] M.D. Rutherford, and M.T. Towns, "Scan Path differences and similarities during Emotion Perception in those with and without Autism Spectrum Disorders," J. Aut. Dev Disord, vol. 38, pp. 1371-1381, 2008.