Multimodal Adaptive Social Interaction in Virtual Environment (MASI-VR) for children with Autism Spectrum Disorders (ASD)

E. Bekele[1]*, *Student Member, IEEE*, J. Wade[1], D. Bian[1], J. Fan[1], Amy Swanson[2], Z. Warren[2,3], and N. Sarkar[1,4]*, *Senior Member, IEEE*

[1]Electrical Engineering and Computer Science Department, [2]Treatment and Research in Autism Spectrum Disorder (TRIAD), [3]Pediatrics and Psychiatry Department, [4]Mechanical Engineering Department, Vanderbilt University, Nashville, TN, USA

Abstract

Difficulties in social interaction, verbal and non-verbal communications as well as repetitive and atypical patterns of behavior, characterizes Autism spectrum disorders (ASD). A number of studies indicated that many children with ASD prefer technology and this preference can be explored to develop systems that may alleviate several challenges of traditional treatment and intervention. As a result, recent advances in computer and robotic technology are ushering in innovative assistive technologies for ASD intervention. The current work presents design, development and a usability study of an adaptive multimodal virtual reality-based social interaction platform for children with ASD. It is hypothesized that endowing a technological system that can detect the processing pattern and mental state of the child using implicit cues from eye tracking and electrophysiological, including peripheral physiological and electroencephalography (EEG), signals and adapt its interaction accordingly is of great importance in assisting and individualizing traditional intervention approaches. The presented VR system is based on a virtual reality based social environment, a school cafeteria, where an individual with ASD interacts with virtual characters. An eye tracker, an EEG monitor and biosensors to measure peripheral electrophysiological signals are integrated with the VR task environment to obtain gaze, EEG signals and several peripheral physiological signals in real-time. In the current work, we show how eye gaze and task performance can be used in real-time to adapt intervention in VR. The other signals are collected for offline analysis. The results from a usability study with 12 subjects with ASD are presented to demonstrate the viability of the proposed concepts within the VR system.

Keywords: 3D social Interaction, multimodal interaction, psychology, usability, VR-based adaptive systems.

1 Introduction

Atypical patterns of behavior, communication and social interaction impairments are characteristics of Autism spectrum disorders (ASD). With prevalence rate of 1 in 68 children [1] in the US, it is a public health emergency. Researchers have shown that children with ASD have fewer social communication skills and that this deficit is related to executive function skills [2]. It was shown that children with ASD exhibit severe deficits in facial and vocal affect recognition, social judgment, problem solving and social functioning skills [3] and hence a deficit in social interaction is the major defining feature of ASD. Generally, these common social and communicational deficits are observed in most children with ASD, however, the manifestation of these deficits is quite different from one individual to another [4]. These individual differences call for approaches to individualize the therapy as opposed to one-therapy-fits-all strategies. One of the fundamental social impairments in ASD population is their inability to recognize and understand facial emotional expressions [5-9]. Specifically, individuals with ASD have been shown to have impaired face discrimination, slow and atypical face processing strategies accompanied by reduced attention to eyes [8]. Although the pathology of the underlying neural processes is poorly understood, some neural studies indicated that children with ASD employ different neural networks and rely on different strategies than their control groups for face processing [10]. In general, children with ASD have shown significant impairment in understanding complex facial emotional expressions in the presence of social context or otherwise [11, 12].

There exist traditional intervention paradigms that seek to mitigate these impairments [13]. For example, in a 7-month long behavioral intervention that involves social interaction and social emotional understanding, Bauminger showed that children with ASD showed improved social functioning and understanding, when they recognized and displayed complex emotional expressions [13]. However, traditional behavioral intervention requiring intensive behavioral sessions is not accessible to the vast majority of ASD population due to lack of trained therapists as well as intervention costs. Moreover, accessible traditional intervention results in excessive life time costs [14-16].

Technology has the potential to individualize and increase effectiveness of traditional human-centric autism therapy. Recently computer technology [17], robot-mediated systems [18, 19], and virtual reality (VR) [20-22] have been proposed for ASD intervention. Innovative adaptive technology promises alternative or assistive therapeutic paradigms in increasing intervention accessibility, decreasing assessment efforts, promoting intervention, reducing the cost of treatment, and ultimately helping in skill generalization to real world social interactions [17].

In this context, emerging technology such as virtual reality (VR) [21, 23-26] in particular has the potential to offer useful technology-enabled therapeutic systems for children with ASD. Virtual reality environments offer benefits to children with ASD mainly due to their ability to simulate real world scenarios in a carefully controlled and safe environment [27, 28]. Controlled stimuli presentation, objectivity and consistency, and gaming factors to motivate task completion are among the primary advantages of using VR-based systems for assistive ASD intervention.

VR platforms have been shown to be promising for improving social skills, cognition and overall social functioning in autism [22]. However, existing VR systems as applied to autism therapy primarily focus on performance, explicit user feedback [29], and remote-operation [22] and thus limit adaptive interaction [30, 31]. Furthermore, existing VR-based systems do not incorporate implicit cues from sensors such as electrophysiological signals including peripheral physiological signals [32, 33] and electroencephalography (EEG), and eye tracking [24] within the VR environment, which may further help in facilitating individualization and adaptation, and possibly acceleration of learning emotional cues. Recently, adaptive social interaction using implicit cues from sensors such as peripheral physiological signals [21, 34] and eye tracking [24] was shown to be possible in VR-based autism therapy [35].

The vast majority of earlier VR systems as applied to ASD intervention were performance based systems. However, recent research in VR systems for application of Attention Deficit Hyperactivity Disorders (ADHD), ASD, and cerebral palsy suggest making such feedback based VR systems facilitates increased interactions and individualization [36]. Some of these studies used touch screen feedbacks to teach children with ASD skills such as pretend play, make eye contact, take turns and share skills using a co-located cooperation enforcing interface, called StoryTable [37, 38]. For further reference, the reader is encouraged to refer [27, 29, 31, 39-

IEEE Virtual Reality Conference 2016
19–23 March, Greenville, SC, USA
978-1-5090-0836-0/16/$31.00 ©2016 IEEE

41] for a complete discussion of social interactions in VR. While haptic feedback can be useful in certain contexts, understanding eye gaze and physiological response during emotion recognition can be critical since both eye gaze and physiology have been shown to convey a tremendous amount of information regarding the emotion recognition process [21, 24, 32, 42].

There are a few recent studies that attempted to incorporate peripheral psychophysiological signals [21, 32] and eye gaze monitoring [24] into VR systems as applied to ASD intervention. These systems monitor several channels of physiological signals to determine the underlying affective states of the subject for individualized VR social interactions. In the present study, we extend these preliminary attempts to incorporate implicit cues from various modalities and study within and intergroup variations in psychological state patterns and eye behavioural patterns for online adaptive individualized VR-based multimodal social interaction. The proposed system implements a novel online gaze-sensitive occlusion paradigm that teaches children with ASD proper processing of emotional faces. Moreover, it also collects objective performance metrics for dynamic adaptive individual feedback generation as well as collection of multimodal data for further analysis for affect elicitation and engagement.

In this paper, we present the design, development and a usability study based evaluation of the novel Multimodal Adaptive Social Interaction in VR (MASI-VR) system that can present controlled facial emotional expressions in the presence of a conversational social context. The system tracks eye gaze and generates eye scanning pattern and uses it for online gaze-sensitive adaptation for ASD intervention for social interaction. Moreover, it also collects electrophysiological signals including peripheral physiological signals and EEG data related to emotion recognition in a synchronous manner. We believe that such an ability will provide insight to the emotion recognition process of the children with ASD and eventually help designing new intervention paradigms to address specific core social impairments and emotional deficits.

The objective of this work is twofold: (1) to design and develop an innovative adaptive multimodal VR-based social interaction platform for ASD intervention. The platform integrates eye gaze, EEG signals and peripheral psychophysiological signals of the participants to understand their emotion processing and engagement. The system is designed to create dynamic social situation where the participants can interact with characters using a spoken dialog management system, and (2) to perform a usability study to demonstrate the usefulness of the designed social VR system for emotional face processing.

