VIGART APPLICATION TO ASSIST ASD CHILDREN

Miss A W. Chorey, Dr. A. D. Gawande

Computer Science of Engineering Sipna College of Engineering Amravati, India.

Abstract— Children with Autism Spectrum Disorder (ASD) are often characterized by inadequacy in social communication skills. In recent time, many assistive technologies, particularly Virtual Reality, have been investigated to promote social. It is well known that children with ASD demonstrate atypical viewing patterns during social interactions and thus monitoring eye-gaze can be valuable to design intervention strategies. While several researchers have used eye-monitoring technology to track eye-gaze for analysis, there exists no real time system that can monitor eye-gaze dynamically and provide customized feedback. Researchers are increasingly employing technology to develop more feasible, accountable, and individualized intervention tools to address core vulnerabilities related to autism. The study states the development and preliminary application of a Virtual Reality (VR) technology aimed at facilitating improvements in social communication skills for adolescents with autism. Virtual Interactive system with Gaze-sensitive Adaptive Response Technology (VIGART) is capable of delivering individualized feedback based on a child's dynamic gaze patterns during VR-based interaction. I present preliminary data from the usability study of this technological application for six adolescents with autism and discuss potential future development and application of adaptive Virtual Reality technology within an intervention framework.

Index Terms: Autism Spectrum Disorder; Virtual Reality; Eye-tracking; Fixation duration

INTRODUCTION

Autism is a developmental disorder that appears in the first 3 years of life, and affects the brain's normal development of social and communication skills [1]. While a specific cause or specific causes of autism spectrum disorders has yet to be found, many risk factors have been identified in the research literature that may contribute to the development of an ASD. These risk factors include genetics, prenatal and perinatal factors, neuroanatomical abnormalities, and environmental factors. To understand the social communication vulnerabilities of individuals with Autism Spectrum Disorder (ASD), research has examined how they process salient social cues, specifically from faces [2].

The ability to originate socially relevant information from faces is a primary skill for facilitating mutual social interactions. However, person with ASD demonstrate poor eye contact during social interaction in which they tend to fixate less towards human faces and more towards other objects within visual stimulus. There is increasing unanimity that intensive behavioral intervention programs can significantly improve long term outcomes for these individuals.

Professional resources, researchers are developing and deploying technology e.g., computer technology, virtual reality environments, and automated systems like robotic systems to develop more accessible individualized intervention. In this, I studied a novel virtual reality (VR)-based system for social interaction to be used with to grow up with ASD. I chose Virtual

reality (VR) technology possesses several strengths in terms of potential application for individuals with ASD, namely, malleability, controllability, reliability, modifiable sensory stimulation, and an ability to pragmatically individualize intervention approaches and reinforcement strategies [3]. In usability study, present VR-based tasks on a computer screen as VR is often effectively experienced using standard desktop computer input devices for ASD intervention. VR can illustrate scenarios which can be changed to accommodate various situations that may not be feasible in a given therapeutic setting with space limitations and resource deficits [4]. VR can also illustrate and replicate scenarios which can be changed to accommodate various situations that may not be feasible in a given therapeutic setting [5]. Thus, VR represents a medium well-suited for creating and intensively studying the impact of interactive technological intervention paradigms for ASD.

Despite potential advantages, current VR environments as applied to assistive intervention for children with ASD are designed in an open-loop fashion. These VR systems may be able to chain learning via aspects of performance; however, they are not capable of a high degree of individualization. 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. Specifically, performance based VR-based interactions do not often involve measurements of or necessitate appropriate subtle, yet critically important, aspects of effective social communication (such as eye-gaze, and other forms of social convention). In fact, while many individuals with ASD are capable of yielding correct performance on objective task measures, it is their vulnerabilities

surrounding elements of social communication that is so closely tied to their functional social impairments.

In the current work, I emphasis on a novel Virtual Interactive system with Gaze-sensitive Adaptive Response Technology (VIGART) using VR-based social situation as a platform for delivering individualized feedback based on one's dynamic gaze patterns. Additionally, I report the results from a usability study with VIGART to investigate the impact of this system on the engagement level, behavioral viewing, and eye physiological indexes of six adolescents with ASD. In the present work, specifically, I describe the development of a system that allows dynamic eye-tracking with VR interaction, such that the VR system pin-points quantitatively, where and how long a child is looking using metrics such as, fixation duration.

