Design of a Virtual Reality Based Adaptive Response Technology for Children With Autism

Uttama Lahiri, *Member, IEEE*, Esubalew Bekele, *Student Member, IEEE*, Elizabeth Dohrmann, Zachary Warren, and Nilanjan Sarkar, *Senior Member, IEEE*

Abstract—Children with autism spectrum disorder (ASD) demonstrate potent impairments in social communication skills including atypical viewing patterns during social interactions. Recently, several assistive technologies, particularly virtual reality (VR), have been investigated to address specific social deficits in this population. Some studies have coupled eye-gaze monitoring mechanisms to design intervention strategies. However, presently available systems are designed to primarily chain learning via aspects of one's performance only which affords restricted range of individualization. The presented work seeks to bridge this gap by developing a novel VR-based interactive system with Gaze-sensitive adaptive response technology that can seamlessly integrate VR-based tasks with eye-tracking techniques to intelligently facilitate engagement in tasks relevant to advancing social communication skills. Specifically, such a system is capable of objectively identifying and quantifying one's engagement level by measuring real-time viewing patterns, subtle changes in eye physiological responses, as well as performance metrics in order to adaptively respond in an individualized manner to foster improved social communication skills among the participants. The developed system was tested through a usability study with eight adolescents with ASD. The results indicate the potential of the system to promote improved social task performance along with socially-appropriate mechanisms during VR-based social conversation tasks.

Index Terms—Autism spectrum disorder (ASD), blink rate (BR), eye-tracking, fixation duration (FD), pupil diameter (PD), virtual reality (VR).

I. INTRODUCTION

AUTISM spectrum disorder (ASD) is a neurodevelopmental disorder characterized by core deficits in social interaction and communication accompanied by restricted patterns of interest and behavior [1]. Individuals with ASD commonly demonstrate difficulties related to complex fluid reciprocal social interactions, including difficulties related to social information processing and understanding nonverbal aspects of communication [2]. Given these vulnerabilities, a specific focus of autism research has been to understand how people with autism process salient social and emotional cues from faces [3], [4]. The ability to derive socially relevant information from faces is considered to be a very important skill for facilitating fluid interpersonal communication [5]. In addition to providing indexes of social information processing during interactions, eye-gaze and physiological measurement of aspects of that gaze may simultaneously provide markers for understanding how an individual is engaging with a task. Specifically, such markers may provide cues as to whether an individual is attending to specific aspects of interaction and for the nature of his affective experiences (e.g., enjoyment, anxiety, boredom, etc.) during tasks. Such a capability may enable the development of intelligent systems capable of providing real-time adaptations that greatly bolster improvements in the core areas of deficit related to ASD [6]–[8].

In recent years, virtual reality (VR) has been proposed as a viable medium to administer ASD intervention through VR-based human–computer interaction tasks [9], [10]. This is because VR technology possesses several strengths namely, malleability, controllability, replicability, modifiable sensory stimulation, and an ability to pragmatically individualize intervention approaches and reinforcement strategies [11]. A number of VR applications have investigated social skill training for children with ASD [9], [12]. While these studies are innovative by pioneering the applicability of VR to ASD intervention, they have not yet demonstrated the capacity to index real-time viewing patterns to assess engagement. This is important, because engagement defined as "sustained attention to an activity/person" [13] is one of the key factors for children to make substantial gains in communication and social domains [14]. Recently, Wilms *et al.* [15] demonstrated the feasibility of linking the gaze behavior of a virtual character with a human observer's gaze position during joint-attention tasks. However, the current VR environments as applied to assistive intervention for children with ASD are designed to chain learning via aspects of performance alone (i.e., correct, or incorrect) thereby limiting individualization of application. Even these recent systems that measure eye gaze do not adaptively respond to engagement predicted from looking patterns or physiological indexes, which could provide important cues in this regard. For example, one's looking pattern quantified as fixation duration (FD) while looking towards the face of the speaker has been

shown to be a predictor of engagement [4]. Several eye physiological indexes, such as blink rate (BR) and pupil diameter (PD) have been considered as reliable autonomic measures of one's engagement. Studies have reported spontaneous inhibition in BR [16] and PD constriction [4] with increased engagement to tasks.

The primary goal of our research was to develop a VR-based ASD technology with potential relevance to ASD intervention. Specifically, in this work we attempted to create a tool that could 1) allow real-time measurement of one's looking pattern and associated eye physiology when engaged in a VR based social communication task; and 2) adapt the tasks in this environment and provide individualized feedback based on the inferences regarding engagement made from these measurements. Realization of such a technology may pave the way for more intensive, intelligent, and individualized intervention paradigms by designing enhanced, individualized feedback and reinforcement strategies during VR-based social skill training tasks. Also, such a facility would provide an engagement profiling system that is capable of adapting to one's predicted engagement level in controlled environments and thereby reinforcing skills in core domains gradually but automatically, which could in turn be applied to a large class of potential intervention paradigms for individuals with ASD.

In this context, our initial work [6] reported the development of a novel VR-based social system integrated with computationally enhanced eye-tracking technology that could seamlessly integrate a VR platform with eye gaze measurement technique to measure one's looking pattern and eye physiological indexes in real-time during a VR-based social communication task. However, there were two main limitations in that work. First, the social communication task was basically passive in the sense that the participant with ASD only viewed and listened to a story narrated by the avatar and then answered a few questions with no bi-directional social discourse between the avatar and the participant. As such, it was a very limited social communication task with the avatar having no ability to respond to a question asked by the participant. Second, the system was not able to respond adaptively, based on one's engagement during a VR-based interaction. The second limitation is critical since without the ability to respond adaptively based on one's engagement, it would be difficult to deploy the technology for targeted and personalized intervention and cannot be incorporated into complex intervention paradigms aimed at improving functioning and quality of life for individuals with ASD. The present work overcomes these limitations.

This paper is organized as follows. In Section II, we describe our objectives. In Section III, we present the system design. Section IV discusses the experiment conducted and the methods used. Section V presents the results obtained in our usability study. Section VI summarizes the contributions of this work and indicates the direction of future work.

