VIRTUAL-REALITY BASED GAZE-SENSITIVE ADAPTIVE RESPONSE TECHNOLOGY FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

By

Uttama Lahiri

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Mechanical Engineering

May, 2011

Nashville, Tennessee

Approved:

Dr. Nilanjan Sarkar, Chair

Dr. Zachary Warren

Dr. Carol Rubin

Dr. Mitch Wilkes

Dr. Robert Webster

To my source of Inspiration, My parents

ACKNOWLEDGEMENTS

Many thanks are extended to my dissertation committee for their support. Particularly my advisor, Dr. Nilanjan Sarkar, has been a great mentor and is perpetually prepared with guidance. This work would not have been possible without the commitment from Dr. Zachary Warren to join in this cross-disciplinary project as well as his continued guidance of the work. Also, I thank Dr. Carol Rubin, Dr. Mitch Wilkes, and Dr. Robert Webster for their helpful suggestions that substantially improved this research. I am grateful for the Autism Speaks Pilot grant, National Science Foundation Grant, National Institute of Health Grant, and Vanderbilt Institute for Clinical/Translational Research grant that financially supported this work.

I would like to thank Gary Walker, Robin Midgett, and Tony Costi for their technical support throughout the user studies. Also, many thanks go to Elizabeth Howell Dohrmann for her valuable contribution in this work and her serving as a reliable clinical observer. Working with my fellow lab members of the Robotics and Autonomous Systems Laboratory has been both enjoyable and educational. In particular, I thank Karla Conn Welch, Changchun Liu, and Esubalew Tamirat Bekele for their hard work and support.

Mostly, I thank my mother for her inspiration and her patient support through both bright and dark days of this journey. I could not have accomplished this goal without her encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADI-R	Autism Diagnostic Interview-Revised
ADOS-G	Autism Diagnostic Observation Schedule-Generic
ASD	Autism Spectrum Disorder
BR	Blink Rate
BVP	Blood Volume Pulse
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders-4th edition
ECG	Electrocardiogram
EDA	Electrodermal Activities
EMG	Electromyogram
FNFR	Face-to-nonFace Ratio
GSR	Galvanic Skin Response
HCI	Human-computer Interaction
HRI	Human-robot Interaction
IBI	Inter Beat Interval
ICG	Impedance Cardiogram
IRB	Institutional Review Board
OFR	Object-to-Face Ratio
PCG	Phonocardiogram
PD	Pupil Diameter
PEP	Pre-ejection Period
PPG	Photoplethysmogram
PPVT	Peabody Picture Vocabulary Test

PTT	Pulse Transit Time
ROI	Region of Interest
SCQ	Social Communication Questionnaire
SD	Standard Deviation
SRS	Social Responsiveness Scale
SVM	Support Vector Machines
VR	Virtual Reality

CHAPTER I

INTRODUCTION

Introduction

This dissertation presents the development and application of virtual-reality based gaze-sensitive system with adaptive response technology for children with Autism Spectrum Disorder (ASD). Such a system can intelligently adapt itself in an individualized manner to encourage a child to participate in social communication tasks while trying to improve his/her level of engagement and performance in the social task. Children with ASD are characterized by core deficits in social interaction and communication accompanied by restricted patterns of interest and behavior (APA, 2000), infrequent engagement in social interactions (APA, 1994), atypicalities surrounding eye-gaze and social information processing (Rutherford, and Towns, 2008; Jones, Carr, and Klin, 2008), and impaired understanding of mental states of others (Baron-Cohen, 1997; Frith, and Frith, 1999). Clinicians[1] involved in interventions must overcome these communication impairments generally exhibited by children with ASD by adeptly inferring the affective cues of the children to adjust the intervention accordingly. There is growing consensus that appropriately individualized intensive behavioral and educational interventions can improve core social communication vulnerabilities seen in individuals with ASD (NRC, 2001). However, there are potent barriers related to accessing and implementing appropriately individualized intensive intervention services such as limited

[1] We use the terms "clinician," "clinical observer," and "therapist" interchangeably to mean an expert with skill in making judgments, such as rating affective states, about the meaning of observable behaviors from individuals with autism.

access to and availability of appropriately trained professionals, lack of available data suggesting which interventions will work better for specific children, and exorbitant costs (Ganz, 2007; Goodwin, 2008). Given these barriers, researchers are now employing technology to develop more accessible, quantifiable, intensive and individualized interventions for core deficit areas related to ASD (Goodwin, 2008). Thus development of an intelligent system with an ability to objectively identify the affective and attentive states of the children with ASD and adapt itself targeted to the specific child is critical. This can pave the way for the development of an individualized, intensive, and cost-effective ASD intervention tool.

Even though there is increasing research in technology-assisted autism intervention, there is a paucity of published studies that specifically address how to automatically detect and respond to affective and attentive cues of children with ASD. The currently available systems as applied to tasks involving children with ASD are capable of modifying tasks based only on objective performance characteristics (i.e., correct or incorrect) of responses (Parsons et al., 2004; Strickland et al., 1996). Though being able to adapt tasks based on performance is an important aspect of potential intervention systems for children with ASD, such adaptation based solely on task performance limits the individualization of application and likely potential generalization of skills. Specifically, performance based social communication skill-training tasks do not often involve measurements of or necessitate appropriate subtle, yet critically important, aspects of effective social communication (e.g., such as eye-gaze, and other forms of social convention). In fact, while many children with ASD are capable of yielding correct performance on objective tasks measures, it is their vulnerabilities surrounding elements of social communication that is so closely tied to their functional social impairments. Thus to foster effective social communication tasks, the system should be capable of intelligently responding to the subtle aspects of social communication to engage the child in the social task through a high degree of individualization. We believe that such ability could be critical given the importance of affective information in human-computer interaction (Picard, 1997) and the significant impacts of the affective (Ernsperger, 2003; Seip, 1996; Wieder, and Greenspan, 2005), attentive (Rutherford, and Towns, 2008; Jones, Carr, and Klin, 2008), and task performance (Blackorby, and Cameto, 2005) factors of children with ASD on the intervention practice.

Thus, there is a need to develop a technologically-advanced social interactive system capable of automatic detection of affective and attentive states and adapting itself to address some of the core social vulnerabilities of these children in an individual-specific manner. Motivated by this need to develop a system that can objectively identify one's attentive indices and provide individualized services, our ongoing research has demonstrated the feasibility of Virtual Reality (VR) based social interaction to elicit variations in the attentive indices of the children with ASD. Also, these indices can be correlated to the affective state that underlies the presumed core social impairments associated with ASD. The work presented in this dissertation utilizes and merges (i) the technological advances in the area of virtual reality, (ii) dynamic eye-gaze tracking and (iii) intelligent adaptive response technology with an aim to provide a technology-based tool that can intelligently adapt itself in an individualized manner to encourage a child to engage in social communication task. In addition, this would also help us to better understand the underlying affective and attentive mechanisms associated with some of the core social vulnerabilities of children with ASD.