I. OBJECTIVES AND SCOPE

The construct of behavioral engagement of children with ASD is the basis for many evidence based intervention paradigms focused on bolstering core skill areas [6]. Given the importance of engagement, viewing patterns and psychophysiology, developing a gaze-sensitive VR-based system for ASD that can provide individualized feedback based on a child's real-time gaze pattern and can potentially elicit changes in interaction level, behavioral viewing, and eye physiological indexes of the child can be important. This can be an approach towards achieving realistic social interaction to challenge, and scaffold skill development in particular areas of vulnerability for these children.

Instead the current work represents a first-step in demonstrating the feasibility of a potentially more complex, sophisticated, and robust intervention system designed to detect patterns of gaze, as well as other subtle and necessary components of social communication. In this manner the current work demonstrates initial technological capacity to develop more robust and sophisticated technological systems that are capable of adjusting task characteristics in order to potentially modify and enhance aspects of social communication. In addition, also investigated whether VIGART can measure behavioral viewing indexes and eye physiological indexes dynamically that have the potential to be used as objective measures of one's engagement as well as emotion recognition capability; though our usability study was not designed to improve the participants' ability to correctly recognize an emotional expression. The behavioral viewing indexes that I studied were the fixation duration and fixation frequency because of their role in predicting attention on a specific region of visual stimulus. Similarly, eye physiological indexes such as the blink rate and pupil diameter were also measured because of their sensitivity to emotion recognition.

II. DESIGN OF VIGART

The dynamic closed-loop interaction provided by VIGART has three main subsystems: First is, a VR based platform that can present social tasks; second is a real time eye-gaze monitoring system; and third is a consolidation module that establishes link in the first two systems.

A. VR-Based Task showcase

The VR experiment was created using the Vizard VR design package (from Worldviz Llc) and Arrington eye-tracker (from Arrington Research Inc). Our participants viewed the VR environment and avatars within the system from first-person perspective. Gaze data along with task-related event markers (e.g., trial start/stop, responses of participants to question asked by the system, etc.) were logged in a time-synchronized manner. However, a number of enhancements were made on the VR-platform. I studied 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.

Fig. 1. Screenshots of avatars demonstrating neutral (left side), happy (middle), and angry (right side) facial expression.

Three dimensional avatar heads were created from contemporary photographs of teenagers using the '3DmeNow' software package. These new avatar heads were used instead of pre-supplied Vizard creations, in order to create avatars: a) with age range close to our participant pool and b) with more authentic facial features (e.g., realistic brow line, nose dimensions, etc.). Facial expressions (e.g., "neutral," "happy," and "angry") (Fig. 1) are morphed by "PeopleMaker" software. 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 [7]. The first-person stories shared by avatars are adapted from dynamic indicators of the basic early literacy skills reading the assessments and include content thought to be related to potential topics of school presentations (e.g., reports on experiences, trips, etc.). Audio files are developed first by using the text-to-speech converter and then recorded using recorder. For the avatars to speak the content of the story, these audio files are lip-synched using a Vizard-based speak module. Additionally, the VR-based visual stimuli (e.g., avatar's face, objects of interest, etc.) are characterized by a set of regions of interest (ROIs). These ROIs are used by the dynamic eye-tracking algorithm I present in this seminar to keep track of the eye-gaze of the participants as they interact with the VR-based tasks.

B. Real Time Gaze Monitoring system

The VIGART captures eye data of a participant interacting with an avatar using Eye-Tracker goggles. This eye-tracker comes with some basic feature acquiring capabilities for offline analysis. In this study, I correlate the extracted eye-gaze features reflecting the behavioral viewing patterns of a participant with ASD with his/her engagement level because engagement, which is defined as "sustained attention to an activity or person," is one of the key factors for these children to make substantial gains in communication and social domains. In addition, I correlate the extracted features reflecting the behavioral viewing

patterns and eye physiological indexes, with emotion recognition capability of the participants because emotion recognition is characterized as one of the core deficits indicating ASD as well as its importance in social communication .
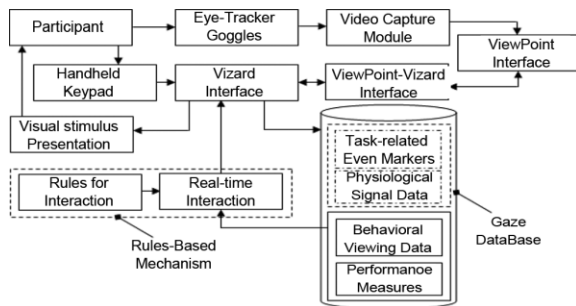

Fig. 2. Schematic of data acquisition and the control mechanism used.