II. OBJECTIVES AND SCOPE

The objective of this paper is two-fold: 1) present the design and development of a VR-based engagement-sensitive system (ES) with adaptive response technology that is based on the composite effect of one's viewing pattern, eye physiology, and

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

performance metric which can be applied to social communication task for children with ASD; and 2) present the results of a usability study designed to analyze the difference between the ES and a system that is based on one's performance metric alone [performance-sensitive system (PS)]. While not an intervention study, here we present the results of a usability study as a proof-of-concept of our designed system.

III. SYSTEM DESIGN

The VR-based system with adaptive response technology comprised of 1) a VR-based social communication task module, 2) a real-time eye-gaze monitoring module, and 3) an individualized adaptive response module utilizing a rule-governed intelligent engagement prediction mechanism.

A. VR-Based Social Communication Task Module

In this work, we used desktop VR applications, because they are accessible, affordable, and minimize potentially distressing sensory response for this ASD population [17]. Vizard [18], a commercially available VR design package, was used to develop the virtual environments and the assistive technology. We developed social situations with context-relevant backgrounds, and avatars whose age and appearance resembled those of the participants' peers without trying to achieve exact similarities. Also, we designed conversation threads for bi-directional social communication between the avatars and the participants. Our social communication task module was comprised of 1) a task presentation module, and 2) a bi-directional conversation module.

1) Design of a VR-Based Task Presentation Module: We developed 24 social tasks in which avatars narrated 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, etc.). These stories were adopted from an online database [19] of term papers comprising essays and research papers written by teenagers based on their personal experience. We conducted a small focused survey of teenagers from the neighboring areas to select stories that they found interesting, out of the pool of stories. The voices for the avatars were recorded from teenagers. An avatar could make pointing gestures and move dynamically in a context-relevant VR environment. For example, when an avatar narrated his trip to a beach, the VR environment reflected the view of the beach [Fig. 1(a)]. When the avatar narrated some of his favorite activities on the beach such as, tanning during the day, the VR world displayed such a situation [Fig. 1(b)] to the participant. Subsequently, when the avatar narrated his experience of the remarkable view of the sunset he witnessed on the beach, the VR scene
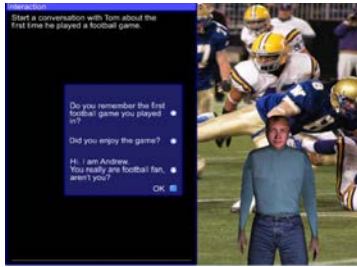
Fig. 2. Snapshot of a bi-directional conversation module.

changed smoothly to display such a situation [Fig. 1(c)]. Thus, realistic situations, relevant to the topic being narrated by the avatar, were presented to the participants. We used 12 avatars, six males and six females, distributed randomly over the 24 task presentation modules. Avatar heads were created from 2-D photographs of teenagers, which were then converted to 3-D heads by "3DmeNow" software for compatibility with Vizard. The 12 most neutral heads from a sample of 26 heads based on a survey of 20 undergraduate students [20] were chosen for this study. One can view the avatars within our system from first-person perspective while the avatars narrate personal stories, which is comparable to research on social anxiety and social conventions [21].

2) Design of Bi-Directional Conversation Module: Our system was designed to encourage social interaction in the form of bi-directional conversation because conversation between interlocutors has been recognized as a fundamental vulnerability related to ASD itself [22]. In our study, the participant was asked to watch and listen to an avatar narrating a personal story during the VR-based task presentation. Then the participant was asked to extract a piece (or multiple pieces) of information from the avatar using a bi-directional conversation module with varying levels of interaction difficulty (e.g., Type 1—"Easy," Type 2—"Medium," and Type 3—"High"). The bi-directional conversation module followed a menu-driven structure used by the interactive fiction community [23], in which the participant was required to converse with the avatar by choosing statements/questions in a particular sequence. Fig. 2 shows an example. One side of the screen displayed the avatar while the menu of questions appeared on the other side. We monitored eye gaze of the participant during this conversation to assess how much time was spent on looking at the face of the avatar. The menu-driven structure was used for developing conversation threads because it offered a number of advantages, e.g., dimensional control by limiting the range of inputs [24], and providing guidance to the participant thereby minimizing repetitive dialogue [23]. The lawnmower problem in which a user can see the entire conversation tree by moving through all the menu branches and which is a disadvantage associated with menu-driven systems [23] was overcome in our present work by preventing the participant from tracing back the conversation tree. The degree of interaction difficulty was controlled by requiring the number of questions a participant needed to ask in order to obtain a desired piece of information from the avatar, and by the nature of conversation. In the present work, we chose conversation tasks with three levels of

interaction difficulty. If a participant could acquire the needed information in a particular interaction difficulty level by asking the minimum number of questions for that level, he would achieve the highest performance score. On the other hand, if the participant could not choose the right questions in the right sequence, causing him to ask more questions to find the needed information, he would acquire a proportionately lower performance score. After a certain number of attempts if the participant was still unable to obtain the right information, the system would terminate the task and provide a task of lesser challenge to the participant. Note that our system is not limited to either how many questions can be asked or the types of conversation that can be constructed. As such, these factors can be modified in the future depending on the intervention paradigm.

In order to ensure consistency among the 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. These clinicians work at Vanderbilt University TRIAD (Treatment and Research Institute for Autism Spectrum Disorders) and have extensive experience in providing therapy to children and adolescents with ASD. There were two kinds of interaction between the avatar and the participant depending on the session. In the performance-based session (i.e., using the PS), the avatar only answered the questions asked by the participant. If the participant asked an inappropriate question, the avatar tried to guide him towards the correct question as a part of the response, thereby serving the role of a facilitator. In the engagement-based session (i.e., using the ES), on the other hand, the avatar not only served the role of a facilitator in the same way as the PS, but also the avatar was made aware of the participant's looking pattern during the VR-based social conversation. Thus, while using the ES, the system provided individualized feedback based on the participant's viewing pattern (details in Section III-C2).