Our system is significantly different from existing systems and this paper contributes in two important ways: (1) employment of an online gaze-based adaptation using occlusion paradigm that occludes the emotional face and reveals parts of the face progressively as the subject looks on context relevant regions of interest (ROI) in an effort to guide the subject to properly process the emotional face, and (2) the presence of a domain restricted question-answer based conversational dialog with the virtual characters to teach proper emotion processing in the presence of social contextual interaction. The conversation was used as a backdrop for the main task of this study, i.e., emotion recognition in social context.

The remainder of the paper is organized as follows. Section 2 describes the overall system design and development and each component of the VR system in detail. Section 3 highlights the methods and procedures followed in the usability study. Results are presented in Section 4. Finally, the overall contributions are summarized in Section 5.

2 System Design

We have designed the VR system to model the different aspects of an emotional social interaction. Social interaction is defined by various components including emotional expressions and verbal communication. The VR system for adaptive multimodal social interaction is composed of 5 major components: (1) an adaptive social task presentation VR module, (2) a spoken conversation management module (Q/A-based dialog management module), (3) a synchronous physiological signal monitoring module, (4) a synchronous EEG monitoring, and (5) a synchronous eye tracking and online adaptive gaze feedback module (Fig. 1).

The system is a distributed system with a central supervisory controller. Each peripheral interface components connects to the MASI-VR task engine using a distributed modular network interface. The peripheral components get events happening during the social interaction with a central timestamp via the supervisory controller.

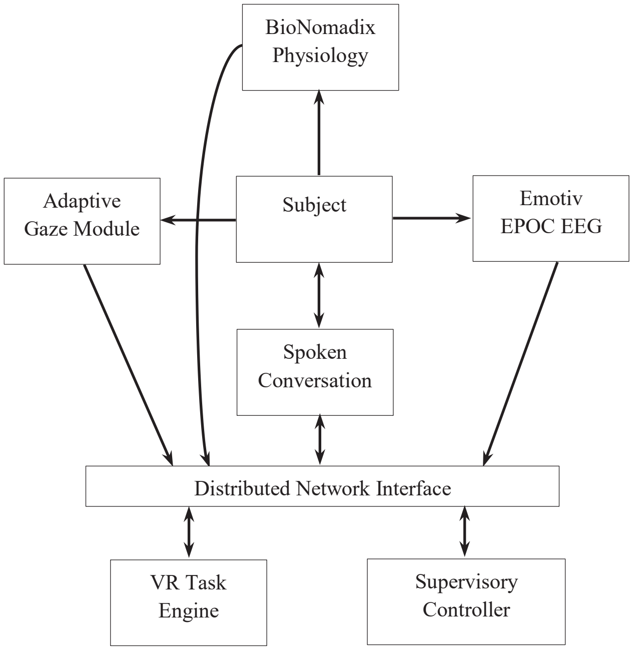

Fig. 1. System architecture of MASI-VR

The supervisory controller facilitates the event synchronization between the VR task presentation engine and the peripheral interfaces. In addition to the implicit cues collected from physiology, eye tracking and EEG, we have designed a spoken conversational dialog management module that interacts with the VR engine as one of the peripheral interfaces to provide speech recognition and dialog management services. In order to undertake naturalistic social interaction several components including conversational dialog, body language (gesture), facial emotional expressions and eye contact need to be considered. Conversational dialog is an important part of social interaction. Recently spoken conversational modules have been incorporated to VR systems to achieve more natural interaction instead of menu driven dialog management. Instead of large vocabulary, domain independent natural language understanding, limited vocabulary question-response dialog management, which is focused on the specific domain, has been shown to be effective [43, 44]. Such multimodal interaction helps in individualization and in cases of inaccessibility of trained therapists, it may serve as a self-contained therapeutic system. Proper facial emotional expressions recognition and appropriate gaze fixation pattern are considered to be an important building block for individuals with ASD to alleviate their overall social interaction impairment. Although isolated emotion recognition in VR was demonstrated earlier, proper processing of emotional faces in the presence of a social context, a dialog in this case, is of paramount importance and appropriate for the VR-based core skills training to generalize into real-world interactions. To aid in proper processing of emotional faces, we have designed a facial occlusion paradigm in which the subject sees an oval occlusion on the face and the occlusion gets revealed progressively as the subject scans the face appropriately by looking at context relevant regions of the face such as the eyes and the mouth.

2.1 The MASI-VR task presentation engine

The development of the virtual reality environment involved a pipeline of design and animation software packages. Characters were customized and rigged in online animation and rigging service, mixamo (www.mixamo.com), and Autodesk Maya (www.autodesk.com). They were animated in Maya and imported into the Unity game engine (www.unity3d.com) for final task presentation.

2.1.1 Characters rigging and animations

The characters used in this project were customized in mixamo to suit the 13-17 year age group targeted for the usability study. A total of 7 characters including 4 boys and 3 girls were selected and customized for the embedded facial expression display at the end of each conversational mission and another set of 12 characters for primary social interaction. Fig. 2 shows some of the characters with various animations.

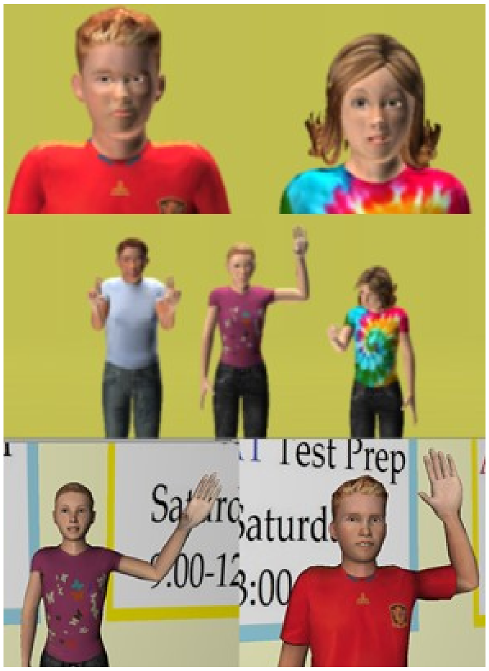

Fig. 2. Various emotion and gestural animations.

Facial emotional expressions and lip-syncing were the major animation of this project. The universally accepted seven emotional expressions proposed by Ekman were used in this project [45]. The expressions are: enjoyment, surprise, contempt, sadness, fear, disgust, and anger. Each facial expression had 4 arousal levels: low, medium, high, and extreme. The four levels were chosen by careful evaluation by clinical psychologists involved in this project. In addition to these facial expressions, seven phonetic viseme poses were created. The phonemes were L, E, M, A, U, O, and I. These phonemes were used to create lip-synced speech animations for storytelling. For further details, please refer to [46].

2.1.2 The VR Environment

The VR task presentation engine used the popular game engine Unity (www.unity3d.com) by Unity Technologies. The peripheral psychophysiological monitoring used the wireless BioNomadix physiological signals acquisition device by Biopac Inc. (www.biopac.com). The eye tracker employed in the study was the Tobii X120 remote desktop eye tracker by Tobii Technologies (www.tobii.com). The venue for the social interaction task is a virtual school cafeteria (Fig. 3).