1)   Data Acquisition

The Eye-Tracker that I use comes with a Video Capture Module with refresh rates of 60 Hz (low precision) and 30 Hz (high precision) to acquire a participant's gaze data using software called Viewpoint. They have designed Viewpoint-Vizard Handshake module (Fig. 2), 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 physiological signal data (pupil diameter and pupil aspect ratio), behavioral viewing data (fixation duration and 2D gaze coordinates), and performance measures (e.g., a participant's responses to questions asked by this system) with a refresh rate of 30 Hz in a time-synchronized manner. Signal processing techniques such as windowing and threshold are used to filter these data to eliminate noise and subsequently extract the relevant features.
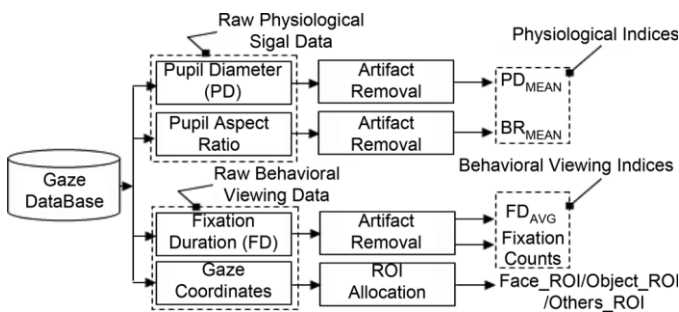
2)   Feature Extraction

The Gaze Database (Fig. 2) is processed to extract four features, which are mean pupil diameter, mean blink rate, sum of fixation counts (SFC), and average fixation duration, for each ROI from each segment of the signals monitored (Fig. 3).

Computation of Mean Pupil Diameter:
The raw normalized pupil diameter is recorded by the Viewpoint software in terms of normalized value (0-1) with respect to the Eye Camera window of the eye-tracker. Literature

indicates the importance of measuring actual pupil diameter of typically developing (TD) and ASD participants. For example, Anderson et al. report larger tonic actual pupil size in children with ASD than their TD counterparts. In the present work, in order to extract the actual pupil diameter at each instant, use the recorded data on pupil aspect ratio (i.e., the ratio of the major and minor axes of the pupil image) defining the eye image (with 1 indicating a perfect circle) to compute optimal pupil diameter. Artifact removal incorporates elimination of the discontinuities in the raw pupil diameter data due to blinking effects and other minor artifacts. Then using the actual Eye- Camera window dimensions, the true pupil diameter (in millimeters) is computed. Using the record of ROIs visited by the eye at each instant, the mean pupil diameter corresponding to each ROI is computed.

Computation of Mean Blink Rate:
The blink rate is determined using the pupil aspect ratio data recorded by the Viewpoint software as mentioned by Arrington. In the present work, I computed the blink rate by considering the number of times the pupil aspect ratio value falls below the lower threshold of 0.5 for a 1 min window, which allowed us to detect blink rate with an accuracy of subsequently, the mean blink rate corresponding to each ROI is computed.

Computation of Fixation Counts and Average Fixation Duration:
The recorded data on fixation duration corresponding to each ROI is first filtered to remove the artifacts due to blinking and noise spikes are eliminated by thresholding. Literature indicates that fixations typically last between 200–600 ms. In the present study, I compute the fixation duration by using a thresholding window of 200 ms as the lower limit to eliminate the blinking effects and 450 ms as the upper limit (i.e., up to 1.5 standard deviations from the lower threshold). These limits provided a reliable data range restricted by noise due to glare effects of cameras of the eye-tracker that I used. Subsequently, the sum of fixation counts (SFC) and average fixation duration are computed for each ROI.

Detection of ROIs Viewed:
Prior to study, I carried out a systematic validation of the overall system performance in terms of accuracy and repeatability of the eye-gaze data with lab members. In this validation study, participants were asked to fixate on each of 15 predefined target points appearing randomly on the visual stimulus each for 5 s. The viewing area was normalized on a 0–1 scale, where the left-hand bottom corner and the right-hand top corner of the screen were assigned the (0, 0) and (1, 1) coordinates, respectively. The average mean square errors for the and gaze coordinates for the 15 predefined points were found to be 0.0053 and 0.0077 (i.e., eight pixels along and seven pixels along for the entire presented visual stimulus), respectively.
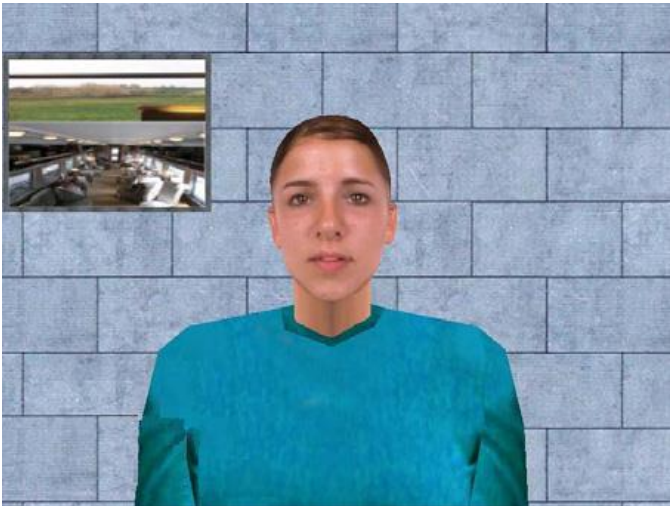
Fig 2 Allocation of ROIs (Face_ROI, Object_ROI, and Others_ROI).