Both the PS and ES systems featured social conversation task modules categorized into easy, medium, and high levels of interaction difficulty. In an easy (Type 1) task, which was considered as an interaction with lowest difficulty level, one was required to ask/make three correct questions/comments in a correct sequence to get a single piece of information from the avatar. The medium (Type 2) task required five correct questions/comments to be asked in the correct sequence to obtain multiple pieces of information from the avatar. The high (Type 3) task required seven 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.

B. Real-Time Eye-Gaze Monitoring Module

The system captures eye data of a participant interacting with an avatar using eye-tracker goggles from Arrington Research [25]. This eye-tracker comes with some basic features (e.g., it provides raw pupil diameter, raw pupil aspect ratio i.e., ratio of minor to major axes of pupil image, etc.) acquiring capability for offline analysis. Its video capture module has a refresh rate of 30 Hz (in precision mode) to acquire a participant's gaze data using the "Viewpoint" software. We designed the Viewpoint-Vizard handshake module to serve as an interface
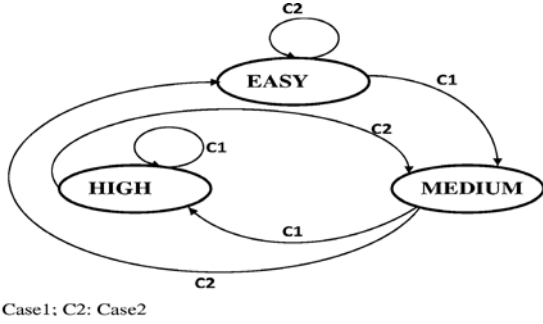
C1: Case1; C2: Case2

Fig. 3. Finite state machine representation of dynamic decision task switching based only on performance PS.

between the eye-tracker and the VR programming platforms. We acquired the raw gaze data using Viewpoint, transformed it to the Vizard compatible format using the handshake interface at a refresh rate of 30 Hz in a time synchronized manner. Subsequently, we applied signal processing techniques to eliminate noise and extract the relevant features at an interval of 33 ms. The Gaze Database was processed to extract three features: mean pupil diameter (PD_{MEAN}), mean blink rate (BR_{MEAN}), and average fixation Duration (FD_{AVG}) for each region of interest (e.g., face region of the avatar, etc.) from each segment of the signals monitored (for details on feature extraction, please refer [6]).

C. Individualized Adaptive Response Module

1) Performance-Sensitive System (PS): For the PS, a task-switching mechanism adjusted the interaction difficulty by switching tasks among Type 1, Type 2, and Type 3 tasks (Section III-A2) based on one's performance in the VR-based social communication task (details in Section IV-D). Note that these strategies to classify performance were chosen as a first approximation to quantify social interaction with full recognition that such quantification could be varied in future study.

In order to individualize social interaction, in this mode, we present a task modification strategy based on one's performance metric alone. We implemented the dynamic task switching by a finite state machine representation [26] (Fig. 3). Fig. 3 shows that our interactive system features three levels of difficulty ("Easy," "Medium," and "High"). When a participant's task performance is "Adequate," the task progression continues stepwise while increasing the task difficulty level (represented by C1 in Fig. 3). But, if on the other hand, the participant's task performance is "Inadequate," then the system lowers the task difficulty (represented by C2 in Fig. 3). This task switching continues until the number of tasks in a particular difficulty level is exhausted.

2) Engagement-Sensitive System (ES): For the ES, however, our goal was to switch task not only based on performance but also on how much the individual was engaged in the task. In other words, it was not sufficient to move to a higher level of task simply based only on one's performance—additionally, one had to look at the avatar appropriately to indicate that he was engaged in the conversation. One's engagement to the VR-based

social task was predicted based on objective metrics such as, dynamic viewing patterns characterized by FD and two eye physiological indexes—BR, and pupil diameter (PD) [13], [27], [28]. In order to discretize the engagement space, we assigned numeric values of 1, 2, and 3 [Table I(a), (b), and (c)] to these indexes to quantify engagement in three levels—"not engaged," "moderately engaged," and "highly engaged," respectively. Although there is tremendous individual difference in gaze patterns, there has been some work attempting to quantify typical looking patterns while listening to conversation with percentages of approximately 70% suggested by other groups as common patterns of gaze during conversation. [29], [30]. For this protocol, as can be seen from Table I(a), if the percent FD while looking towards the face of the communicator during VR-based social conversation is greater than 70%, we allocate a value of 3 (implying "highly engaged"). Also, there are studies that have reported spontaneous inhibition in BR [16] and decrease in PD [4] with increased engagement to a social task. Thus, in our present work, besides the FD of the participants, we have also considered the PD and BR, as can be seen from Table I(b) and (c). Note that the discretization through use of numeric values [Table I(a), (b), and (c)] is a first approximation to quantify engagement. Also note that while designing our system, we tested the system with typically developing children. We chose the variation of the eye physiological indexes through trial and error. We used this incremental increase/decrease in one's engagement level as predicted from one's looking pattern (such as, fixation duration) and eye physiological indexes (e.g., PD and BR) during a social task trial with respect to his previous task to encourage progressive improvement from one task trial to the next. This is similar to the computer games as designed by the interactive fiction community where the computer decides to move a player from one level of difficulty to the next based on his performance during each game trial. Our system is not limited to the three levels of engagement—a much finer resolution is possible depending on the task requirement. A rule-governed task-switching mechanism adjusted the interaction difficulty by switching tasks among Type 1, Type 2, and Type 3 tasks based on the composite effect of one's engagement (as predicted from the FD, PD, and BR) and performance in the VR-based social communication task. If the cumulative sum of the engagement level obtained by real-time monitoring of FD, PD, and $BR \geq 6$, then the engagement is considered as "Good Enough;" otherwise this is considered as "Not Good Enough."

The ES fuses the information on the engagement (i.e., "Good Enough" or, "Not Good Enough") and the task performance (i.e., "Adequate," or "Inadequate") to dynamically switch tasks of different difficulty levels by implementing an individualized task modification strategy (Table II). There are two cases when both the metrics—engagement and task performance—agree with each other and the rule design becomes easier. If one's engagement to a social task based on eye gaze indexes is "Good Enough" and the performance is "Adequate," then our system recommends "Move Up" action (Case 1). In this case, the task progression will continue stepwise to increase the task difficulty level. On the other hand, if one's engagement to a social task is "Not Good Enough" and the performance is "Inadequate," then our system recommends "Move Down" action (Case 4).