The research work presented in this dissertation utilizes the attentive factors, namely, the behavioral viewing patterns, and eye physiological parameters, and the performance metric of an individual, to achieve the primary objective of developing technology-based assessment tools capable of identifying specific aspects of interaction that induce an affective (e.g., engagement) response in individuals with ASD. Additionally, the presented system is capable of adaptively responding to the engagement level as predicted from the behavioral viewing pattern, eye physiological indices, and performance of a child with ASD during social interaction of the child with the VR-based system. We use engagement as the target affective state, because, engagement, defined as ‘‘sustained attention to an activity or person’’ (NRC, 2001), is one of the key factors for children with ASD to make substantial gains in communication and social domains (Ruble, and Robson, 2006). Infrequent or no engagement in social interaction is one of the defining characteristics of ASD (APA, 1994). The engagement of children with ASD is the ground basis for the 'floor-time-therapy' to help them develop relationships and improve their social skills (Wieder and Greenspan, 2005). Thus, if we can engage these children to a social task, then we can teach them social skills. The behavioral viewing patterns speak of one’s attention and interest in a target (Denver, 2004; Poole, and Ball, 2005; Just, and Carpenter, 1976) and children with ASD often demonstrate atypical viewing patterns by attending more towards non-human objects than the human faces during social interaction (Anderson, Colombo, and Shaddy, 2006). In addition, eye physiology based methodologies (Libby, Lacey, and Lacey, 1973; Partala, and Surakka, 2003) have compelling advantages over other observational modalities (e.g., facial expression, vocal intonation, or gesture) in evaluating the affective responses of children with ASD, since they permit continuous gathering of rich data in the face of potential communicative limitations of these children, particularly regarding expression of affective states. Further, we also consider the participant's task performance, because, clinicians involved in ASD intervention, often look out for the task performance metric which is positively correlated to the participant's engagement level (Blackorby, and Cameto, 2005). The presented system can be employed to develop new intervention paradigms, which can promote interventions for individuals with ASD that are practical, widely available, and specific to the unique strengths and vulnerabilities of individuals with ASD. Thus this can serve as a valuable tool which can provide important information to caregivers and clinicians. Also, it can be utilized to adaptively drive behavioral interventions in an individualized manner towards achieving realistic social interaction to challenge, and expectantly promote scaffolded skill development in particular areas of vulnerability while improving the engagement level and the task performance of these children. Additionally, the presented technology with a behavioral engagement profiling system is capable of adapting to one's predicted engagement level in controlled environments and thereby reinforcing skills in core domains gradually but automatically, which can prove an effective tool for developing tailored interventions for individuals with ASD. The research work presented here has the following two objectives:

- **Objective 1: To design and evaluate a VR-based gaze-sensitive social interactive system capable of delivering individualized feedback based on one's dynamic viewing patterns**

We plan to design social interaction modules on a VR platform. These are to be integrated to computationally-enhanced eye-tracker to provide individualized feedback. Specifically, the designed VR-based gaze-sensitive system will be capable of quantifying eye-gaze patterns of a child with ASD detected in real-time during virtual social interaction and utilizing this data to provide specific feedback aimed at altering viewing patterns (e.g., fixation counts, fixation duration, face-to-nonface ratio, and object-to-face ratio) at each instant of time. Also the developed system would be capable of communicating some of these indices to the participant at a preferred time as the task proceeds depending on the study design.

We plan to investigate the effectiveness of the VR-based gaze-sensitive social interactive system to elicit variations in the participants' behavioral viewing patterns, scanning patterns of the visual stimulus, and the engagement level, measured by ratings from the observers, during virtual social interaction as a result of the individualized feedback. Further, we plan to evaluate (in an off-line manner) the potential of such a system to have an impact on the participants' eye physiological indices (e.g., blink rate, pupil diameter) while recognizing emotions of their virtual peers (i.e., the avatars).

- **Objective 2: To enhance the developed VR-based gaze-sensitive social interactive system with adaptive response technology based on one's behavioral viewing, eye physiological indices, and performance metrics**

Our aim is to enhance the system (as mentioned in Objective 1) by designing a VR-based gaze-sensitive system with adaptive response technology which can be applied to social communication task for children with ASD. We plan to formulate the system to present VR-based social tasks coupled with gaze-sensitive feature to monitor the behavioral viewing and eye physiological indices of the participants in real-time, as they interact with the virtual social scenarios. Additionally, this system will feature bidirectional interaction in the form of social conversations between a participant and his/her virtual peer. Also, we plan to develop VR-based social tasks equipped with varying degrees of task difficulty (e.g., Low, Medium, and High) for social communication between the participant and his/her virtual peer. The system will monitor the performance of a participant while he/she interacts with the system using the bidirectional social conversation module of different degrees of task difficulty. Based on the participant's behavioral viewing, eye physiological indices, and performance metric, the system will adaptively and socially respond by using a rule-governed strategy generator.

We will assess the potential of the designed social interactive system using the rule-governed strategy generator to adaptively respond and encourage a participant to continue virtual social interaction. The rule-based strategy generator will fuse the participant's behavioral viewing, eye physiological indices, and the performance metric to implement an individualized task modification strategy. We plan to investigate the ability of the strategy generator to enhance the participant's performance (e.g., whether the participant's task performance improves on interacting socially with tasks of higher degree of interaction difficulty) via adaptively modifying the task difficulty (i.e., increasing/decreasing).

The dissertation is organized as follows: The motivation for the present research work is briefly discussed in Chapter II. Chapter III presents the design and development of VR-based social communication task and its feasibility to influence one's peripheral physiological signals and affective states (e.g., engagement, enjoyment/liking, and anxiety). In addition, this presents the mapping of one's physiological responses with the affective states while an individual participates in a VR-based social communication task. In Chapter IV, the design and development of a VR-based social communication system seamlessly integrated with technologically-enhanced eye-tracking technology is presented. The system is capable of computing one's real-time behavioral viewing patterns during social communication and thereby delivering individualized feedback. Here, the impact of the individualized feedback on one's behavioral viewing patterns has also been investigated. Chapter V elaborates on the detailed design specifications of the VR-based gaze-sensitive system along with bidirectional conversation module and adaptive response technology. Also this describes the rationale behind the rule-governed strategy generator that administers the dynamic switching of the social communication tasks. Chapter VI describes the design of the usability study to demonstrate the feasibility of such a system. Also, this presents the implication of such a VR-based gaze-sensitive system that adaptively responds based on the composite effect of one's real-time behavioral viewing, eye physiology and performance metric. Specifically, this describes the effect of interaction with such a system, on one's engagement level and performance while participating in the VR-based social communication task. Chapter VII shows the efficacy of such a system to influence the physiological signals, whether it is peripheral physiology or the eye physiology of the participants while they interact with the VR-based social situations. In addition, this presents the correlation of the physiological signals with the affective state of the participants as rated by the clinical observer/therapist. Finally, chapter VIII summarizes the contributions of the present work and describes the scope for future work.