Afterwards, I designed a computational algorithm, the real-time gaze-based feedback algorithm (RGFA) that determines whether the gaze coordinates correspond to our task-specific segmented regions of the visual stimulus presented to participants. Subsequently, RGFA assigns numeric tags (e.g., 1, 2, etc.) for each ROI. In the present study, I segmented the VR-based visual stimulus into three 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 outdoor games, the context-relevant object is a picture displaying collage of snapshots of narrated games). The task-related event markers along with the ROI tags are then used by Real-time Gaze-based Feedback Algorithm to compute the physiological and behavioral indexes during viewing of different ROIs by a participant.

C. Communication link between VR-Based Task show case and Real Time Gaze Monitoring Modules

1) Rationale behind Gaze-Based Individualized Feedback Mechanism

The VIGART 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. In dyadic communication, gaze information underlying one's expressive behavior plays a vital role in regulating conversation flow. Looking behavior is a function of the cultural upbringing of an individual. Specifically, some cultures may give a negative connotation to looking directly at others' eyes, while others may consider this as appropriate. Our presented system has the flexibility of changing the gaze parameter to suit different cultural requirements. In our present work I use gaze definition suitable for western culture with a listener looking at the speaker 70% of the time during an interaction identified as "normal while listening" [8]. In the present study, a participant can serve as an audience while the avatars narrate personal stories and display context relevant facial expressions (Fig. 1) to capture the mood inherent in the content of the story. Subsequently, the participant's fixation duration for Face_ROI viewing as a

percentage of total fixation duration, extracted from the behavioral viewing data and the performance measure (the participant's response to question asked by VIGART) initiate a rule-based mechanism (Fig. 2) to trigger the VIGART to provide feedback to the participant using the individualized real-time gaze-based feedback algorithm (Section IV). The VIGART 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 VIGART. A feedback is generated based on whether the participant could recognize the avatar's emotion and how much he/she looked at the avatar's face.

2) Overview of the Individualized Real-Time Gaze-Based Feedback Algorithm (RGFA)

The data flow diagram for the RGFA (Fig. 5) presents a brief overview of the logic used by VIGART. Real-time gaze coordinates of a participant interacting with an avatar are acquired using the Viewpoint software and converted to VR (Vizard) compatible format using a Viewpoint-Vizard handshake module (Fig. 2). A computer, where the VR-based tasks are presented, runs Viewpoint Software at the background and Vizard software at the foreground. The RGFA triggers a 33 ms timer to acquire the gaze coordinates. Based on the participant's 2D gaze-coordinates, the RGFA then computes the specific ROI looked at by the participant. Times spent by the participant looking at different ROIs are stored in the respective buffers that are added up at each instant during participant-avatar interaction. This determines the participant's looking time towards the presented stimuli in terms Face_ROI, Object_ROI, of and Other_ROI 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 from which the percentage of time spent by a participant in looking at Face_ROI is computed. In this specific usability study, four different responses (S1-S4) (Table I) are generated based on how much a participant looked at the face of the avatar and whether he/she correctly recognized the feeling of the avatar.

TABLE I: Rationale behind Attention-based Real-time Motivational Feedback.