TABLE I
(A) PREDICTION OF ENGAGEMENT BASED ON FIXATION DURATION,
(B) PREDICTION OF ENGAGEMENT BASED ON PUPIL DIAMETER,
(C) PREDICTION OF ENGAGEMENT BASED ON BLINK RATE

a

Fixation Duration	Value
$0\% \leq T \leq 50\%$	1
$50\% < T < 70\%$	2
$T \geq 70\%$	3

T : Percent FD (Fixation Duration) towards Face region i.e., Face_ROI (during conversation) out of total FD.

b

Pupil Diameter	Value
$PD_{Now} > PD_{Prev}$	1
$PD_{Prev} \geq PD_{Now} \geq 0.95 PD_{Prev}$	2
$PD_{Now} < 0.95 PD_{Prev}$	3

PD_{Now}, PD_{Prev}: Pupil Diameter during present and previous situation respectively.

c

Blink Rate	Value
$BR_{Now} > BR_{Prev}$	1
$BR_{Prev} \geq BR_{Now} \geq 0.95 BR_{Prev}$	2
$BR_{Now} < 0.95 BR_{Prev}$	3

BR_{Now}, BR_{Prev}: Blink Rate during present and previous situation respectively.

TABLE II
TASK MODIFICATION STRATEGY BASED ON COMPOSITE EFFECT OF
BEHAVIORAL VIEWING, EYE PHYSIOLOGY, AND PERFORMANCE 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 III
RATIONALE BEHIND FEEDBACK BASED ON ONE'S FIXATION PATTERN

Fixation Duration	Feedback
$T \geq 90\%$	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.
$90\% > T \geq 70\%$	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.
$T \leq 30\%$	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.
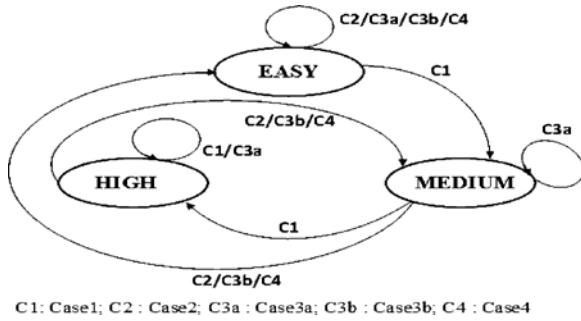

Fig. 4. State machine representation of dynamic decision task switching based on composite effect of behavioral viewing, eye physiology, and performance ES.

In this case the task may be too difficult for the participant and the strategy generator will reduce the task difficulty level. However, when the two metrics contradict, it becomes difficult to design the rules. In these cases, we assign more importance to the task performance metric. If one's engagement to a social task is "Not Good Enough" but the performance is "Adequate," then our system recommends "Move Same" action (Case 3a) or "Move Down" action (Case 3b). Thus, the strategy generator will maintain tasks at the same level of difficulty ("Move Same") and look out for an improvement in the next cycle (Case 3a). In case of no further improvement in the next cycle, the strategy generator will reduce the task difficulty level (Case 3b). Finally, if one's engagement to a social task is "Good Enough" and the performance is "Inadequate," then our system recommends "Move Down" (Case 2). In this case, the strategy generator will decrease the task difficulty level. Our system

takes this decision as it considers one's task performance as the superordinate variable in order to avoid one's escape mechanism. We implemented the dynamic task switching by a finite state machine representation (Fig. 4).

Additionally, for the engagement-sensitive system, the system 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 III. Note that if the participant looks 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.

IV. EXPERIMENT AND METHODS

A. Participants

Eight adolescents (ASD1-ASD8) with high-functioning ASD participated in this study (Table IV). All the participants scored ≥ 80 on Peabody Picture Vocabulary Test (PPVT) [31] indicating that their language skills were adequate to participate in the current protocol. Data on core ASD related symptoms, e.g., Social Responsive Scale (SRS) [32] with a cutoff T-score ≥ 60 T, Social Communication Questionnaire (SCQ) [33] with a cutoff score of 13, Autism Diagnostic Observation Scheduled-Generic (ADOS-G) [34] with a cutoff score of 7, and Autism Diagnostic Interview-Revised (ADI-R) [35] with a cutoff score of 22 were used to select participants. All research procedures

TABLE IV
PARTICIPANT CHARACTERISTICS

	Age (y)	PPVT	SRS T-score	SCQ	ADOS-G (cutoff = 7)	ADI-R (cutoff = 22)
ASD1	17.58	134	80	12	13	49
ASD2	16.92	110	73	13	7	33
ASD3	14.25	130	89	16	15	34
ASD4	13.83	170	92	14	13	53
ASD5	16.50	92	87	20	-	-
ASD6	18.25	97	63	17	9	49
ASD7	13.00	133	90	10	7	25
ASD8	18.25	97	63	17	9	49

were approved by the Vanderbilt University Institutional Review Board.

B. Experimental Setup

The experiment was created using the VR design package discussed earlier (Section III-A). The participants' eye movements were tracked by the eye frame eye-tracker (Section III-B). Stimuli were presented on a 17-in task computer monitor. Uniform room illumination was maintained throughout the study. The task computer presented the VR-based social tasks in the foreground and computed dynamic gaze information in the background using the eye-tracking data. Gaze data along with task-related event markers (e.g., trial start and trial stop, participant's response, etc.) were logged in a time-synchronized manner. A therapist watched the participant from a video camera view, whose signal was routed to a television, hidden from the participant's view. Signal from the task computer was routed to a separate monitor for the therapist to view how the task progressed. Based on these two observations, the therapist rated the participant's engagement level.