References

Anderson, C.J., Colombo, J., and Shaddy, D.J. "Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder," J. of Cli. and Exp. Neuropsy., vol. 28, pp. 1238–1256, 2006.

APA, "Diagnostic and statistical manual of mental disorders," (4th ed.), Washington, DC: American Psychiatric Association, 1994.

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

Baron-Cohen, S. "Mindblindness: An essay on autism and theory of mind," Cambridge: MIT, 1997.

Blackorby, J. and Cameto, R. "Changes in the School Engagement and Academic Performance of Students with Disabilities," in Engagement and Academics, pp. 1-24, 2005.

Denver, J.W. "The Social Engagement System : Functional Differences in Individuals with Autism," Dissertation, 2004. *www.lib.umd.edu/drum/bitstream/1903/1351/1/umi-umd-1486.pdf*

Ernsperger, L. "Keys to Success for Teaching Students with Autism," Future Horizons, 2003.

Frith, C. D., and Frith, U. "Interacting minds—a biological basis," Science, vol. 286, no. 5445, pp. 1692–1695, 1999.

Goodwin, M.S. "Enhancing and accelerating the pace of Autism Research and Treatment: The promise of developing Innovative Technology," Focus on aut. and other dev. dis., vol. 23, pp. 125-128, 2008.

Jones, W., Carr, K., and Klin, A. "Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder," Arch. Gen. Psychiatry, vol. 65, no. 8, pp. 946-954, 2008.

Just, M.A., and Carpenter, P.A. "Eye fixations and cognitive processes," Cognitive Psychology, vol. 8, pp. 441-480, 1976.

Libby, W.L., Lacey, B.C., and Lacey, J.I. "Pupillary and cardiac activity during visual attention," Psychophysiology, vol. 10, no. 3, pp. 270-294, 1973.

NRC. (2001). Educating children with autism. National Academy Press, Washington, DC.

Partala, T., and Surakka, V. "Pupil size variation as an indication of affective processing," Int. J. Human-Computer Studies, vol. 59, pp. 185-198, 2003.

Picard, R. W. "Affective Computing," Cambridge, MA: MIT Press. 1997.

Poole, A., and Ball, L.J. "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects," In C. Ghaoui (Ed.), Encyclopedia of Human Computer Interaction, Idea Group, 2005.

Ruble, L.A., and Robson, D.M. "Individual and environmental determinants of engagement in autism," J. of Aut. and Dev. Dis., vol. 37, no. 8, pp. 1457–1468, 2006.

Rutherford, M.D., and Towns, M.T. "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.

Seip, J. "Teaching the Autistic and Developmentally Delayed: A Guide for Staff Training and Development," Delta, BC: Author, 1996.

Wieder, S., and Greenspan, S. "Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective?," The Journal of Developmental and Learning Disorders, vol. 9, 2005.

CHAPTER II

SIGNIFICANCE AND BACKGROUND

Significance

Emerging research suggests prevalence rates in the United States recorded as high as approximately 1 in 110 for the broad autism spectrum (CDC, 2009). Impairments in social communication skills are thought to be core deficits in children with ASD (APA, 2000). Specifically, these children demonstrate atypical viewing patterns in part characterized by greater fixation towards non-social objects than faces of individuals during social communication. These are thought to contribute to difficulties in social interaction, including difficulties reading others' nonverbal emotional cues. To understand the social communication vulnerabilities of individuals with ASD, research has examined how they process salient social cues, specifically from faces (Rutherford, and Towns, 2008; Jones, Carr, and Klin, 2008). The ability to derive socially relevant information from faces is thought to be a fundamental skill for facilitating reciprocal social interactions (Trepagnier, Sebrechts, and Peterson, 2002) and an early deficit may contribute in part to the developmental cascade associated with core vulnerabilities of the disorder (Dawson, 2008). As children with ASD show (Baron-Cohen et. al., 1999; Carpenter, Pennington, and Rogers, 2002) difficulties in social judgment (e.g., deciding on appropriate social behaviors, understanding others' emotions, etc.), attenuated attention with increased engagement in atypical behavior and non-social tasks (McGee, Feldman, and Morrier, 1997; Sigman, and Ruskin 1999), focus in autism research has been to devise affect-sensitive interactive techniques to address some of the core deficits of these children in communication and social domains.

While there is at present no single accepted intervention, treatment, or known cure for ASD, there is growing consensus that intensive behavioral and educational intervention programs can significantly improve long-term outcomes for individuals and their families (Cohen, Amerine-Dickens, and Smith, 2006; Rogers, 1998; NRC, 2001). In response to this need, a growing number of studies have been investigating the application of advanced interactive technologies to address core deficits related to autism, namely computer technology (Bernard-Opitz, Sriram, and Nakhoda-Sapuan, 2001; Moore, McGrath, and Thorpe, 2000; Swettenham, 1996), VR environments (Parsons, Mitchell, and Leonard, 2004; Strickland et al., 1996; Tartaro, and Cassell, 2007), and robotic systems (Dautenhahn, and Werry, 2004; Kozima, Nakagawa, and Yasuda, 2005; Michaud, and Theberge-Turmel, 2002; Pioggia et al., 2005; Scassellati, 2005). Computer- and VR-based intervention may provide a simplified but exploratory interaction environment for children with ASD (Moore, McGrath, and Thorpe, 2000; Parsons, Mitchell, and Leonard, 2004; Strickland et al., 1996).

A computer that can detect the affective states of a child with ASD and interact with him/her based on such perception could have a wide range of potential impacts. Interesting activities likely to retain the child's attention could be chosen when a low level of engagement is detected. The engagement of children with ASD is the ground basis for the 'floor-time-therapy' to help them develop relationships and improve their social skills (Wieder and Greenspan, 2005). Clinicians who work with children in autism intervention intensely monitor affective cues, e.g., engagement in order to make appropriate decisions about adaptations to their intervention and reinforcement strategies.