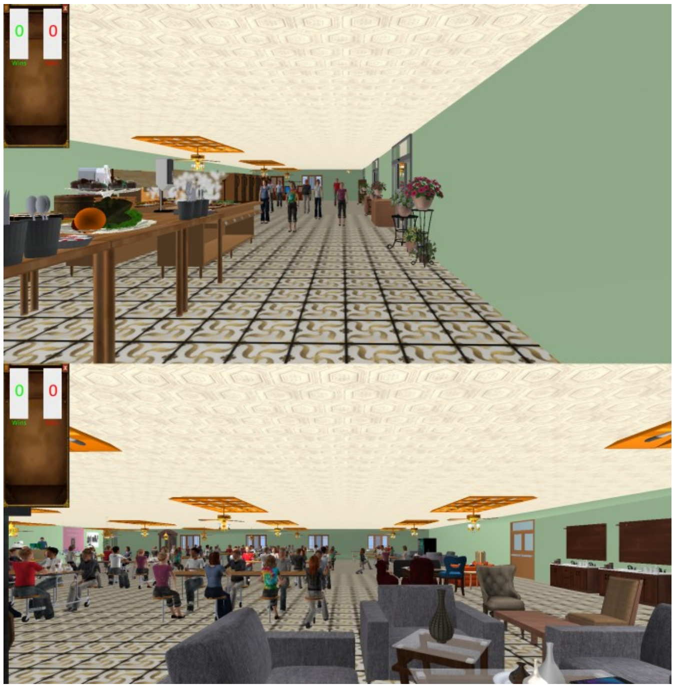

Fig. 3. The VR cafeteria environment for the social task.

2.2 The Dialog Management System

We have developed domain dependent conversation threads for more reliable speech-based interactions and a dialog management engine that parses these threads and performs a lexical comparison between each of the dialog options and the user utterance as captured from a speech interface module within a specified time interval. The speech interface module was developed using the Microsoft Speech API (MS SAPI) SDK. The speech recognition engine expects a domain knowledge supplied to it in the form of a context free grammar (CFG). Given the grammar XML file, and a speech utterance by the user, the recognition engine produces a hypothesized speech sentence together with confidence probability. We then use individually trained probability threshold to accept or reject the speech recognition result. For further details on the dialog management system, see [46].

2.3 Peripheral physiological monitoring

Affective state recognition using peripheral physiological signals was formally introduced in [47, 48]. Estimating the psychological affective states of subjects is important for technology-assisted therapy and it enables implicit and meaningful human-machine interaction. The physiological affective module for this study collected 8 channels of physiological signals for later offline analysis of affect. We collected labelled data for preliminary feature level analysis and future supervised online affect recognition module for interaction adaptation based on affective state of the subject. We utilized wireless sensors form Biopac Inc. (www.biopac.com) called BioNomadix. The collected physiological signals were analysed for feature level statistical comparisons to see any pattern differences between the pre and post conditions of the same group and between the two groups (See details of the experimental protocol in Section 3).

2.4 EEG Monitoring Module

We also introduced an electroencephalography (EEG) monitoring module as part of the electrophysiological monitoring. 14 channels of EEG signals were collected from the scalp of the subjects using the Emotive EPOC neuroheadset (www.emotiv.com) with a sampling rate of 128Hz. The channel locations were AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, according to the international 10-20 system of electrode placement [49]. The common mode sense and driven right leg references were placed at locations P3 and P4 respectively. Similar to physiological monitoring, EEG monitoring module received trial and session markers from the task presentation engine via socket communication. For the purpose of this study EEG analysis is limited to offline feature level comparisons across groups and conditions. Six EEG features that related to affective stimuli processing were extracted from pre-processed EEG signals, including theta band (4-8Hz) power at right parietal area (P8), averaged theta2 band (6-8Hz) power at left anterior areas (AF3, F7, F3, and FC5), averaged alpha2 band (10-12Hz) power at anterior areas (AF3, F7, F3, FC5, FC6, F4, F8, and AF4), alpha2 band (10-12Hz) power at right parietal area, averaged beta1 band (12-18Hz) power at anterior areas, and gamma band (30-45Hz) power at right parietal area. Increase in power of theta ranges and high frequency activity, beta and gamma, were found to associate with processing of the affective stimuli, while the alpha2 bands were related to cognitive involvement during emotional stimuli processing [50, 51].

2.5 Eye Gaze Monitoring and Online Adaptation

The gaze data analysis was performed to determine behavioral viewing patterns of children with gaze group as compared to that of the control. The behavioral indices such as where they were looking in terms of screen coordinates were clustered into ROIs defined around the key facial bones. The clustering results were then averaged over trials for each subject and the aggregate results were used. The defined ROIs represented the following regions: forehead, eyes (left and right), nose, and mouth. The face region was modelled by a combination of an ellipsoid and a rectangular forehead region (Fig. 7). Facial regions outside of the 5 defined regions of interest were categorized as "other face regions" while all the background environment regions outside of the face regions were defined as "non-face regions". This gave a total of 4 regions into which all the gaze data points were clustered.

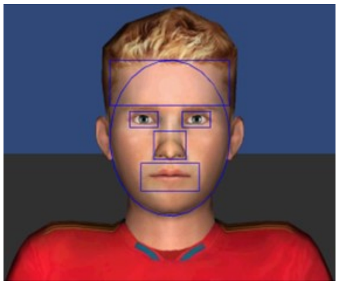

Fig. 4. The five facial ROIs defined on the face region.

The other behavioral index considered for analysis was the fixation duration. The raw fixation duration was computed for each gaze point during the online interaction. The raw data were first filtered to remove excessively small and large fixation durations. Typical fixation duration and saccades last between 200 and 600 ms and 30 and 120 ms, respectively [52]. The filtered fixation duration data were used to compute the average fixation duration (FDave). Another important behavioral eye index associated with fixation duration, called the total sum of fixation counts (SFC), was also computed from the filtered fixation duration data.

The eye physiological indices, i.e., the blink rate (BR) and the pupil diameter (PD) were also post processed. Missing PD data due to blinks and presence of noise was filtered from the collected PD data. First a threshold was established to segment missing noisy data from the actual PD. Then the missing data points were constructed by linearly interpolating the neighbouring data points. The BR data were also filtered based on typical blink ranges. Typical human blinks range between 100 and 200 ms [53]. The PD data were used to reject blinks at missing data points.

3 Methods and Procedure

A usability study was conducted to validate the system and to study the behavioral and physiological pattern difference of children with ASD that participated with a gaze-sensitive version of the system and control group that participated without the online gaze adaptation and occlusion paradigm. Subjects were randomly assigned to one of the two groups to control for bias.

3.1 The MASI-VR Protocol

Each subject in the experimental protocol went through a typical 3 visits protocol. For some of the subjects that were unable to perform the pre and the post-tests together with the social task training, the sessions were spaced out to 5 visits. Table I describes what the subject was doing in each visit. First of all, the subject performs a pre-test of isolated emotion recognition task in VR without the conversational dialog context and a standardised "A Developmental NEuroPSYchological (NEPSY) Assessment" with emotion recognition components followed by the first session of the social task in visit 1.

Table I Summary of Visits

V1	Informed consent
	FE Pre-test + NEPSY Pre
	1st exposure to conversation paradigms - emotions at high valence
V2	2nd exposure to conversation paradigms - emotions at medium valence
V3	3rd exposure to conversation paradigms - emotions at low valence
	FE Post-test + NEPSY Post

FE: Facial Expression

Then, in visit 2, the subject performed the second session of the social task. Finally, in visit 3, the subject goes through the third session of the social task, the post-test isolated emotion recognition in VR which is the same as the pre-test, and the post-test of NEPSY.

The difference between the three sessions of the social task is the emotion intensity level of low, medium and high, respectively. The pre-post-test contains isolated emotion recognition of 28 trials whereas the social tasks have 12 conversational missions (goals). At the end of each mission, 2 emotional faces were presented with the occlusion paradigm for the gaze group and without for the control group for a total of 24 emotion recognition trials in the presence of social context.

3.2 Experimental setup

The VR environment was run on Unity. Eye tracking and peripheral physiological monitoring were performed in parallel on separate applications that communicated with the Unity-based VR engine via a network interface as described in Section 2. The VR task was presented using a 24'' flat LCD panel monitor. The experiment was performed in a laboratory with two rooms separated by one-way glass windows for parent observation. The parents sat in the outside room. In the inner room, the subject sat in front of the task computer. A therapist was present in the inner room to monitor the process. The task computer monitor was also routed to the outer room for parent observation. The session was video recorded for the whole duration of the participation (Fig. 5).