C. Tasks and Procedures

We designed a usability study to investigate the implications of the designed VR-based PS and ES with the adaptive response technology. The commitment required of interested participants was a total of two sessions (sessions 1 and 2) on two different days lasting for approximately 2.5 h in total. First, a brief adaptation phase was carried out. In the first phase of adaptation, the experimenter briefed the participants and their caregivers about the experiment, showed the experimental setup, and the eye-tracker goggles. Then they were asked to sign the assent and the consent forms. This phase took 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 experimenter told him that he could choose to withdraw anytime from the experiments for any reason, especially if he was not comfortable interacting with the system. The participant was also introduced to a visual schedule to provide him with a brief overview of the steps involved in the study, e.g., eye-tracker calibration, followed by his role as an audience to the virtual peer's (i.e., avatar's) task presentation followed by his interaction with the avatar in the form of social conversation. The participant was then asked to rest for 3 minutes to acclimate himself with 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. The participants viewed an initial instruction screen followed by an interaction with the avatar narrating a personal story. Each storytelling task trial was approximately 1.5 min long. The participants were asked to imagine that the avatars were his 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. They were also asked to try and make their classmate (i.e., the avatar) feel as comfortable as possible during the presentation. The first three VR-based social communication task trials (one of each difficulty level) helped in selecting the baseline for each participant. Also, when the avatar responded to the participant's choice of menu option, the list of menu choices displayed on the screen disappeared to remove any confounding effects. At the end of each social communication task trial, the therapist rated the participant as to what she thought the level (using a 1–9 scale, with 1—not at all, and 9—very much) of engagement was for the participant during the recently completed task trial.

Each subject participated in both sessions, one with a PS and the other with the ES. Among the eight adolescents who participated in the study, four participants (ASD1-ASD4) first interacted with the PS-followed by-ES (henceforth referred to as Group1). The other group (henceforth referred to as Group2) of four participants (ASD5-ASD8) was exposed first to the ES-followed by-PS. This was done to determine whether there existed any ordering effect [36] due to the order of presentation of VR-based social tasks of PS and ES.

D. Rationale Behind Quantitative Estimation of Performance Score of Participants During VR-Based Social Communication Tasks

The VR-based social communication tasks used in our study were categorized in three levels of difficulty ("Easy," "Medium," and "High") (Section III-A2) requiring participants to ask three, five, and seven questions, respectively, in a particular sequence in order to retrieve the intended information from the avatar during the VR-based social conversation. We assigned six points for each correct choice and a penalty of three points for each incorrect choice. Thus the maximum points achievable corresponding to "Easy (Type 1)," "Medium (Type 2)," and "High (Type 3)" difficulty levels were 18, 30, and 42, respectively. Also, if after a certain number of attempts the participant was still unable to obtain the right information, the system would terminate the task and provide a task of lesser challenge to the participant. If a participant scored $\geq 70\%$ of the maximum score possible in a task, then the performance was considered as "Adequate," otherwise, this was considered as "Inadequate." A participant was allowed to make up to two, three, and four irrelevant choices in each step for the "Type 1," "Type 2," and "Type 3" tasks, respectively, beyond which the strategy generator switched to tasks of a lower difficulty level.

In order to quantitatively estimate the performance of the participants, each of whom could potentially participate in different VR-based social communication tasks (of varying numbers of trials and difficulty levels), we needed to compute normalized values (similar to other studies [37], [38]) of the performance

scores achieved by the participants. 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 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 is

$$
\text{PERF. SCORE (Normalized)} = \\
\frac{\left(\frac{x}{x+y+z}X\mathrm{AVG}\right) + \left(\frac{y}{x+y+z}Y\mathrm{AVG}\right) + \left(\frac{z}{x+y+z}Z\mathrm{AVG}\right)}{\left(\frac{x}{x+y+z}X\mathrm{Max}\right) + \left(\frac{y}{x+y+z}Y\mathrm{Max}\right) + \left(\frac{z}{x+y+z}Z\mathrm{Max}\right)}. \quad (1)
$$

If we consider $x = 1, y = 2$, and $z = 3$, then (1) becomes

$$
\text{PERF. SCORE (Normalized)} \\
= \frac{\left(\frac{1}{6}X\mathrm{Avg}\right) + \left(\frac{2}{6}Y\mathrm{Avg}\right) + \left(\frac{3}{6}Z\mathrm{Avg}\right)}{\left(\frac{1}{6}X\mathrm{Max}\right) + \left(\frac{2}{6}Y\mathrm{Max}\right) + \left(\frac{3}{6}Z\mathrm{Max}\right)}. \quad (2)
$$

Thus, if one achieves maximum possible scores in tasks of each level of difficulty, then his normalized performance score is 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.

V. RESULTS AND DISCUSSION

A. System Acceptability

In the current study, we wanted to investigate whether our VR-based interactive system with adaptive response technology was acceptable to our participants with ASD. In spite of being given the option of withdrawing from the experiment at any time during their interaction with the system, all the participants completed the sessions. An exit survey at the end of the experiment revealed that all the participants liked interacting with the system particularly while using the bi-directional conversation module, had no problems in wearing the eye-tracker goggles, and understanding the stories narrated by their virtual classmates. When asked about any take-home lesson that they had from the conversation between them and their virtual classmates, most of them (six out of eight) said that they learned that they should introduce themselves first while speaking to a new friend for the first time and that they should look towards the faces of their friends during conversation. Thus, it is reasonable to infer from this study that our system has a potential to be accepted by the target population.

B. Quantitative Analysis of Performance of Participants While Interacting With Trials of PS and ES

The normalized performance scores achieved by the participants were examined. Results (Fig. 5) indicate an improvement in the performance score for all the participants (except ASD8)
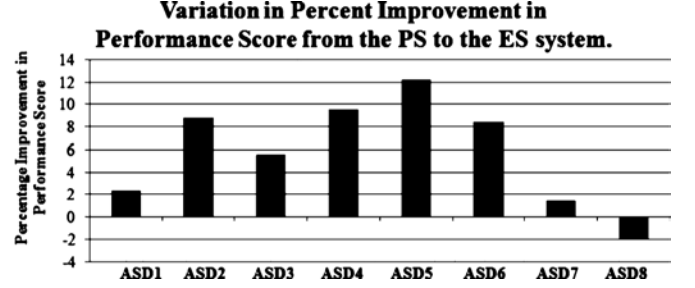

Fig. 5. Comparative analysis of the percentage improvement in the performance score of participants from the PS to the ES system.

from PS to ES with the range of improvement being approximately 1.5%–12.16%.