Response to Q1	t>=70%	VIGART Response [Label]
Right	Yes	Your classmate really enjoyed having you un the audience. You have paid attention to her and made her feel comfortable. Keep it up![S1]
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. [S2]
Wrong	Yes	Your classmate felt comfortable in having you in the audience. However you may try.[S3]
Wrong	No	Your classmate would have felt more comfortable if you had paid more attention to her.[S4]

3) Real-Time Interaction Module

The gaze metrics as obtained from the Real-time Gaze-based Feedback Algorithm and any predefined interaction rules can be combined using the real time interaction module (Fig. 2) to provide individualized feedback in an attempt to improve the participant's engagement to the social tasks. While not utilized in the usability study presented in this paper, the system has the capability of communicating with the VR-based platform using our Viewpoint-Vizard interface to modify the avatar's states to create an interactive interaction. I have augmented the avatar's capabilities in several ways. 1) The avatar can change his/her facial expressions during communication; e.g., the avatar may start with a neutral face and then based on the content of the story gradually change to happy/angry facial expression. 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. 2) The avatar can walk towards, point to, and turn and look at an object of interest 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 I have developed and modified in vizard. 3) The avatar can also provide verbal feedback during the trial interrupting its original conversation/storytelling by initiating a separate conversation thread within vizard, if the protocol demands such a feedback. These capabilities can be accessed at every 33 ms if required. However, in the present usability study I 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.

III. DESIGN OF THE USABILITY STUDY

A. *Experimental Set-Up*

Experimental setup included a 17" task computer monitor that was customized to present VR-based social communication tasks in the foreground and compute dynamic gaze information in the background. The VR experiment was created using the Vizard VR design package (from Worldviz Llc) and Arrington eye-tracker (from Arrington Research Inc). Our participants viewed the VR environment and avatars within the system from first-person perspective. Gaze data along with task-related event markers (e.g., trial start/stop, responses of participants to question asked by the system, etc.) were logged in a time-synchronized manner. Uniform room illumination was maintained throughout the experiment and sessions were video-recorded to afford the potential to cross-reference experimental observation.

Participant (Gender)	Age (Years)	PPVT	SRS (Cutoff=60)	SCQ (Cutoff=13)	ADOS-G
ASD 1(m)	13.83	126	69	23	11
ASD2(m)	15.5	110	73	13	7
ASD3(f)	15.17	83	90	28	10
ASD4(m)	16.5	97	63	17	9
ASD5(m)	15.08	92	87	20	NA
ASD6(m)	17.5	103	83	31	20
Means(SD)	15.60(1.27)	102(15)	78(11)	22(7)	11(5)

B. *Participant Recruitment and Inclusion Criteria*

Six adolescents (Male: n =5 , Female: n=1) with high functioning ASD, ages 13–17 y(M=15.60y, SD=1.27y) participated in the usability study of the designed system. 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 on Peabody Picture Vocabulary Test-3rd Edition (PPVT-III) 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) profile sheet and Social Communication Questionnaire (SCQ) with all participants falling above the clinical thresholds (Table II). The SRS (cutoff T-score) generates a total score reflecting the severity of social deficits in the autism spectrum. From the SRS scores, I find that ASD1, ASD2, and ASD4 can be considered to be in mild to moderate ASD range, and ASD3, ASD5, and ASD6 were in the severe range of ASD (Table II). The SCQ (cutoff of 13) is a parent report questionnaire aimed at evaluating critical autism diagnostic domains of qualitative impairments. Autism Diagnostic Observation Schedule-Generic (ADOS-G) scores (cutoff of 7) were available for five of the six participants from prior evaluations (Table II).

C. *Tasks and Procedures*

In the present usability study, I constructed five VR-based social interaction scenarios (trial1-trial5) in which the avatars (and henceforth referred to as the "virtual classmates" of the participants) narrated personal stories on diverse topics such as, outdoor sports, travel, favorite food, etc. Each participant took part 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, and showed the experimental setup. 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 chair was adjusted so that his/her eyes were collinear with the center of the task computer, C1 (Fig. 6). The experimenter told him/her that he/she could choose to withdraw anytime from the experiments for any reason, especially if he/she was not comfortable interacting with VIGART. The participant was then asked to rest for 3 min to acclimate him/her to the experimental setup. 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 their virtual classmate 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 [Fig. 1(a)] and narrating a personal story. This trial was followed by four other trials that were similar to the trial1 except that in these subsequent trials the virtual classmate displayed "happy" [Fig. 1(b)] or "angry" [Fig. 1(c)] facial expressions to capture the mood inherent in the content of the story. In the present study I used three female, two male avatar heads with one neutral, three happy, and one angry facial expression. I randomized the trial2-trial5 among the participants and the avatars displayed context-relevant facial expressions. 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 with three answer choices (A: Happy; B: Angry; C: Not Sure). The Q2 was about some basic facts as narrated in the story. It also had three answer choices. The participant responded with a keypad. Q2was asked to encourage a participant to pay attention to the story content. 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, VIGART encouraged the participant to either pay more or keep the same attention towards the presenter (Table I). After each trial, the observer rated about

what he/she thought about how engaged the participant was during the trial using a 1–9 scale (1: least engaged; 9: most engaged). Each participant was compensated with a gift card for completing a session.