Thus, allowing a computer to recognize the engagement level of a child in terms of his/her performance, behavioral viewing pattern, and eye physiological indices during social tasks and applying this information as a means of taking appropriate decisions about the adaptation of the child to the intervention may be important. Complex social stimuli, sophisticated interactions, and unpredictable situations could be gradually, but automatically, introduced when the computer recognizes that the child is engaged at a certain level of interaction dynamics for a reasonably long period of time. A clinician could use the history of the child's affective information to analyze the effects of the intervention approach. With the record of the activities and the consequent emotional changes in a child, a computer could learn individual preferences and affective characteristics over time and thus could alter the manner in which it responds to the needs of different children.

The current research as presented in this dissertation describes development of a gaze-sensitive virtual interactive platform that can dynamically adapt itself based on an individual's engagement level predicted by the performance metric, real-time behavioral viewing pattern and eye physiological indices during a child's virtual socially-oriented tasks. In addition, we assess the effectiveness of this system with adaptive response technology to enhance the child's performance (e.g., the participant's performance improves on interacting socially with tasks of higher degree of interaction difficulty) with improved engagement to the social interaction tasks. Thus, this will provide an integrated computer and eye physiological profiling system which would serve as a tool for designing intervention strategies. In the future, such an integrated intelligent system could be effective for use in developing adaptive controlled environments that can systematically manipulate various aspects of social communication and thereby help individuals to explore social interaction dynamics gradually and automatically.

Background

Use of Eye Physiology for Affect Recognition of Children with ASD

Explicit as well as implicit channels of communication with presumed underlying affective states are thought to characterize human interactions with technology (Picard, 1997). While the explicit channel transmits overt messages, the implicit one transmits hidden messages about the communicator (e.g., his/her intention and attitude). However, children with ASD often have communicative impairments (both verbal and nonverbal), particularly regarding expression of affective states (APA, 2000; Green et al., 2002; Schultz, 2005). Typically, observation of facial emotional expressions automatically prompts imitation, termed as mimicry (Canon, Hayes, and Tipper, 2009) due to emotional contagion, social perception, and embodied effect (Moody et al., 2007). But, children with ASD often show an absence of quick, automatic matching of others' emotional expressions (McIntosh et al., 2006) leading to communicative impairments. They often experience states of emotional or cognitive stress measured as Autonomic Nervous System activation without external expression (Picard, 2009) challenging their interests in learning and communicating. These vulnerabilities characterizing the communicative impairments place limits on traditional conversational and observational methodologies. There is a growing consensus that endowing a computer with an ability to understand implicit affective cues should permit more meaningful and natural human-computer interaction (Picard, 1997; Reeves, and Nass, 1996). There are several modalities such as facial expression (Bartlett et al., 2003), vocal intonation (Lee, and Narayanan, 2005), gestures and postures (Asha et al., 2005; Kleinsmith et al., 2005), and eye physiology (Bradley et al., 2008; Partala, and Surakka, 2003; Wilbarger, McIntosh, and Winkielmanc, 2009) that can be utilized to evaluate the affective states of individuals interacting with a computer. However, as children with ASD often have communicative impairments, particularly regarding explicit expression of affective states, we plan to choose the implicit measure by using the eye physiological signals. The physiological signals are continuously available and are not necessarily directly impacted by the communicative impairments (Ben Shalom et al., 2006; Groden et al., 2005; Toichi, and Kamio, 2003). As such, physiological signal acquisition may represent a methodology for gathering rich data despite the potential communicative impairments of children with ASD. In addition, physiological data may offer an avenue for recognizing aspects of affect that may be less obvious for humans but more suitable for computers by using signal processing and pattern recognition tools. Furthermore, there is evidence that the dynamic shifts in indicators of Autonomic Nervous System activity are accompanied with transition from one affective state to another (Bradley, 2000).

When estimating human affective response, an important question is how to operationalize the affective state. Although much existing research on affective computing categorizes physiological signal data into "basic emotions," there is no consensus on a set of basic emotions among the researchers (Cowie et al., 2001). This fact implies that practical choices are required to select target affective states for a given application (Cowie et al., 2001). In part of our completed preliminary research work, we chose anxiety, engagement, and liking to be the target affective states. Anxiety was chosen for two primary reasons. First, anxiety plays an important role in various human-machine interaction tasks that can be related to task performance (Brown et al., 1997). Second, anxiety frequently co-occurs with ASD and plays an important role in the behavior difficulties of children with autism (Gillott, Furniss, and Walter, 2001). Engagement, defined as "sustained attention to an activity or person" (NRC, 2001), has been regarded as one of the key factors for children with ASD to make substantial gains in academic, communication, and social domains (Ruble, and Robson, 2006). With "playful" activities during the intervention, the liking of the children (i.e., the enjoyment they experience when interacting with the computer) may create urges to explore and allow prolonged interaction for the children with ASD, who are susceptible to being withdrawn (Papert, 1993).

A review of literature provides a rich history in support of physiology based methodologies for studying stress (Zhai, and Barreto, 2006), engagement (Anderson, Colombo, and Shaddy, 2006; Jensen et. al., 2009), and other similar mental states based on eye physiological measures such as those derived from blink rate (BR), and pupil diameter (PD). Meehan et al. reported that changes in physiological activity are evoked by different amounts of presence in stressful VR environments (Meehan et al., 2005). Jensen et al. has demonstrated the measurement of BR as important to indicate engagement, with increased BR being observed in ASD participants during task-free periods, but not in the higher engagement state (Jensen et al., 2009). Also, PD is an important indicator of affective processing with significant pupillary constriction for children with ASD while being engaged in attending to static face stimulus (Anderson, Colombo, and Shaddy, 2006). Therefore, the development of a VR-based gaze-sensitive adaptive response technology system for exploration of physiological signals and the target affective state of engagement that may be associated with core social deficits for children with ASD is scientifically and technologically valid and feasible.

Necessity for Monitoring Behavioral Viewing Patterns of Children with ASD

Eye-gaze is a richly informative behavior in face-to-face interaction. In dyadic communication, eye-gaze serves at least five distinct communicative functions (Argyle, and Cook, 1976; Kendon, 1967): regulating conversation flow, providing feedback, communicating emotional information, communicating the nature of interpersonal relationships and avoiding distraction by restricting visual input. Eye-gaze helps control the flow of turn taking in conversations. For example, the person who is listening uses eye gaze to indicate whether he/she is paying attention, while the person who is speaking uses it to track whether the listener is still engaged in the conversation (Colburn, Drucker, and Cohen, 2000). Kendon (Kendon, 1967) reports that a typical pattern of interaction when two people converse with each other consists of the listener maintaining fairly long gazes at the speaker, interrupted by short glances away. In contrast, the speaker makes longer gazes away from the listener with shorter gazes at the listener. For example, a listener looking at the speaker 70 percent of the time during an interaction has been identified as '*normal while listening'* and a speaker looking at the listener 30 percent of the time has been defined as '*normal while speaking*' (Colburn, Drucker, and Cohen, 2000; Argyle, and Cook, 1976).