Additionally, we carried out statistical analysis to determine whether the improvement in the normalized performance scores of the participants while interacting with the PS and ES were statistically significant. A dependent sample T-test on the participants' normalized performance score while interacting with the PS and the ES indicates that they were statistically significantly different ($p = 0.0106$) with an effect size (Cohen's d) of 0.4614. To determine whether our choice of the scoring rationale (i.e., a score of six points for a relevant choice and a penalty of three points for an irrelevant choice) influenced the statistical significance of the performance scores, we performed offline analysis of the performance scores of the participants using different scoring rationale. For example, if the present system used a performance score of 10 points for a relevant choice and a penalty factor of two points for an irrelevant choice for each of the three difficulty levels, then comparison of the performance scores between the PS and the ES systems would indicate that they were statistically significantly different ($p = 0.0122$). In addition, if the penalty factors were three, or four, or five points, for each irrelevant choice and a score of 10 points for each relevant choice, the performance scores of the participants while interacting with the PS and the ES systems would also be statistically significant ($p = 0.0116$; $p = 0.0102$; and $p = 0.0152$, respectively). Also, to determine whether our choice of weight factors e.g., $x = 1, y = 2$, and $z = 3$ (Section IV-D) contributed to the participants' normalized performance scores being statistically significantly different, we performed offline analysis on the performance scores with different weight factors. Thus, considering the scoring rationale of six points for a relevant choice and three points for an irrelevant choice along with weight factors of $x = 1.1$, $y = 1.2$, and $z = 1.3$, we find that the performance scores remain statistically significantly different ($p = 0.0098$). This analysis indicates that the performance improvement in ES as compared to PS was not due to the scoring scheme.

Furthermore, we performed a statistical analysis to determine whether there were any ordering effects due to the order of presentation of PS and ES. An independent sample T-test on the normalized performance scores across Group1 and Group2 for each of PS and ES indicates that they were not statistically significantly different ($p = 0.8550$ and effect size of 0.1356 for PS and $p = 0.7606$ and effect size of 0.1115 for ES). Thus, we can say that there were no significant ordering effects due to the
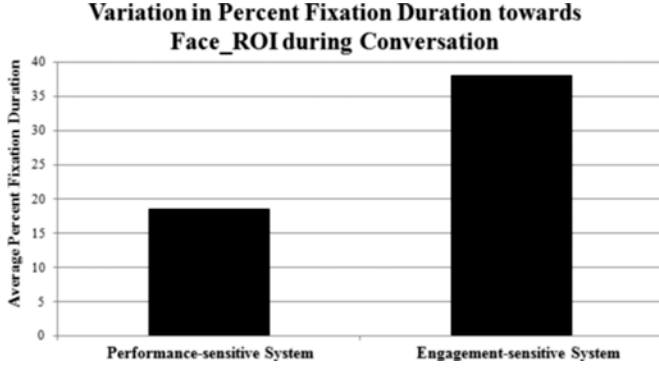
Fig. 6. Variation in percent FD while looking towards the face region (Face_ROI) of the avatars during VR-based social conversation (PS-left bar, ES-right bar).
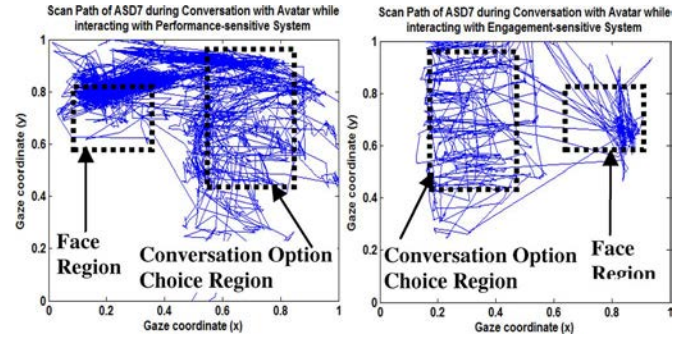

Fig. 7. Variation in scan path of ASD7 while conversing with the avatar during interaction with performance-sensitive system and engagement-sensitive system.
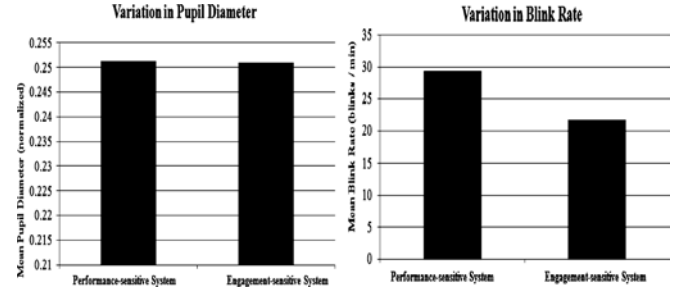

Fig. 8. Comparative analysis of the eye physiological indexes (a) PD and (b) BR of participants while interacting with performance-sensitive system and the engagement-sensitive system.

order of presentation of VR-based social tasks of the PS and the ES systems.

C. Variation in Behavioral Viewing Pattern of Participants While Interacting With PS and ES

We investigated the impact of the developed systems on the viewing patterns, since besides acquiring improved task performance, one needs to be able to carry on conversation in a socially appropriate way (e.g., looking towards the face of the communicator). We found an improvement in the viewing pattern in terms of increased FD of all the participants while looking towards the face region of the avatar during interaction with the ES compared to that with the PS. A dependent sample T-test between the percent FD while looking towards the face of the avatars while interacting with PS and ES indicate that the improvement in the viewing pattern for the group was statistically significantly different ($p = 0.002$) with effect size being 0.4824. Also note that although the average percentage fixation duration of the participants while looking towards the face (Face_ROI) of the avatar during social conversation was below 50% (as can be seen from Fig. 6), there were multiple trials when some of the participants looked at the Face_ROI greater than 50% of the time.

We also studied the scan paths of the participants as individuals with ASD have been shown to exhibit atypical scan paths during social interaction [3]. Here we present the scan path of one of the participants (e.g., ASD7) during the last trial while interacting with the PS and with the ES (Fig. 7), both representing the "High" interaction difficulty level. Fig. 7 indicates that ASD7 fixated on different regions of the visual stimulus while interacting with PS. However, while interacting with ES, ASD7 fixated mainly on the face region of the avatar and the menu option choices of the bi-directional conversation module while conversing with the avatar. Note that, these scan paths were analyzed in the background and they were not visible to the participant.