IV. RESULTS

Here I present the results of our usability study with six adolescents with ASD to examine the acceptability of VIGART by the target population, and investigate the potential of VIGART to elicit variation in participant's engagement level. Subsequently, I also analyze whether VIGART can be used to define objective measures of engagement and emotion recognition capabilities so that such objective measures can be utilized in future intervention strategies using VIGART. All the results reported here are descriptive only. More definitive results supported by inferential statistics would require a larger sample size and comparison to a control group.

A. System Acceptability

1) In the current usability study, I wanted to investigate whether VIGART was acceptable to adolescents with ASD. I tested our system with a small sample of six adolescents with ASD. In spite of being given the option of withdrawing from the experiment at any time during their interaction with VIGART, all the participants completed the session. An exit survey revealed that all participants liked interacting with VIGART, and had no problem in wearing the eye-tracker goggles, understanding the stories narrated by their virtual classmates, or responding to questions asked by VIGART. In fact five of them inquired whether there would be any future participation possibilities with VIGART. Thus it is reasonable to infer from this small usability study that the VIGART has a potential to be accepted by the target population.

B. Potential of VIGART to Elicit Variation in Participants' Engagement Level

I wanted to assess whether the VIGART 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 VIGART, the participants' caregivers rated as to what they felt regarding the participants' engagement level while interacting socially with their virtual classmates. As engagement can be represented in terms of "sustained attention to an activity or person" [7], I asked the caregivers to rate the participants' engagement by observing their attention to the social task by using a 1–9 scale. Table III indicates that interaction with the VIGART was able to elicit a range of different engagement levels among the participants causing an increase of 64% from the lowest to the highest reported engagement levels, considering two out of five trials for each participant. Note that the lower rating on engagement given by the caregivers for five of the six participants (except ASD2) occurred at the initial trials and they

received higher individual ratings as the session progressed. 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. Only ASD2 among all the participants did not inquire about future participation possibilities with VIGART. Thus, it is reasonable to infer that the individualized feedback provided by VIGART using dynamic gaze measurement as the session progressed contributed towards the higher engagement ratings.

C. Objective Measures of Engagement through behavioral viewing Indexes [Sum of Fixation Counts (SFC) and Fixation Duration (FD)] Using VIGART

In the usability study I investigated the potential of VIGART to influence the behavioral viewing indexes of the participants. While caregivers' ratings showed an increase in engagement (Table III), I also wanted to investigate whether the objective measurement of behavioral viewing indexes, namely sum of fixation counts and fixation duration, showed corresponding variations. If such variations can be observed then these indexes can be used to map one's engagement within the VIGART, which may lead to new future intervention where objective metrics for engagement can be utilized within an intervention paradigm. In the present study, for our analysis, I labeled trials as eliciting low engagement (LE) and high engagement (HE) states based on the lowest and highest ratings on engagement for each participant, as reported by the caregiver. Analysis of sum of fixation counts for the participants with respect to Face_ROI and nonFace_ROI (i.e., Object_ROI and Others_ROI) corresponding to trials reported as eliciting low engagement and high engagement states, reveals improved sum of fixation counts percentage towards Face_ROI along with reduced fixation towards nonFace_ROI with higher engagement for every participant (Table III). Group analysis shows an increase of 13.77% towards Face_ROI and a decrease of 61.37% towards nonFace_ROI with higher engagement
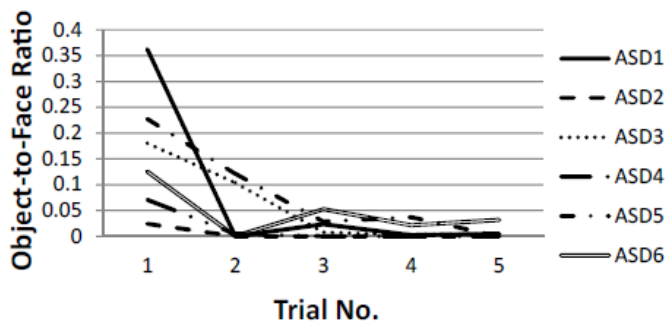

TABLE III COMPARATIVE ANALYSIS OF VARIATION IN THE BEHAVIORAL VIEWING INDEXES WITH REPORTED ENGAGEMENT LEVEL OF PARTICIPANTS

(Table III). Not only did the participants visit the Face_ROI more frequently, the average fixation duration of the participants indicates a trend towards improvement (i.e., increasing average fixation duration towards Face_ROI) in viewing patterns with increased engagement (Table III). In short, our investigation reveals that the potential of VIGART to bring about an overall improvement in behavioral viewing patterns in terms of increased sum of fixation counts and average fixation duration towards the Face_ROI accompanied with reduced distraction by NonFace_ROI. Thus, both the sum of fixation counts and the average fixation duration could be potential objective metrics for inferring engagement in VIGART-based intervention.