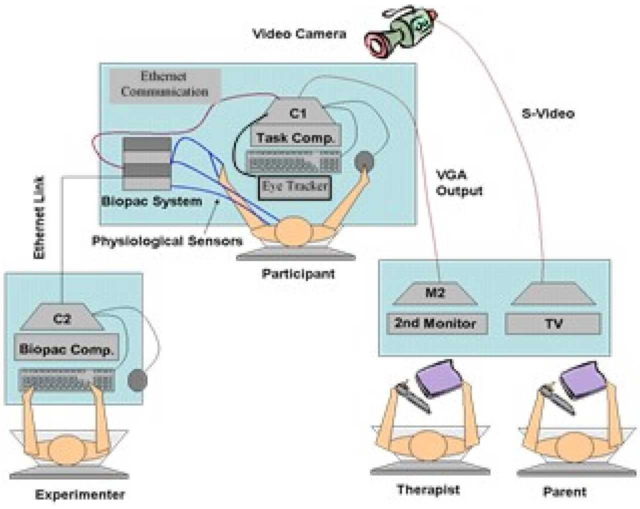

Fig. 5. The experimental setup.

3.3 Subjects

A total of 6 high functioning subjects with ASD with no mental retardation record (Male: n=6,) of ages 13 – 17 (M=15.77, SD=1.87) and an age matched 6 (Male: n=6) controls of ages 13 – 17 y (M=15.20, SD=1.68) were recruited and participated in the usability study. All ASD subjects were recruited through existing clinical research programs and had established clinical diagnosis of ASD. All subjects with ASD fall well above the clinical threshold. The gold standard in clinical ASD diagnosis, the Autism Diagnostic Observation Schedule-Generic (ADOS-G) the new algorithm score and the severity score (ADOS-SS) were used to recruit the ASD subjects. IQ of the ASD subjects was obtained from existing clinical research database.

The control group were trained using the system without any online gaze feedback and without the occlusion paradigm. In addition to ADOS-G, we asked parents of the subjects to fill out the social responsiveness scale (SRS)[54] and the social communication questionnaire (SCQ) [55]. Parents of both groups have completed these forms. In addition, WASI [56] was used to measure IQ of the subjects. The IQ measures were used to potentially screen for intellectual competency to complete the tasks. Moreover, we used a measure of facial recognition called NEPSY together with an isolated facial expression recognition was administered to the subjects as pre and post measures before and after their repeated training with the social task. Again all the CTR subjects were well above the clinical cut-offs for the SRS and the SCQ as well as ADOS-G.

3.4 The MASI-VR Social Task

Initially the subject is seated in front of the task computer. The eye tracker is calibrated and the peripheral interfaces are all connected via network. Once this is finished, the subject information is entered into the system. Then the subject is ready to start the task. At the start of the social task, the subject will go through a sequence of instructions on how to perform the task. Once the instruction is finished, the subject gets to move around the virtual cafeteria and approach a character to interact with. Each conversational character has a personal space which was implemented using Unity's collider triggers. Once the subject enters the invisible personal space of the character, the character invites the subject to interact by indicating it is ready to talk to the subject. If the subject chooses not to interact with that specific character, the message gets reset once the subject moves out of the personal space of the character. Once the subject chooses to interact with a specific character, then the subject is presented with the available 4 missions in low level. The subject can choose any of the 4 missions. Once the subject chooses the mission, the subject is asked to choose the conversational topic in order to measure if the subject understood the mission. Then the subject will enter the conversation dialog and the conversational dialog management engine goes through series of trials of conversation depending on the subject's success and failure as shown in Fig. 6.

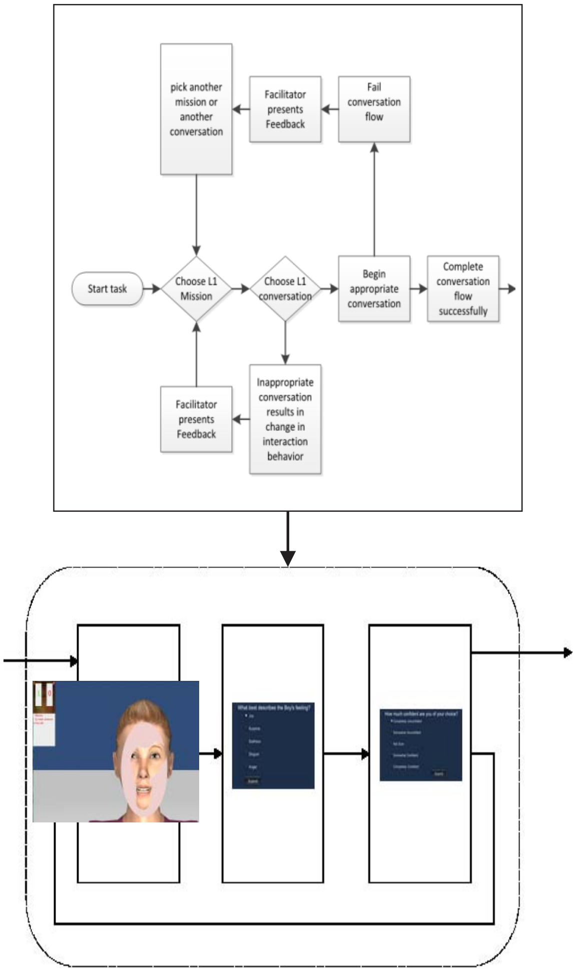
pick another mission or another conversation
Facilitator presents Feedback
Fail conversation flow
Start task
Choose L1 Mission
Choose L1 conversation
Begin appropriate conversation
Complete conversation flow successfully
Facilitator presents Feedback
Inappropriate conversation results in change in interaction behavior

Fig. 6. The spoken conversation and the emotion recognition

The VR-based facial emotional recognition in the presence of conversational dialog system presented a total of 12 conversational dialog missions. At the end of each dialog mission two facial expressions were presented with the face occluded with oval occlusion as shown in Fig. 9. As the subject scans the face, the occlusion erodes by the gaze of the subject to give an online adaptive gaze feedback. If the subject pays attention to the context relevant areas of eyes and mouth beyond a threshold, the face reveals with the emotion and the subject gets to choose what the emotion was. If the subject was not successful in revealing the face in 15s, the face reveals itself and the process continues normally.

The emotions in the pre-post-tests consisted of a total of 28 trials corresponding to the 7 emotional expressions with each expression having 4 levels. Each trial was 30-45 s long. For the first 25-40 s, the character narrated a lip-synced context story that was linked to the emotional expression that followed for the next 5 s. The character exhibited a neutral emotional face during story telling. Subjects were expected to rate the emotions based on the last 5 seconds of interaction. The story was used to give context to the displayed emotions. The context of the stories ranged from incidents at school to interactions with families and friends that were suitable for the targeted age group. However, since this VR task was used as purely a pre and post measure, the story was not interactive and the audios of the stories were recorded with monotonous tone so the subject gets to decide the emotion solely based on the isolated expression and not get influenced by the story.

A typical laboratory visit was approximately one hour long. During the first 15 minutes, a trained therapist read approved ascent and consent documents to the subject and the parent, and explained the procedures. Once the subject finished signing the ascent document, he/she began the task. While the parent completed the SRS and SCQ forms, the subject wore the wearable physiological sensors with the help of a researcher. Before the task began the eye tracker was calibrated. The calibration was a fast 9 points calibration that took about 10-15 s. At the start of the task, a welcome screen greeted the subject and described what was about to happen and how the subject was to interact with the system. Immediately after the welcome screen, the trials started. At the end of each trial, questionnaires popped up asking the subject what emotion he/she thought the character displayed and how confident he/she was in his/her choice. The total participation time was about 20-25 minutes. The emotional expressions presentations were randomized for each subject across trials to avoid ordering effects. To avoid other compounding factors arising from the context story, the story was recorded with monotonous tone and there was no facial expression displayed by the character during that period.