Thus one's fixation pattern with respect to different components of a visual stimulus plays an important role in communication. Fixation duration is an important indicator of affective processing (Anderson, Colombo, and Shaddy, 2006). Another important indicator of behavioral viewing pattern is the number of fixations of eye-gaze. The higher the fixation frequency on a region as measured by Sum of Fixation Counts (Denver, 2004; Poole, and Ball, 2005), the greater the attention and interest (Just, and Carpenter, 1976) in the target.

However, children with ASD exhibit lower fixation duration (FD) while viewing human faces than the non-human face stimuli (Anderson, Colombo, and Shaddy, 2006). Children with ASD tend to fixate less towards faces and more to other objects (Jones, Carr, and Klin, 2008; Dawson et al., 1998; Pelphrey et. al., 2002; Cohen, and Volkmar, 1997) in the environment. Study reveals that children with ASD exhibit reduced FD while viewing faces with fewer shifts from object to face (Swettenham et al., 1998). Atypical viewing patterns of individuals with ASD may emerge early in childhood (Jones, Carr, and Klin, 2008). Many children with autism are delayed in early, face-related social milestones, such as looking to another person's face to reference that person's reactions or to share their own experience of objects and events (Mundy, Sigman, and Kasari, 1994; Joseph, and Tager-Flusberg, 1997). There is considerable amount of work using static faces (Joseph, and Tanaka, 2003; Trepagnier, Sebrechts, and Peterson, 2002) with offline analysis of gaze information while viewing static scene (Klin et. al., 2002). Eye-tracking techniques have been used to capture one's behavioral viewing patterns to the presented stimuli in terms of instantaneous gaze coordinates (Scassellati, 1998) and visual fixation patterns (Klin et al., 2002). Eye-tracking has great potential for application to technological intervention as a) aytpicalities surrounding eye-gaze and processing of salient social cues, specifically cues and information from faces are thought to be inherent to the disorder (Rutherford, and Towns, 2008; Jones, Carr, and Klin, 2008) and may potentially contribute to the underlying developmental mechanisms of the disorder itself (Dawson, 2008) and b) this technology makes exact location of gaze easily quantifiable with specifically designed regions within the visual stimuli (Anderson, Colombo, and Shaddy, 2006). As such, sophisticated application of eye-tracking technology within complex intervention systems could provide a way for elucidating a wide variety of cognitive processes, from visual–spatial attention to object perception to complex social interactions (Trepagnier et al., 2006). In spite of this potential of eye-tracking technology, development of interactive system based on dynamic gaze patterns of these children to address some of their core deficits in communication and social domains is still at its infancy.

Use of Monitoring Performance Metric for Children with ASD

Performance measurement is an important facet in realizing the success/failure in a particular task and is universally used to assess how well someone has done against some set objectives. Previous study has shown the importance of engagement in determining performance at school for children with ASD with performance being positively correlated with one's engagement (Blackorby, and Cameto, 2005). Studies have shown that communication and social difficulties constitute the primary hindrance to satisfactory job performance among individuals with ASD (Camerena, and Sarigiani, 2009; Ruef, and Turnbull, 2002). These indicate the importance of determining a performance metric and adopting measures to improve performance in tasks. For example, while being engaged in social communication tasks, the performance metric can be the success of an individual to retrieve some intended information from the communicator. Thus, measurement of improvement in task performance is also an important ingredient in ASD intervention.

The novelty of our VR-based gaze-sensitive system with adaptive response technology is that it is individual-specific based on an individual's engagement level predicted by monitoring the eye physiological indices, real-time behavioral viewing pattern, and performance metric of the individual during virtual socially-oriented tasks.

Application of Technology in ASD Intervention

There is growing consensus that appropriately individualized intensive behavioral and educational interventions can improve core social communication vulnerabilities seen in individuals with ASD (NRC, 2001). However, there are potent barriers related to accessing and implementing appropriately individualized intensive intervention services (e.g., limited access to and availability of appropriately trained professionals, lack of available data suggesting which interventions will work better for specific children, concerns about efficacy and generalization regarding certain interventions, and exorbitant costs (Ganz, 2007; Goodwin, 2008)). Given these barriers, researchers are employing technology to develop more accessible, quantifiable, intensive and individualized intervention services for core deficit areas related to ASD (Goodwin, 2008). A growing number of studies are investigating applications of advanced interactive technologies e.g., computer technology, robotic systems, and VR environments to social and communication related intervention (Blocher, and Picard, 2002; Kozima, Nakagawa, and Yasuda, 2005; Parsons, Mitchell, and Leonard, 2004).

Among these alternative interactive technologies, we chose VR because of the numerous reasons for a VR-based intervention system to be particularly relevant for children with ASD. The strength of VR technology for ASD intervention includes controllability, reduced sensory stimuli, individualized approach, safety, and a reduction of human interaction during initial skill training (Strickland, 1997). VR does not necessarily include direct human-to-human interaction, which may work well for an initial intervention to remove the difficulties common in ASD related to mere human interaction that is part of a typical intervention setting involving a child and a clinician (Chen, and Bernard-Opitz, 1993; Tartaro, and Cassell, 2007). Having the controllable complexity of a virtual world with minimized distractions may allow for simplified but embodied social interaction that is less intimidating or confusing for children with ASD than human-to-human interaction (Moore, McGrath, and Thorpe, 2000; Standen, and Brown, 2005). However, VR should not be considered an isolating agent, because dyadic communication accomplished between a child and a VR environment can lead into triadic communication including a clinician, caregiver, or peer and in due course potentially accomplish the intervention goals of developing social communication skills between the child with ASD and another person (Bernard-Opitz, Sriram, and Nakhoda-Sapuan, 2001). Furthermore, the main sensory output of VR is auditory and visual, which may represent a reduction of information from a real-world setting but also represents a full description of a setting without need for imagined components (Sherman, and Craig, 2003; Strickland, 1997). Individuals with ASD can improve their learning skills related to a situation if the proposed setting can be manifested in a physical or visual manner (Kerr, and Durkin, 2004). Since VR mimics real environments in terms of imagery and contexts, it may allow for efficient generalization of skills from the VR environment to the real world (Cromby, Standen, and Brown, 1996). However, since limited social insight and social cognition are vulnerabilities that are often part of the core deficits associated with ASD, individuals may lack the skills to envision abstract concepts or changes to situations on their own. Virtual environments can easily change the attributes of, add, or remove objects in ways that may not be possible in a real-world setting but could be valuable to teach abstract concepts. Therefore, VR can offer the benefit of representing abstract concepts through visual means (e.g., thought bubbles with text descriptions of a virtual character's thoughts) and seamlessly allows for changes to the environment (e.g., changing the color of a ball or making a table disappear) that may be difficult or even impossible to accomplish in a real-world setting (Sherman, and Craig, 2003; Strickland, 1997). Furthermore, the spectrum nature of autism means an individual approach is appropriate, and computers can accommodate individualized treatment (Strickland, 1997). The highly versatile VR environment can illustrate scenarios which can be changed to accommodate various situations that may not be feasible in a given therapeutic setting because of space limitations, resource deficits, safety concerns, etc. (Parsons, and Mitchell, 2002). VR has also shown the capacity to ease the burden, both time and effort, of trained clinicians in an intervention process as well as the potential to allow untrained personnel (e.g., parents or peers) to aid a participant in the intervention (Standen, and Brown, 2005). Therefore, VR represents a medium well-suited for creating interactive intervention paradigms for skill training in the core areas of impairment for children with ASD (i.e., social interaction, social communication, and imagination). However, to date the capability of VR technology has not been fully explored to examine the factors that lead to difficulties in impairments such as social communication, which could be critical in designing an efficient intervention plan.