D. Variation in Eye Physiological Indexes (e.g., BR and PD) of Participants While Interacting With PS and ES

In our present work we used the two eye physiological indexes, namely, PD and BR as indicators of one's engagement to a task. This is because individuals with ASD have been shown

to demonstrate reduction in PD [28] and BR [27] while being engaged to a task. In our present work, we performed a comparative analysis of the participants' eye physiological indexes while they interacted with the PS and the ES. Although the PS system switched tasks based on the performance only, we collected the eye physiological signals (PD and BR) for later offline analysis. We analyzed our data to determine whether the variation in the eye physiological indexes of the participants while interacting with our PS and ES systems was indicative of the participants' improved engagement level while interacting with the ES than that with the PS.

From Fig. 8(a), we find that the mean PD (normalized on a 0-1 scale with respect to the eye camera window of the eye-tracker that we used) of the participants almost remained the same—it decreased marginally with no statistical significance between the PS and the ES. The BR changed more between the PS and ES [Fig. 8(b)]—it decreased from approximately 29 bpm to 22 bpm. However, this change was also not statistically significant with an effect size of 0.15. In the present usability study, with a limited sample size, we find that the BR of the participants is more sensitive to their engagement while interacting with the VR-based social situations than the PD.

VI. CONCLUSION

In the present work we developed a VR-based engagement-sensitive system with an adaptive response technology for intervention of individuals with ASD. We presented the system development and results from a usability study to test the efficacy of the developed system as a first step to technology-assisted intervention. The developed system can communicate with the

participants and detect subtle variations in one's eye-physiological features, and viewing pattern in real-time. Also, it seamlessly integrates this information with the VR-platform to provide intelligent individualized adaptive response.

Results from a usability study show the capability of the system to contribute to improving one's social task performance (e.g., quantitative improvement of performance metric) along with encouraging socially-appropriate mechanisms (e.g., increased viewing of the face of the communicator) to foster effective social communication skills among the participants with ASD. However, there are several limitations of the current study that warrant consideration. Certainly a much larger study of the current system would be needed to understand how our current findings impact areas of core deficit for individuals with ASD and how such impact can be generalized to the heterogeneous population of individuals with the disorder. There also remain many questions regarding the specific impact of this system on learning over time and what specific mechanism may need adjustment in order to optimize efficient and relevant task adjustment. Our findings certainly provide initial support for the potential of our intelligent adaptive system to enhance learning via real-time measurement of gaze behavior and verbal communication exchanges with an avatar in a controlled VR environment. However, a fundamental challenge to this system is that realistic social interactions, a likely target for this intervention tool, do not exist within controlled environments. Fluid, meaningful, and highly relevant social interactions are quite challenging to reproduce within virtual environments. A technical limitation of this system is that this study used a wearable eye-tracker and the bi-directional conversational module required use of a mouse. These may not be suitable for individuals with low functioning ASD. In the future we plan to incorporate a remote desktop mounted eye-tracker and a speech recognizer module with built-in natural language understanding facility for the bi-directional conversation module to mitigate these problems. In our present study we have designed twenty four VR-based social conversation modules. In the future we plan to develop an inventory of such VR-based social conversation modules so that the participants can interact with the system for a prolonged period.

The current study as presented here shows that such an integrated system has a potential to be incorporated into complex intervention paradigms. We hope that in the future such an integrated system can pave the way for developing a powerful complementary tool in the hands of the interventionist so as to contribute to our endeavors in improving quality of life for individuals with ASD.

Acknowledgment

References

[1] APA, American Psychiatric Association, Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC, 2000.

[2] M. Carpenter, B. F. Pennington, and S. J. Rogers, "Interrelations among social-cognitive skills in young children with autism," *J. Autism Develop. Disorders*, vol. 32, no. 2, pp. 91–106, 2002.

[3] 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. Develop. Disorders*, vol. 38, pp. 1371–1381, 2008.

[4] 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.

[5] C. Trepagnier, M. M. Sebrechts, and R. Peterson, "Atypical face gaze in autism," *Cyberpsychol. Behav.*, vol. 5, no. 3, pp. 213–217, 2002.

[6] U. Lahiri, Z. Warren, and N. Sarkar, "Design of a gaze-sensitive virtual social interactive system for children with autism," *Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 4, pp. 443–452, 2011.

[7] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Arch. Gen. Psychiatry*, vol. 59, no. 9, pp. 809–816, 2002.

[8] 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," *Cyberpsychol. Behav.*, vol. 9, no. 2, pp. 213–217, 2006.

[9] S. Parsons, P. Mitchell, and A. Leonard, "The use and understanding of virtual environments by adolescents with autistic spectrum disorders," *J. Autism Develop. Disorders*, vol. 34, no. 4, pp. 449–466, 2004.

[10] A. Tartaro and J. Cassell, "Using virtual peer technology as an intervention for children with autism," in *Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations*, J. Lazar, Ed. New York: Wiley, 2007.

[11] D. Strickland, "Virtual reality for the treatment of autism," in *Virtual Reality in Neuropsychophysiology*, G. Riva, Ed. Lansdale, PA: IOS Press, 1997, pp. 81–86.

[12] A. Tartaro and J. Cassell, "Playing with virtual peers: Bootstrapping contingent discourse in children with autism," in *Int. Conf. Learn. Sci.*, Utrecht, The Netherlands, 2008.

[13] *Educating Children With Autism*. Washington, DC: Nat. Acad. Press, 2001.

[14] L. A. Ruble and D. M. Robson, "Individual and environmental determinants of engagement in autism," *J. Autism Develop. Disorders* , vol. 37, no. 8, pp. 1457–1468, 2006.

[15] 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 Affective Neurosci.*, vol. 5, no. 1, pp. 98–107, 2010.