4 Results and Discussion

4.1 Performance

Performance of the subjects in identifying the emotions in the five sessions (three social task training sessions, S1, S2, and S3, and pre and post isolated facial emotional recognition sessions) was measured using three metrics. Their total score was gauged as a percentage of the total trials in each session. The subjects also indicated how confident they were in their choices immediately after they made their choices. The time they took to indicate their choices was also considered as a performance metric and calculated as latency of response in seconds.

Fig. 7 shows that both the gaze and the control group were able to improve their performance from pre to post while the gaze group improves by 3% more than the control group's improvement. However, the performance score differences were not statistically significant. The confidence metrics indicated that gaze group subjects were relatively as confident as the ones in the control group while taking slightly more time across all the sessions.

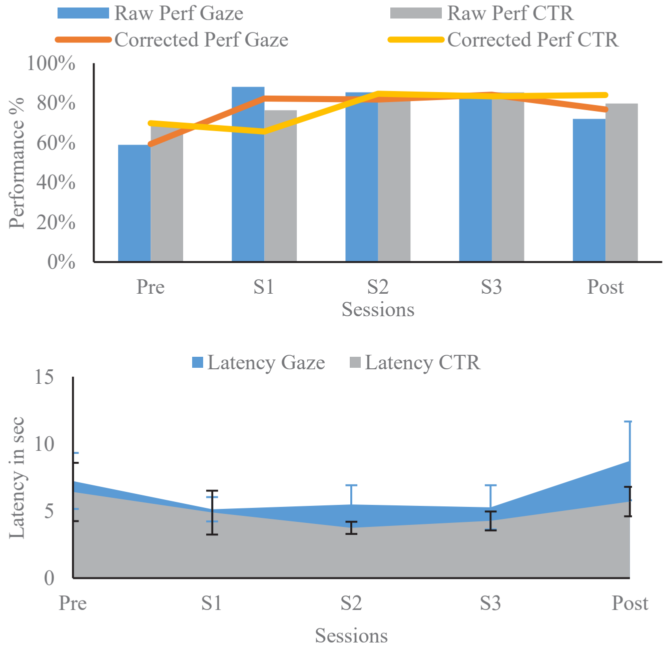

Fig. 7. Generalized performance metrics

To further control for the effect of choice bias, i.e. subjects choosing a specific emotion selection regardless of stimuli, we computed the bias for each emotion and corrected the raw performance by removing the bias. Fig. 8 shows the bias for each emotion across subjects for the pre and post sessions with the expected bias, which is 14.29% for 7 emotions, for reference. We can clearly see that subjects were more biased for some emotions such as fear over the others. The bottom figure shows the raw performance and the bias corrected performance overlaid on top as lines. In summary, the results indicated that MASI-VR gaze-sensitive system was effective in making the gaze group close the performance difference by more than 3% from pre to post without significant difference in confidence and latency with the control group.

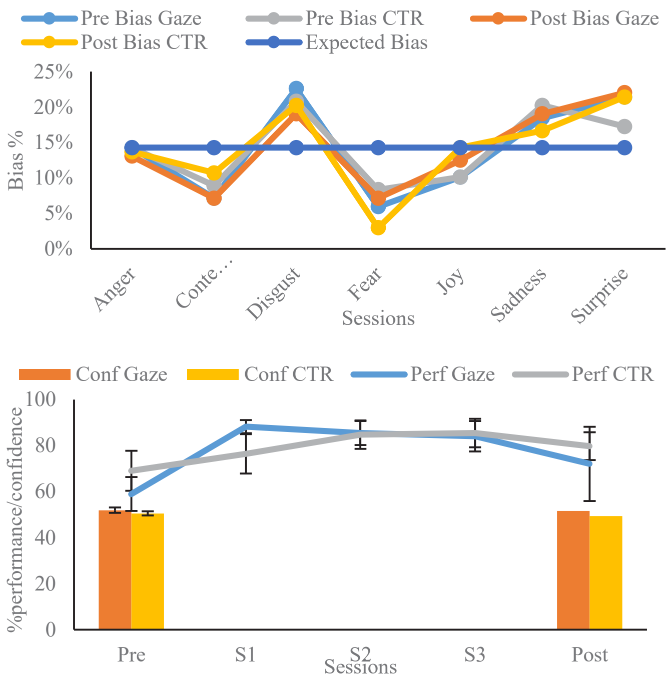

Fig. 8. (top) bias per emotion, and (bottom) raw and bias corrected performance

4.2 Gaze towards ROIs

This is the core gaze analysis to determine if subjects improved due to the social emotion recognition training by MASI-VR.

4.2.1 ROIs and gaze analysis

Fig. 9 presents the gaze towards the various ROIs that were defined for the VR task as shown in Fig. 4. The plot shows time spent in that specific ROI as a percentage of total trial time.

Data was averaged across trials for each subject and across subjects in each group. In the context relevant areas such as the eyes, both groups increased paying attention from pre to post and decreased gaze towards the mouth ROI from pre to post with the gaze group decrease of 10%, $p<0.05$, while the control group decreased gaze to the mouth by 5%. In a similar manner the gaze towards other parts of the face also increased as the forehead ROI. Since there are many facial bones that were animated on the forehead that move during the emotional expression trials, the increase in the forehead was not surprising. The gaze group actually increased more than the control group towards the forehead ROI. There were no statistically significant differences between the two groups. Most of the significant changes were within group changes of the gaze group towards the mouth and the control group towards the eyes.

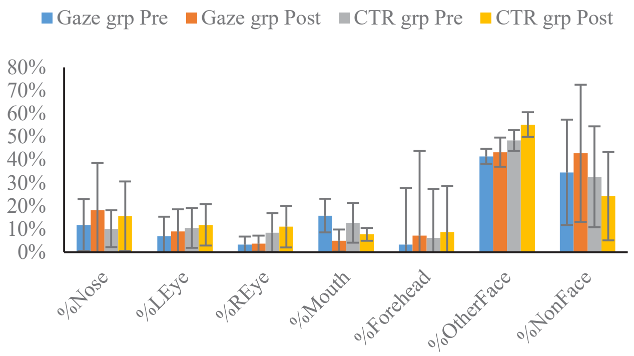

Fig. 9. Gaze towards ROIs defined in Fig. 4.

To further enhance the differences, we have combined several ROIs. The right and left eyes were collapsed into one measure towards the eyes, all the face region, and the core face ROIs including eyes and mouth. Further we combined all the ROIs on the face into a single ROI, roiFace and the ROIs that might interfere with the eyes such as the nose and the forehead combined together with the eyes as comboFace. Both groups increased in the eyes, roiFace and the comboFace ROIs. The gaze group increase in the comboFace was specifically larger than the control group. The total face area ROIs were combined and compared with non-face ROIs as well. The gaze group decreased on the face while increasing on the non-face ROIs while the control group displayed the opposite behaviour.

4.2.2 Gaze visualizations

We have generated heat map and masked map visualizations to qualitatively compare the differences from pre to post in both groups collapsed across all trials and all subjects. The heat map was smoothed using a Gaussian filter after the fixation duration was accumulated and computed for each region to remove unnecessary discontinuities and minimize the effect of noise on the visualization.

Fig. 10 shows representative gaze pattern towards a character across trials and across all subjects in the gaze group. The gaze group increased their gaze towards the eyes with a more symmetric gaze in the post test (bottom) than the pre-test. The change in symmetry in the gaze processing pattern is consistent over all 7 characters the subjects identified the emotions with.

The control group also had a more symmetric and slightly increased gaze towards the eyes in the pre-test than the gaze group. However, the slightly decreased and the gaze slightly faded towards the eyes (Fig. 11).