Despite potential advantages, current VR environments as applied to assistive intervention for children with ASD are designed based only on performance metrics (Parsons, Mitchell, and Leonard, 2004; Tartaro, and Cassell, 2007). Various VR environments have been developed and applied to address specific deficits associated with autism (e.g., understanding of false belief (Swettenham, 1996), attention (Trepagnier et al., 2006), expression recognition (Silver, and Oakes, 2001), social problem solving (Bernard-Opitz, Sriram, and Nakhoda-Sapuan, 2001), and social conventions (Parsons, Mitchell, and Leonard, 2005)). These systems may be able to chain learning via aspects of performance; however, they are not capable of a high degree of individualization. Specifically, these systems cannot automatically detect and respond based on behavioral viewing and eye physiological indices, and thus cannot objectively identify and predict social engagement targeted to the specific child. Given the importance of social engagement (Pan, 2009), behavioral viewing (Trepagnier et al., 2006), eye physiological (Anderson, Colombo, and Shaddy, 2006; Jensen et al., 2009) indices, and performance metrics (Blackorby, and Cameto, 2005), developing a VR-based gaze-sensitive social interactive system that can adaptively respond based on these indices can be critical. Thus the development of such a system can be a step towards achieving realistic social interaction to challenge, and expectantly promote scaffolded skill development in particular areas of vulnerability for the children with ASD.

References

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

Anderson, C.J., Colombo, J., and Shaddy, D.J. "Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder," J. of Cli. and Exp. Neuropsy., vol. 28, pp. 1238–1256, 2006.

Argyle, M., and Cook, M. "Gaze and Mutual Gaze," Cambridge, MA: Cambridge Univ. Press, 1976.

Asha, K., Ajay, K., Naznin, V., George, T., and Peter, D. "Gesture-based affective computing on motion capture data," In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, 2005.

Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., Simmons, A., and Williams, S.C.R. "Social intelligence in the normal and autistic brain: An fMRI study," The European Journal of Neuroscience, vol. 11, no. 6, pp. 1891–1898, 1999.

Bartlett, M.S., Littlewort, G., Fasel, I., and Movellan, J. R. "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," Conference on Computer Vision and Pattern Recognition Workshop, 2003.

Ben Shalom, D., Mostofsky, S. H., Hazlett, R. L., Goldberg, M. C., Landa, R. J., Faran, Y., McLeod, D. R., Hoehn-Saric, R. "Normal physiological emotions but differences in expression of conscious feelings in children with highfunctioning autism," Journal of Autism and Developmental Disorders, vol. 36, no. 3, pp. 395-400, 2006.

Bernard-Opitz, V., Sriram, N., and Nakhoda-Sapuan, S. "Enhancing social problem solving in children with autism and normal children through computer-assisted instruction," Journal of Autism and Developmental Disorders, vol. 31, no. 4, pp. 377-384, 2001.

Blackorby, J., and Cameto, R. "Changes in the School Engagement and Academic Performance of Students with Disabilities," in Engagement and Academics, pp. 1-24, 2005.

Blocher, K., and Picard, R.W. "Affective social quest: emotion recognition therapy for autistic children," In: K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds (Eds.), Socially Intelligent Agents: Creating Relationships with Computers and Robots, Kluwer Academic Publishers, 2002.

Bradley, M.M. "Emotion and motivation," In J.T. Cacioppo, L.G. Tassinary, and G. Berntson (Eds.), Handbook of Psychophysiology, pp. 602-642. New York: Cambridge University Press, 2000.

Bradley, M.B., Miccoli, L., Escrig, M.A., and Lang, P.J. "The pupil as a measure of emotional arousal and autonomic activation," Psychophysiology, vol. 45, no. 4, pp. 602-607, 2008.

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., and Brown, N. L. "Gender and video game performance," Sex Roles, vol. 36, pp. 793-812, 1997.

Camerena, P., and Sarigiani, P. "Postsecondary educational aspirations of high-functioning adolescents with autism spectrum disorders and their parents," Focus on Autism and Other Developmental Disabilities, vol. 24, no. 2, pp. 115-128, 2009.

Canon, P.R., Hayes, A.E., and Tipper, S.P. "An Electromyographic Investigation of the Impact of Task Relevance on Facial Mimicry," Cognition and Emotion, vol. 23, no. 5, pp. 918-929, 2009.

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

CDC, "Prevalence of Autism Spectrum Disorders-ADDM Network," United States, 2006. MMWR Surveill Summ, vol. 58, pp. 1-20, 2009.

Chen, S. H., and Bernard-Opitz, V. "Comparison of personal and computer assisted instruction for children with autism," Mental Retardation, vol. 31, no. 6, pp. 368-376, 1993.

Cohen, H., Amerine-Dickens, M., and Smith, T. "Early intensive behavioral treatment: replication of the UCLA model in a community setting," J Dev Behav Pediatr, vol. 27, no. 2 Suppl, pp. S145-155, 2006.

Cohen, D.J., and Volkmar, F.R. "Handbook of Autism and Pervasive Developmental Disorders," 2nd Ed., New York, NY: John Wiley and Sons, 1997.