D. Objective Measures of Emotion Recognition Capability Through Behavioral Viewing Index Using VIGART

Further data analysis was carried out to examine the impact of the VIGART on the behavioral viewing indexes of our participants with ASD as function of their capability of recognizing emotions while they viewed their virtual classmates narrating personal stories and exhibiting context-relevant facial expressions. Note that our experiment was not designed to improve a participant's ability to correctly recognize an emotional expression, which will require a comprehensive investigation once the potential of VIGART is realized. However, since at the end of each trial, an emotion related question was asked to understand whether the participant was able to identify the facial expressions of his/her virtual classmates, I further performed an analysis to assess whether the emotion recognition capability and the gaze data were connected. In our usability study, I investigated fixation duration as a measure of emotion recognition because of its greater importance when recognizing faces and emotional expressions as compared to the number of fixations as discovered in non-VR-based studies. Our experimental findings revealed that all participants (except ASD5) fixated more on the face_ROI of their virtual classmates displaying emotions, either positive (i.e., Happy) or negative (i.e., Angry), as compared to no emotion (i.e.,Neutral). I found that there was an overall increase of 5.18% average fixation duration for Neutral-to-Happy expressions and an overall increase of 8.85% average fixation duration for Neutral-to-Angry expressions (Table IV). I examined the participants' responses to the emotion-identification questions and found that ASD5 could not identify the Neutral expression of his virtual classmate misidentifying it as Happy and I believe that this may account for his not looking more towards the Face_ROI. The above results indicate that VR-based social interaction can use average fixation duration as a quantitative measure to infer whether a participant with ASD has the ability to recognize emotion represented by facial expression, similar to the non-VR based tasks. In addition, the results show the potential of VIGART to map average fixation duration, measured in real-time to the participants' ability to recognize facial emotional expression of their virtual classmates. Thus, VIGART can be a step towards exploiting the behavioral viewing index in real-time as an objective measure which can serve as a tool while designing intervention strategies. However, a much larger study must be conducted before such findings can be generalized.

E. Eye Physiological Indexes (e.g., Blink Rate and Pupil Diameter) as Indicators of Emotion Recognition Capability

Children with ASD often experience states of emotional stress with autonomic nervous system activation without external expression [8] challenging their ability to communicate. However, their eye physiological indexes could be a valuable source to indicate their process of emotion recognition. In fact, literature indicates an important role of blink rate and pupil diameter in emotion recognition.

Past research shows startle potentiation for both positive and negative stimuli (static pictures) for children with ASD. Additionally, pupil has been considered as an indicator of emotion recognition. Studies indicate pupil to have sympathetic innervations with pupillary dilation to both pleasant and unpleasant auditory stimuli, being greater to unpleasant than pleasant visual stimuli.

Eye physiological indexes, namely blink rate and pupil diameter can be made continuously available within the VIGART using the feature extraction method (Section III-B2) when a participant socially interacts with the avatars. As a result, I analyzed to see whether the eye physiological indexes are also influenced by the interaction with VIGART.

Analysis of Changes in Blink Rate:

In our present study, results as presented in Table V, reflect a higher change in mean blink rate for Neutral-to-Angry (an overall increase of 107.93%) than that for Neutral-to-Happy (an overall increase of 49.91%) for all participants, except ASD3. In addition, mean blink rate for all participants (except ASD5) while viewing the Angry and Happy facial expressions of their virtual classmates were greater than that while viewing Neutral expression. A detailed analysis revealed that ASD3 and ASD5 could not identify the avatar's Angry facial expression. ASD3 responded to the angry face as Not Sure while ASD5 responded to it as Happy. Also, ASD5 could not identify the Neutral facial expression and misidentified this as Happy and he possessed a much higher mean blink rate in general, as compared to the other participants. Note that other researchers found similar trends in blink rate for non-VR based applications.