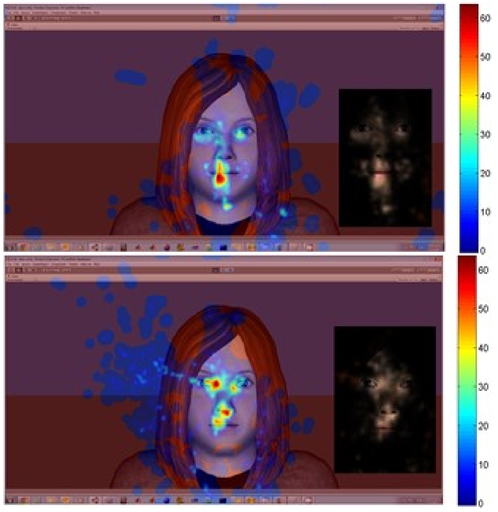

Fig. 10. Gaze Group Pre (top) and Post (bottom).

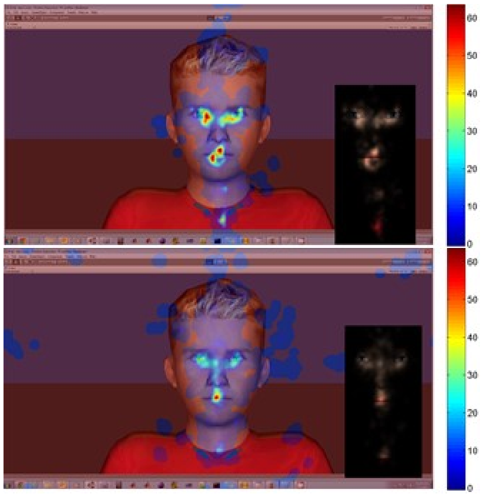

Fig. 11. Control (CTR) Group Pre (top) and Post (bottom).

4.3 Analysis of Eye Features

In addition to the gaze towards ROIs and qualitative visualizations, we analysed 5 features from the eye tracking data. There were: the pupil diameter (PD), the average fixation duration (FDave), sum of fixation counts (SFC), blink rate (BR), and saccade path length (SPL). These are measures of behavioral viewing patterns in this study. Section 2 describes how these indices were computed. These behavioral indices are indicative of engagement to particular stimuli and are correlated with social functioning for individuals with autism [57, 58]. Generally, children in the gaze group had slightly lower FD and slightly higher SFC than the control group in the pre and post sessions. However, the variations were not as such statistically significant and hence we computed correlation between these features in order to come up with a combined gaze engagement index. Fig. 12 shows the correlation matrix and it is evident that features 3, 4, and 5 are highly correlated. So, we choose these highly correlated features and the fixation duration as it is directly correlated with engagement. Then we combined the three together with inverse and add it to the normalized fixation duration. From the figure it is evident that the three features are inversely correlated with fixation duration.

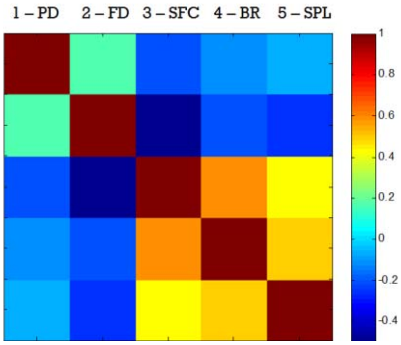

Fig. 12. Correlations of physiological eye indices.

The physiological patterns of the eyes of the subjects were represented by the average pupil diameter (PDave) and the average blink rates (BRave). PD is indicative of how engaged a subject is and literature suggests that there are variations of these indices among individuals with ASD [24] given the same stimuli. Individuals with autism were shown to have abnormal eye blink conditioning compared [59]. Generally, children in the gaze group exhibited lower pupil diameter and blink rates in all the sessions.

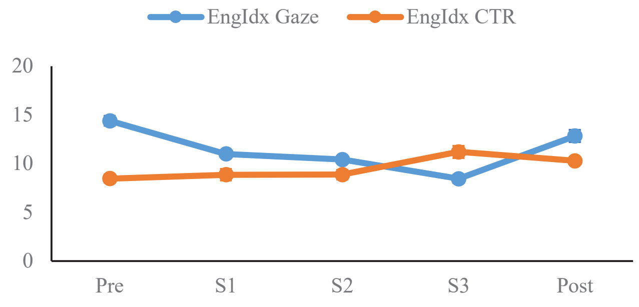

Fig. 13. Combined engagement index.

Fig. 13 shows that with the combined engagement index the gaze group engagement was slightly lower in the post-test compared to the pre-test and the control group was the opposite. However, the system was able to maintain engagement in both groups across all the sessions and ending up with almost similar pattern of engagement in the post-test with the gaze group with higher engagement than the control group.

4.4 Physiological Features Analysis

The physiological data of the children in both groups were processed and five features were extracted as described in Section 2 in the data analysis section. These were the heart rate (HR), the skin temperature (SKT), the respiration rate (RSPR), the galvanic skin conductance rate (SCR), and the skin conductance level (SCL).

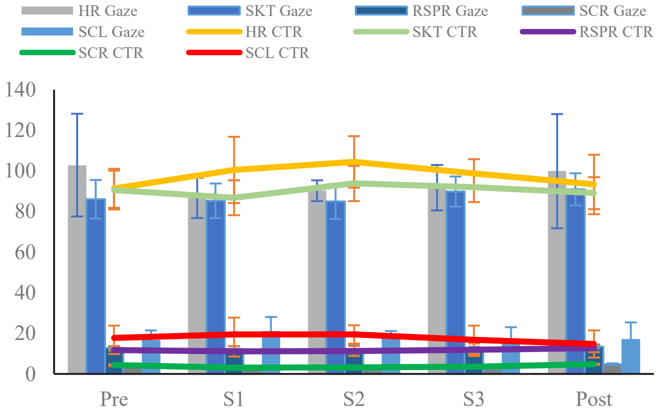

Fig. 14. Comparisons of physiological features.

Fig. 14 shows that the general trend in both groups (the bars are for the gaze group whereas the line plots are for the control group for clarity) decreased activity from pre-to-post with the control group slightly higher than the gaze group. However, none of these changes were statistically significant. Lower heart rate, lower skin conductance and respiration rate are all indicative of lower emotion reflection activity. These seem to be consistent with the gaze features as well.

4.5 EEG Features Analysis

EEG features extracted from each trial were divided by corresponding baseline measures to remove individual variations. Table II listed the mean and standard deviation of each individual feature for gaze and control group during pre and post-tests. For gaze group, the values of all the features decreased. For control group, all but right parietal gamma increased. Right parietal theta was significantly different (t-test, P < 0.05) between two groups for pre, while for post right parietal alpha2 was significantly different (t-test, P < 0.05). Within the control group, right parietal theta increased significantly from pre to post (t-test, P < 0.05). Since the correlation coefficients among six features were all positive, we normalized each feature to range [0, 1] and combined them together as one single feature. Fig. 15 shows the change of the combined feature for the five sessions. From pre to post, the combined feature decreased for gaze group and increased for control group. Even though the gaze group had higher feature value for pre, it was not statistically significant. After training with social task, the gaze group had lower feature value and it was statistically significant (t-test, P < 0.05). As these features relate to the subjects ability to process emotion and cognitive task load, the gaze group was found to be less engaged in the emotion recognition mental process. This result also goes with the eye tracking features in which the gaze group decreased engagement from pre to post-test although the change was not as significant.