[16] D. Palomba, M. Sarlo, A. Angrilli, A. Mini, and L. Stegagno, "Cardiac responses associated with affective processing of unpleasant film stimuli," *Int. J. Psychophysiol.*, vol. 36, pp. 45–57, 2000.

[17] S. V. G. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson, "Virtual reality induced symptoms and effects," *Presence*, vol. 8, no. 2, pp. 169–186, 1999.

[18] WorldvizLlc. [Online]. Available: http://www.worldviz.com/

[19] AllFreeEssays [Online]. Available: http://www.allfreeessays.com/

[20] K. Welch, U. Lahiri, N. Sarkar, and Z. Warren, "An approach to the design of socially acceptable robots for children with autism spectrum disorders," *Int. J. Social Robot.*, vol. 2, no. 4, pp. 391–403, 2010.

[21] A. F. Pereira, C. Yu, L. B. Smith, and H. Shen, "A first-person perspective on a parent-child social interaction during object play," presented at the 31st Annu. Meeting Cog. Sci. Soc., Amsterdam, The Netherlands, 2009.

[22] C. Psaltis and G. Duveen, "Conservation and conversation types: Forms of recognition and cognitive development," *Br. J. Develop. Psychol.*, vol. 25, pp. 79–102, 2007.

[23] M. J. Roberts, Choosing a conversation system [Online]. Available: http://www.tads.org/howto/convbkg.htm, 2001

[24] J. Fisher, Advanced NPC implementation [Online]. Available: http://www.onyxring.com/InformGuide.aspx?article =20 2004

[25] Arrington Research Inc. [Online]. Available: http://www.arringtonresearch.com/

[26] T. Booth, *Sequential Machines and Automata Theory*. New York: Wiley, 1967.

[27] C. J. Anderson, J. Colombo, and D. J. Shaddy, "Visual scanning and pupillary responses in young children with autism spectrum disorder," *J. Clin. Exp. Neuropsy.*, vol. 28, pp. 1238–1256, 2006.

[28] B. Jensen, B. Keehn, L. Brenner, S. P. Marshall, A. J. Lincoln, and R. A. Müller, "Increased eye-blink rate in autism spectrum disorder may reflect dopaminergic abnormalities," *Int. Soc. Autism Res.*, 2009.

[29] M. Argyle and M. Cook, *Gaze and Mutual Gaze*. Cambridge, MA: Cambridge Univ. Press, 1976.

[30] A. Colburn, S. Drucker, and M. Cohen, "The role of eye-gaze in avatar-mediated conversational interfaces," in *SIGGRAPH Sketches Appl.*, New Orleans, LA, 2000.

[31] L. M. Dunn and L. M. Dunn, *PPVT-III: Peabody Picture Vocabulary Test*, 3rd ed. Circle Pines, MN: Am. Guidance Service, 1997.

[32] J. N. Constantino, *The Social Responsiveness Scale*. , CA: Western Psych. Serv., 2002.

[33] M. Rutter, A. Bailey, S. Berument, C. Lord, and A. Pickles, *Social Communication Questionnaire*. Los Angeles, CA: Western Psychol. Services.

[34] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, Jr., 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," *J. Autism Develop. Disord.*, vol. 30, no. 3, pp. 205–223, 2000.

[35] M. Rutter, A. Le Couteur, and C. Lord, *Autism Diagnostic Interview Revised WPS Edition Manual*. Los Angeles, CA: Western Psychol. Services, 2003b.

[36] G. W. Heiman, *Research Methods in Psychology*, 3rd ed. New York: Houghton Mifflin, 2002.

[37] S. C.-R. Javier, "Weighted average score of customer needs as critical input for QFD," *QFD Int. J.*, 2007.

[38] S. Hirsch, T. L. Frank, J. L. Shapiro, M. L. Hazell, and P. I. Frank, "Development of a questionnaire weighted scoring system to target diagnostic examinations for asthma in adults: A modelling study," *BMC Family Pract.*, vol. 5, pp. 30–38, 2004.

Esubalew Bekele (S'07) received the M.S. degree in electrical engineering and computer science, in 2009, from Vanderbilt University, Nashville, TN, where he is currently working toward the Ph.D. degree in electrical engineering and computer science.

He served as junior faculty in Mekelle University, Ethiopia, before joining Vanderbilt University. His current research interests include human–computer interaction, robotics, affect recognition, machine learning, and computer vision.

Elizabeth Dohrmann received the bachelor's degree with distinction in psychology at Yale University, New Haven, CT. Following her initial training at the Yale Child Study Center, she continued research in autism spectrum disorders at the University of California, San Diego and Vanderbilt University. She is currently a second year medical student at the University of Tennessee Health Science Center, Memphis.

Zachary Warren received the Ph.D. degree from the University of Miami, Miami, FL, in 2005. He completed his pre-doctoral internship at Children's Hospital Boston/Harvard Medical School in and his postdoctoral fellowship at the Medical University of South Carolina.

Currently he is an Associate Professor of Pediatrics and Psychiatry at Vanderbilt University, Nashville, TN. He is the Director of the Treatment and Research Institute for Autism Spectrum Disorders at the Vanderbilt Kennedy Center as well Director of Autism Clinical Services within the Division of Genetics and Developmental Pediatrics.

Uttama Lahiri (S'11–M'12) received the Ph.D. degree from Vanderbilt University, Nashville, TN, in 2011.

Currently, she is an Assistant Professor at Indian Institute of Technology, Gandhinagar. India. Her research interests include virtual reality based human–computer interaction, eye tracking and physiology based modeling techniques, human robot interaction, adaptive intelligent techniques in cognitive research and robot assisted surgical techniques.

Nilanjan Sarkar (S'92–M'93–SM'04) received the Ph.D. degree in mechanical engineering and applied mechanics from the University of Pennsylvania, Philadelphia, in 1993. He was a Postdoctoral Fellow at Queen's University, Canada.

Then, he joined the University of Hawaii as an Assistant Professor. In 2000, he joined Vanderbilt University, Nashville, TN, where he is currently a Professor of mechanical engineering and computer engineering. His current research interests include human–robot interaction, affective computing, dynamics, and control.

Dr. Sarkar was an Associate Editor for the IEEE TRANSACTIONS ON ROBOTICS.