Analysis of Changes in Pupil Diameter:

Our findings on pupil diameter as presented in Table VI, are also in line with some of the previous non-VR based findings. I found that the mean pupil diameter of each participant was less while viewing the Happy faces than that for the Angry faces and both being greater than that for Neutral faces (except ASD5). Also, the change in mean pupil diameter for Neutral-to-Angry(overall increase of 8.84%) was found to be greater than that for Neutral-to-Happy (overall increase of 3.43%) for all participants (except ASD5) (Table VI).I examined the participants' responses to the emotion-identification questions and found that ASD5 could not identify the Neutral and Angry facial expressions, misidentifying them as Happy. I believe that this can be the reason for his percent increase in mean pupil diameter for Neutral-to-Angry to be lower than that for Neutral-to-Happy.

In the present usability study, with a limited sample size, I find that the blink rate of the participants is more sensitive to their ability to recognize different emotional expressions exhibited by their virtual classmates, than the pupil diameter. For pupil diameter, the percent change for Neutral-to-Happy and that for Neutral-to-Angry though quite small the overall trend is similar to that of other non-VR based studies. More importantly, the above results indicate the ability of VIGART to measure one's blink rate and pupil diameter in real-time and correlate these as function of his/her ability to recognize emotions while interacting socially with virtual classmates. The results suggest that a participant's eye physiological response in VR-based social interaction as presented in VIGART may indicate whether or not one is able to recognize emotion, which is similar to what has been observed in non-VR based tasks. Thus, it is reasonable to believe that VIGART could be used in intervention, perhaps as a supplementary tool, to allow a child with ASD to enhance his/her social skills.

V. Conclusion

Prior research has demonstrated that children with ASD commonly exhibit atypical viewing patterns, such as greater fixation towards non- social objects than human faces during social interaction. These deficits coupled with the wide spectrum of ASD necessitate the use of individualized intervention. Regardless of definitive understanding of origins or sustaining developmental and biological mechanisms, differences in viewing patterns are thought to be on a basic level. Research indicates the potential of emerging technology to play a significant role in providing individualized intervention. With this motivation, in the current study, I design a novel VR, which seamlessly integrates the VR technology with dynamic eye-tracking to create a novel virtual gaze-sensitive social interaction platform and provides individualized feedback based on one's performance and real-time viewing.

Now a days, a usability study with six adolescents with ASD, to test the acceptability of VIGART and also investigate how these adolescents respond in terms of their engagement level, behavioral viewing and eye physiological indexes. From the exit survey, I can infer that the VIGART has a potential to be accepted by the target population. The VIGART can measure behavioral viewing and eye physiological indexes dynamically that have the potential to be used as objective measures of one's engagement as well as emotion recognition capability. Such capabilities, in our opinion, make VIGART unique in the sense that this intelligent system could be effectively used as a potential intervention tool where objective quantifiable measures could guide refinement of intervention strategies.

The preliminary results of the usability study are promising. Overall, our results indicate that the impact of VIGART is similar to other non-VR-based tasks. However, a much larger study must be conducted before such findings can be generalized. Also note that, at this current configuration, VIGART uses a wearable eye-tracker, which may not be suitable for children with low-functioning ASD. However, I plan to use a noncontact desktop-based eye-tracker to mitigate this problem. It is worth mentioning that VIGART is capable of providing extensive closed-loop 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 accordingly the avatars can change their mode of talking to allow a higher level of interactivity between the participant and VIGART. Such

sophisticated interaction will be performed in the future. The present usability study shows, in principle, that VIGART has the potential to be used as a supplement to real-life social skills training tasks in an individualized manner. However, I acknowledge that current findings, particularly toward skill improvement, are preliminary. While demonstrating proof of concept of the technology and trends of "improved" behavioral viewing in a VR-based social task, questions about the practicality, efficacy, and ultimate benefit of the use of this technological tool for demonstrating clinically significant improvements in terms of ASD impairment remain, which will be addressed by empirical investigation in the future.

REFERENCES

[1] A service of the U.S. National Library of Medicine

[2] 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.

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

[4] 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.

[5] Parsons S, Mitchell P (2002) The Potential of Virtual Reality in Social Skills Training
for People with Autistic Spectrum Disorders. J Intellect Disabil Res 46: 430-443.

[6] S.Weider and S. Greenspan, "Can children with autism master the core deficits and become empathetic, creative, and reflective?," J. Develop.Learn. Disorders, vol. 9, 2005.

[7] L. Itti, N. Dhavale, and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in Proc. Int.v Symp. Opt. Sci. Technol., 2003, vol. 5200, pp. 64–78.

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

AUTHORS PROFILE

Miss A W Chorey persuing her Post graduation in Computers Scinece and Engineering from Amravati Univuersity.