Table II EEG features for the VR session

			RPT	LAT2	AA2	RPA2	AB1	RPG
Pre	Gaze	M	0.84	0.83	0.88	0.91	0.94	0.96
		SD	0.13	0.17	0.15	0.10	0.23	0.23
	CTR	M	0.65	0.70	0.84	0.81	0.80	1.16
		SD	0.12	0.23	0.35	0.24	0.15	0.87
Post	Gaze	M	0.61	0.76	0.76	0.67	0.86	0.85
		SD	0.24	0.23	0.15	0.30	0.09	0.23
	CTR	M	0.90	1.18	1.31	1.20	0.99	1.01
		SD	0.24	0.66	0.97	0.46	0.17	0.24

RPT: Right parietal theta, LAT2: Left anterior theta2, AA2: Anterior alpha2, RPA2: Right parietal alpha2, AB1: Anterior beta1, RPG: Right Parietal Gamma

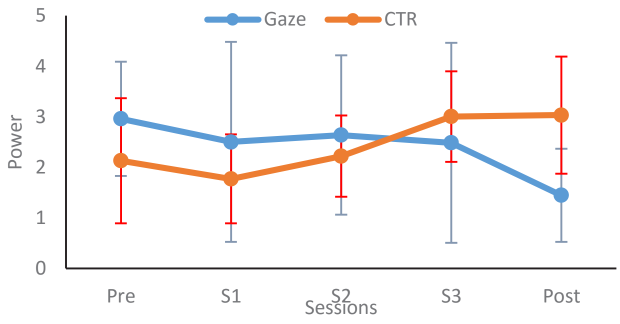

Fig. 15. Combined EEG feature

5 Discussion and Conclusion

In this paper, we have presented the design, development and usability study of a multimodal adaptive social interaction VR environment (MASI-VR). The system was able to collect eye tracking, peripheral psychophysiological and EEG data while the subjects were involved in the emotion recognition training and pre and post isolated emotion recognition tasks over 3 visits over a period of time. Such capabilities are expected to be useful in understanding the underlying deficits individuals with ASD and, in turn, will hopefully help developing new intervention paradigms to improve such impairments. The 3 social task training sessions in between the pre and post-tests were specifically designed to allow the training of proper gaze processing pattern with an online adaptive and individualized gaze feedback mechanism. Moreover, it also incorporated a spoken conversational dialog management for more natural social emotion recognition. These two are the pillars of the system in an effort to teach children with ASD the ability to recognize faces in a social context. A usability study involving 12 children with ASD was performed to evaluate the efficacy of the system as well as to study behavioural and physiological pattern differences. Half of them were randomly assigned to the gaze-sensitive part of the system while the remaining half were used as controls without the occlusion paradigm and hence without the online gaze feedback mechanism to evaluate the effectiveness of the dynamic system as a whole.

The system successfully performed the adaptive social task as well as the pre and post isolated facial expression tasks and collected the synchronized eye gaze, EEG and physiological data.

The results of the usability study indicated that the gaze-sensitive system enabled the gaze group to close the performance gap that existed in the pre-test with the control group by more than 3% points while maintaining relatively similar engagement as depicted by the various modalities with that of the control group. There were several limitations of the system in its current form. Although the system was equipped with measuring EEG as well as physiology signals, only the gaze was used for online adaptation. The gaze group subjects had difficulty in understanding the occlusion paradigm at first as explicit guidance was not given so as to not bias the outcome. However, over the course of the 3 visit social task training, they picked up the importance of the occlusion and what they are supposed to do to reveal the faces. This initial confusion might have contributed to the eventual non-pronounced engagement differences.

Despite these limitations, the system proved that the controllability, ease of interaction without information overload and the game nature of the interaction were useful in training core deficit areas of children with ASD for eventual better social functioning. These preliminary findings will be used to build a more robust adaptive VR-based social interactive environment that enables online adaptation not only by gaze but also using all the multimodal inputs using some form of decision level fusion for children with ASD to improve their emotion recognition abilities and eventual social functioning.

Acknowledgments

We would like to thank all colleagues that helped in evaluating system. Specifically, we would like to acknowledge Amy Swanson, Amy Weitlauf and Nicole Bardett for their help in participant recruitment and serving as therapist for monitoring the experimental sessions. We give special thanks to all participants and their family.

References

[1] *Diagnostic and Statistical Manual of Mental Disorders: Quick reference to the diagnostic criteria from DSM-IV-TR*, Washington, DC: American Psychiatric Association, Amer Psychiatric Pub Incorporated, 2000.

[2] R. E. McEvoy, S. J. Rogers, and B. F. Pennington, "Executive function and social communication deficits in young autistic children," *Journal of Child Psychology and Psychiatry,* vol. 34, no. 4, pp. 563-578, 2006.

[3] C. Demopoulos, J. Hopkins, and A. Davis, "A Comparison of Social Cognitive Profiles in children with Autism Spectrum Disorders and Attention-Deficit/Hyperactivity Disorder: A Matter of Quantitative but not Qualitative Difference?," *Journal of autism and developmental disorders*, pp. 1-14, 2012.

[4] B. O. Ploog, A. Scharf, D. Nelson *et al.*, "Use of Computer-Assisted Technologies (CAT) to Enhance Social, Communicative, and Language Development in Children with Autism Spectrum Disorders," *Journal of autism and developmental disorders*, pp. 1-22, 2012.

[5] R. Adolphs, L. Sears, and J. Piven, "Abnormal processing of social information from faces in autism," *Journal of Cognitive neuroscience,* vol. 13, no. 2, pp. 232-240, 2001.

[6] F. Castelli, "Understanding emotions from standardized facial expressions in autism and normal development," *Autism,* vol. 9, no. 4, pp. 428-449, 2005.

[7] G. Celani, M. W. Battacchi, and L. Arcidiacono, "The understanding of the emotional meaning of facial expressions in people with autism," *Journal of autism and developmental disorders,* vol. 29, no. 1, pp. 57-66, 1999.

[8] G. Dawson, S. J. Webb, and J. McPartland, "Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies," *Developmental neuropsychology,* vol. 27, no. 3, pp. 403-424, 2005.

[9] F. Gosselin, and P. G. Schyns, "Bubbles: a technique to reveal the use of information in recognition tasks," *Vision research,* vol. 41, no. 17, pp. 2261-2271, 2001.

[10] A. T. Wang, M. Dapretto, A. R. Hariri *et al.*, "Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder," *Journal of the American Academy of Child & Adolescent Psychiatry,* vol. 43, no. 4, pp. 481-490, 2004.

[11] S. J. Weeks, and R. P. Hobson, "The salience of facial expression for autistic children," *Journal of Child Psychology and Psychiatry,* vol. 28, no. 1, pp. 137-152, 1987.

[12] R. P. Hobson, "The autistic child's appraisal of expressions of emotion," *Journal of Child Psychology and Psychiatry,* vol. 27, no. 3, pp. 321-342, 1986.

[13] N. Bauminger, "The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes," *Journal of autism and developmental disorders,* vol. 32, no. 4, pp. 283-298, 2002.

[14] G. S. Chasson, G. E. Harris, and W. J. Neely, "Cost comparison of early intensive behavioral intervention and special education for children with autism," *Journal of Child and Family Studies,* vol. 16, no. 3, pp. 401-413, 2007.

[15] M. L. Ganz, "The lifetime distribution of the incremental societal costs of autism," *Archives of Pediatrics and Adolescent Medicine, Am Med Assoc,* vol. 161, no. 4, pp. 343-349, 2007.

[16] M. L. Ganz, "The costs of autism," CRC Press, New York, 2006, pp. 475-502.

[17] M. S. Goodwin, "Enhancing and Accelerating the Pace of Autism Research and Treatment," *Focus on Autism and Other Developmental Disabilities,* vol. 23, no. 2, pp. 125-128, 2008.

[18] E. Bekele, U. Lahiri, J. Davidson *et al.*, "Development of a novel robot-mediated adaptive response system for joint attention task for children with autism," in 20th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN, , Atlanta, GA, 2011, pp. 276-281.

[19] E. T. Bekele, U. Lahiri, A. R. Swanson *et al.*, "A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 21, no. 2, pp. 289-299, 2013.

[20] U. Lahiri, Welch, Karla Conn, Warren, Zachary, Sarkar, Nilanjan, "Understanding psychophysiological response to a Virtual Reality-based social communication system for children with ASD," in International Conference on Virtual Rehabilitation (ICVR), 2011 Zurich, Switzerland, 2011, pp. 1-2.