Colburn, A., Drucker, S., and Cohen, M. "The role of eye-gaze in avatar-mediated conversational interfaces," In SIGGRAPH Sketches and Applications, New Orleans, Louisiana, USA, 2000.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. "Emotion recognition in human-computer interaction," IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32-80, 2001.

Cromby, J. J., Standen, P. J., and Brown, D. J. "The potentials of virtual environments in the education and training of people with learning disabilities," Journal Intellectual Disability Research, vol. 40, pp. 489-501, 1996.

Dautenhahn, K., and Werry, I. "Towards interactive robots in autism therapy: Background, motivation and challenges," Pragmatics and Cognition, vol. 12, pp. 1-35, 2004.

Dawson, G. "Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder," Dev. Psychopathol, vol. 20, no. 3, pp. 775-803, 2008.

Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., and Brown, E. "Children with Autism fail to orient to naturally occuring social stimuli," Journal of Autism and Developmental Disorders, vol. 28, pp. 479-485, 1998.

Denver, J.W. "The Social Engagement System : Functional Differences in Individuals with Autism," Dissertation, 2004. *www.lib.umd.edu/drum/bitstream/1903/1351/1/umi-umd-1486.pdf*

Ganz, M. L. "The lifetime distribution of the incremental societal costs of autism," Arch Pediatr Adolesc Med, vol. 161, no. 4, pp. 343-349, 2007.

Gillott, A., Furniss, F., and Walter, A. "Anxiety in high-functioning children with autism," Autism, vol. 5, no. 3, pp. 277-286, 2001.

Goodwin, M.S. "Enhancing and accelerating the pace of Autism Research and Treatment: The promise of developing Innovative Technology," Focus on autism and other developmental disabilities, vol. 23, pp. 125-128, 2008.

Green, D., Baird, G., Barnett, A.L., Henderson, L., Huber, J., and Henderson, S.E. "The severity and nature of motor impairment in Asperger's syndrome: a comparison with specific developmental disorder of motor function," Journal of Child Psychology and Psychiatry and Allied Disciplines, vol. 43, no. 5, pp. 655–668, 2002.

Groden, J., Goodwin, M.S., Baron, M.G., Groden, G., Velicer, W.F., Lipsitt, L.P., Hofmann, S.G., and Plummer, B. "Assessing cardiovascular responses to stressors in individuals with autism spectrum disorders," Focus on Autism and Other Developmental Disabilities, vol. 20, no. 4, pp. 244–252, 2005.

Jensen, B., Keehn, B., Brenner, L., Marshall, S.P., Lincoln, A.J., and Müller, R.A. "Increased Eye-Blink Rate in Autism Spectrum Disorder May Reflect Dopaminergic Abnormalities," Intl. Society for Autism Research, Poster Presentation, 2009.

Jones, W., Carr, K., and Klin, A. "Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder," Arch. Gen. Psychiatry, vol. 65, no. 8, pp. 946-954, 2008.

Joseph, R.M., and Tager-Flusberg, H. "An Investigation of Attention and Affect in Children with Autism and Down Syndrome," J. Autism Dev. Disord., vol. 27, pp. 385-395, 1997.

Joseph, R.M., and Tanaka, J. "Holistic and part-based face recognition in children with autism," Journal of Child Psychology and Psychiatry, vol. 44, no. 4, pp. 529–542, 2003.

Just, M.A., and Carpenter, P.A. "Eye fixations and cognitive processes," Cognitive Psychology, vol. 8, pp. 441-480, 1976.

Kendon, A. "Some functions of gaze direction in social interaction," Acta Psychologica, vol. 26, pp. 22- 63, 1967.

Kerr, S., and Durkin, K. "Understanding of thought bubbles as mental representations in children with autism: Implications for theory of mind," Journal of Autism and Developmental Disorders, vol. 34, no. 6, 2004.

Kleinsmith, A., Ravindra De Silva, P., and Bianchi-Berthouze, N. "Recognizing emotion from postures: crosscultural differences in user modeling," In Proceedings of User Modeling, pp. 50-59, 2005.

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," Archives of General Psychiatry, vol. 59, no. 9, pp. 809-816, 2002.

Kozima, H., Nakagawa, C., and Yasuda, Y. "Interactive robots for communication-care: A case-study in autism therapy," in Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, pp. 341-346, Nashville, Tennessee, 2005.

Lee, C. M., and Narayanan, S. S. "Toward detecting emotions in spoken dialogs," IEEE Transactions on Speech and Audio Processing, vol., 13, no. 2, pp. 293-303, 2005.

McGee, G. G., Feldman, R. S., and Morrier, M. J. "Benchmarks of social treatment for children with autism," Journal of Autism and Developmental Disorders, vol. 27, pp. 353–364, 1997.

McIntosh, D.N., Reichmann-Decker, A., Winkielman, P., and Wilbarger, J.L. "When the social mirror breaks: Deficits in automatic, but not voluntary mimicry of emotional facial expressions in autism," Developmental Science, vol. 9, pp. 295–302, 2006.

Meehan, M., Razzaque, S., Insko, B., Whitton, M., and Brooks, Jr., F. P. "Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments," Applied Psychophysiology and Biofeedback, vol. 30, no. 3, pp. 239-258, 2005.

Michaud, F., and Theberge-Turmel, C. "Mobile robotic toys and autism," in Socially Intelligent Agents: Creating Relationships With Computers and Robots, K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds (Eds.) Norwell, MA: Kluwer, pp. 125-132, 2002.

Moody, E.J., McIntosh, D.N., Mann, L.J., and Weisser, K.R. "More Than Mere Mimicry? The Influence of Emotion on Rapid Facial Reactions to Faces," Emotion, vol. 7, no. 2, pp. 447-457, 2007.

Moore, D.J., McGrath, P., and Thorpe, J. “Computer aided learning for people with autism - A framework for research and development,” Innovations in Education and Training International, vol. 37, no. 3, pp. 218-228, 2000.

Mundy, P., Sigman, M., and Kasari, C. “Theory of mind and joint-attention deficits in autism,” in S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen, (eds), Understanding Other Minds: Perspectives from Autism, Oxford University Press, London, pp. 181-203, 1994.

NRC. “Educating children with autism,” National Academy Press, Washington, DC, 2001.

Pan, C.Y. "Age, social engagement, and physical activity in children with Autism Spectrum Disorder," Research in Autism Spectrum Disorders, vol. 3, pp. 22-31, 2009.

Papert, S. “Mindstorms: Children, Computers, and Powerful Ideas,” 2nd ed. New York: Basic Books, 1993.

Parsons, S., and Mitchell, P. "The potential of virtual reality in social skills training for people with autistic spectrum disorders," Journal of Intellectual Disability Research, vol. 46, pp. 430-443, 2002.