[21] K. Welch, Lahiri, U., Liu, C., Weller, R., Sarkar, N., Warren, Z., "An Affect-Sensitive Social Interaction Paradigm Utilizing Virtual Reality Environments for Autism Intervention," in Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction, 2009, pp. 703-712.

[22] M. R. Kandalaft, N. Didehbani, D. C. Krawczyk *et al.*, "Virtual Reality Social Cognition Training for Young Adults with High-Functioning Autism," *Journal of autism and developmental disorders*, pp. 1-11, 2012.

[23] U. Lahiri, Warren, Z., Sarkar, N., "Dynamic gaze measurement with adaptive response technology in Virtual Reality based social communication for autism," in 2011 International Conference on Virtual Rehabilitation (ICVR), 2011, pp. 1-8.

[24] U. Lahiri, Z. Warren, and N. Sarkar, "Design of a Gaze-Sensitive Virtual Social Interactive System for Children With Autism," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, no. 99, pp. 1-1, 2012.

[25] P. J. Standen, Brown, D.J., "Virtual reality in the rehabilitation of people with intellectual disabilities: review," *Cyberpsychology & behavior,* vol. 8, no. 3, pp. 272-282, 2005.

[26] A. Tartaro, Cassell, J., "Using virtual peer technology as an intervention for children with autism," in Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations. , Chichester, UK, , 2006, pp. 231-262.

[27] S. Parsons, and S. Cobb, “State-of-the-art of Virtual Reality technologies for children on the autism spectrum,” *European Journal of Special Needs Education,* vol. 26, no. 3, pp. 355-366, 2011.

[28] N. Josman, H. M. Ben-Chaim, S. Friedrich *et al.*, “Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism,” *International Journal on Disability and Human Development,* vol. 7, no. 1, pp. 49-56, 2011.

[29] S. Parsons, Mitchell, P., “The potential of virtual reality in social skills training for people with autistic spectrum disorders,” *Journal of Intellectual Disability Research,* vol. 46, no. 5, pp. 430-443, 2002.

[30] P. Kenny, T. Parsons, J. Gratch *et al.*, "Virtual patients for clinical therapist skills training." pp. 197-210.

[31] P. Mitchell, Parsons, S., Leonard, A., “Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders,” *Journal of autism and developmental disorders,* vol. 37, no. 3, pp. 589-600, 2007.

[32] C. Liu, Conn, K., Sarkar, N., Stone, W., “Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder,” *International Journal of Human-Computer Studies, Elsevier,* vol. 66, no. 9, pp. 662-677, 2008.

[33] C. Liu, Conn, K., Sarkar, N., Stone, W., “Online affect detection and robot behavior adaptation for intervention of children with autism,” *Robotics, IEEE Transactions on,* vol. 24, no. 4, pp. 883 - 896, 2008.

[34] K. C. Welch, Lahiri, U., Sarkar, N., Warren, Z., Stone, W., Liu, C., "Affect-Sensitive Computing and Autism," *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives*, G. Y. Didem Gökçay, ed., pp. 325-343: Information Science Reference, 2010.

[35] E. Bekele, Z. Zheng, A. Swanson *et al.*, “Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments,” *Visualization and Computer Graphics, IEEE Transactions on,* vol. 19, no. 4, pp. 711-720, 2013.

[36] M. Wang, and D. Reid, “Virtual Reality in Pediatric Neurorehabilitation: Attention Deficit Hyperactivity Disorder, Autism and Cerebral Palsy,” *Neuroepidemiology,* vol. 36, no. 1, pp. 2-18, 2011.

[37] E. Gal, N. Bauminger, D. Goren-Bar *et al.*, “Enhancing social communication of children with high-functioning autism through a co-located interface,” *AI & Society,* vol. 24, no. 1, pp. 75-84, 2009.

[38] N. Bauminger, D. Goren-Bar, E. Gal *et al.*, “Enhancing social communication in high-functioning children with autism through a co-located interface,” in Multimedia Signal Processing. IEEE 9th Workshop on, 2007, pp. 18-21.

[39] S. Parsons, Mitchell, P., Leonard, A., “The use and understanding of virtual environments by adolescents with autistic spectrum disorders,” *Journal of Autism and Developmental Disorders,* vol. 34, no. 4, pp. 449-466, 2004.

[40] S. Parsons, Mitchell, P., Leonard, A., “Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments?,” *Autism,* vol. 9, no. 1, pp. 95-117, 2005.

[41] T. D. Parsons, A. A. Rizzo, S. Rogers *et al.*, “Virtual reality in paediatric rehabilitation: A review,” *Developmental Neurorehabilitation,* vol. 12, no. 4, pp. 224-238, 2009.

[42] L. A. Ruble, and D. M. Robson, “Individual and environmental determinants of engagement in autism,” *Journal of autism and developmental disorders,* vol. 37, no. 8, pp. 1457-1468, 2007.

[43] P. Kenny, T. Parsons, J. Gratch *et al.*, “Virtual patients for clinical therapist skills training,” in Intelligent Virtual Agents, 2007, pp. 197-210.

[44] A. Leuski, R. Patel, D. Traum *et al.*, "Building effective question answering characters." pp. 18-27.

[45] P. Ekman, “Facial expression and emotion,” *American Psychologist,* vol. 48, no. 4, pp. 384, 1993.

[46] E. Bekele, J. W. Wade, D. Bian *et al.*, "Multimodal Interfaces and Sensory Fusion in VR for Social Interactions," *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, pp. 14-24: Springer, 2014.

[47] R. W. Picard, “Future affective technology for autism and emotion communication,” *Philosophical Transactions of the Royal Society B: Biological Sciences,* vol. 364, no. 1535, pp. 3575-3584, 2009.

[48] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 23, no. 10, pp. 1175-1191, 2001.

[49] H. H. Jasper, “The ten twenty electrode system of the international federation,” *Electroencephalography and Clinical Neurophysiology,* vol. 10, pp. 371-375, 1958, 1958.

[50] L. Aftanas, N. Reva, A. Varlamov *et al.*, “Analysis of evoked EEG synchronization and desynchronization in conditions of emotional activation in humans: temporal and topographic characteristics,” *Neuroscience and behavioral physiology,* vol. 34, no. 8, pp. 859-867, 2004.

[51] A. Keil, M. M. Müller, T. Gruber *et al.*, “Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials,” *Clinical neurophysiology,* vol. 112, no. 11, pp. 2057-2068, 2001.

[52] R. J. K. Jacob, “Eye tracking in advanced interface design,” *Virtual environments and advanced interface design*, pp. 258-288, 1995.

[53] H. Shiffman, "Fundamental visual functions and phenomena " *sensation and perception: An Integrated Approach*, pp. 89-115: John Welsly and Sons, New York, 2001.

[54] J. Constantino, and C. Gruber, “The social responsiveness scale,” *Los Angeles: Western Psychological Services*, 2002.

[55] M. Rutter, A. Bailey, C. Lord *et al.*, “Social communication questionnaire,” *Los Angeles, CA: Western Psychological Services*, 2003.

[56] D. Wechsler, *Wechsler Abbreviated Scale of Intelligence® – Fourth Edition (WASI®-IV)*, San Antonio, TX: Harcourt Assessment, The Psychological Corporation, 2008.

[57] A. Klin, W. Jones, R. Schultz *et al.*, “Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism,” *Archives of general psychiatry,* vol. 59, no. 9, pp. 809, 2002.

[58] J. W. Denver, “The social engagement system: Functional differences in individuals with autism,” *Cerebral Cortex,* vol. 16, no. 9, pp. 1276-1282, 2004.

[59] L. L. Sears, P. R. Finn, and J. E. Steinmetz, “Abnormal classical eye-blink conditioning in autism,” *Journal of autism and developmental disorders,* vol. 24, no. 6, pp. 737-751, 1994.