Parsons, S., Mitchell, P., and Leonard, A. "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.

Parsons, S., Mitchell, P., and Leonard, A. "Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments?," Autism, vol. 9, pp. 95-117, 2005.

Partala, T., and Surakka, V. "Pupil size variation as an indication of affective processing," Int. J. Human-Computer Studies, vol. 59, pp. 185-198, 2003.

Pelphrey, K.A., Sasson, N.J., Reznick, J.S., Paul, G., Goldman, B.D., and Piven, J. “Visual scanning of faces in autism,” Journal of Autism and Developmental Disorders, vol. 32, pp. 249–261, 2002.

Picard, R. W. “Affective Computing,” Cambridge, MA: MIT Press, 1997.

Picard, R.W. "Future Affective Technology for Autism and Emotion Communication," Philosophical Transactions of the Royal Society B., vol. 364, no. 1535, pp. 3575-3584, 2009.

Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., and De Rossi, D. “An android for enhancing social skills and emotion recognition in people with autism,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 13, no. 4, pp. 507-515, 2005.

Poole, A., and Ball, L.J. “Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects,” In C. Ghaoui (Ed.), Encyclopedia of Human Computer Interaction, Idea Group, 2005.

Reeves, B., and Nass, C. I. “The media equation: how people treat computers, televisions, and new media as real people and places,” New York: Cambridge University Press, 1996.

Rogers, S. J. "Empirically supported comprehensive treatments for young children with autism," Journal of Clinical Child Psychology, vol. 27, pp. 168-179, 1998.

Ruble, L.A., and Robson, D.M. "Individual and environmental determinants of engagement in autism," J. of Aut. and Dev. Dis., vol. 37, no. 8, pp. 1457–1468, 2006.

Ruef, M. B., and Turnbull, A. P. “The perspectives of individuals with cognitive disabilities and/or autism on their lives and their problem behavior,” Research and Practice for Persons with Severe Disabilities, vol. 27, no. 2, pp. 125-140, 2002.

Rutherford, M.D., and Towns, M.T. “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.

Scassellati, B. “Eye Finding via Face Detection for a Foveated Active Vision System,” Fifteenth Nat. Conf. on Artificial Intelligence (AAAI-98), Madison, Wisconsin, USA., pp. 969 – 976, 1998.

Scassellati, B. “Quantitative metrics of social response for autism diagnosis,” in Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, pp. 585-590, Nashville, Tennessee, 2005.

Schultz, R.T. “Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area,” International Journal of Developmental Neuroscience, vol. 23, pp. 125–141, 2005.

Sherman, W. R., and Craig, A. B. “Understanding virtual reality: interface, application, and design,” Boston: Morgan Kaufmann Publishers, 2003.

Sigman, M., and Ruskin, E. “Continuity and change in the social competence of children with autism, down syndrome, and developmental delays,” Monographs of the Society for Research in Child Development, vol. 64, pp. 11–30, 1999.

Silver, M., and Oakes, P. “Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others,” Autism, vol. 5, no. 3, pp. 299–316, 2001.

Standen, P. J., and Brown, D. J. “Virtual reality in the rehabilitation of people with intellectual disabilities: review,” CyberPsychology and Behavior, vol. 8, no. 3, pp. 272-282; discussion pp. 283-288, 2005.

Strickland, D. “Virtual reality for the treatment of autism,” in Virtual reality in neuropsycho-physiology, G. Riva, Ed., pp. 81-86, Amsterdam: IOS Press, 1997.

Strickland, D., Marcus, L. M., Mesibov, G. B., and Hogan, K. “Brief report: two case studies using virtual reality as a learning tool for autistic children,” Journal of Autism and Developmental Disorders, vol. 26, no. 6, 1996.

Swettenham, J. “Can children with autism be taught to understand false belief using computers?,” Journal of Child Psychology and Psychiatry, vol. 37, no. 2, pp. 157-165, 1996.

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., Rees, L., and Wheelwright, S. “The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants,” J Child Psychology and Psychiatry, vol. 39, no. 5, pp. 747–753, 1998.

Tartaro, A., and Cassell, J. “Using Virtual Peer Technology as an Intervention for Children with Autism,” In: J. Lazar (Ed.), Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations. UK: John Wiley and Sons, Chichester, 2007.

Toichi, M., and Kamio, Y. “Paradoxical autonomic response to mental tasks in autism,” Journal of Autism and Developmental Disorders, vol. 33, no. 4, pp. 417–426, 2003.

Trepagnier, C., Sebrechts, M.M., and Peterson, R. “Atypical face gaze in autism,” Cyberpsychology and Behavior, vol. 5, no. 3, pp. 213–217, 2002.

Trepagnier, C.Y., Sebrechts, M.M., Finkelmeyer, A., Stewart, W., Woodford, J., and Coleman, M. “Simulating social interaction to address deficits of autistic spectrum disorder in children,” Cyberpsych Behav, vol. 9, no. 2, pp. 213-217, 2006.

Wieder, S., and Greenspan, S. “Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective?,” The Journal of Developmental and Learning Disorders, vol. 9, 2005.

Wilbarger, J. L., McIntosh, D. N., and Winkielmanc, P. “Startle modulation in autism: Positive affective stimuli enhance startle response,” Neuropsychologia, vol. 47, pp. 1323–1331, 2009.

Zhai, J., and Barreto, A. “Stress Detection in Computer Users through Noninvasive Monitoring of Physiological Signals,” Biomedical Science Instrumentation, vol. 42, pp. 495-500, 2006.

CHAPTER III

VIRTUAL REALITY SYSTEM FOR SOCIAL INTERACTION AND PHYSIOLOGY-BASED AFFECT RECOGNITION

Introduction

The primary objective of this chapter is to present the design and development of a VR-based system for social interaction and to examine a physiology-based approach for affect recognition. The VR-based system discussed in this chapter is capable of systematic manipulation of specific aspects of social communication. The virtual peers (i.e., avatars) within this system can display varying amounts of eye contact, and can vary proximity to the participant, as they interact socially with the participants. The design is evaluated through an experiment that combines ratings reported from a clinical observer with physiological responses indicative of affective states of the participants, both being collected when the participants participate in social tasks with the avatars in the VR environment.

Design Specifications of VR-based Tasks

VR is often effectively experienced on a desktop system using standard computer input devices (Parsons, and Mitchell, 2002) for ASD intervention. Our participants also view the avatars in the VR environment (with avatars narrating personal stories) on a computer monitor from the first-person perspective, which is comparable to research on social anxiety and social conventions (Pereira et al., 2009). Vizard (www.worldviz.com), a commercially available VR design package, is employed to develop the environments.

Within the controllable VR environment, components of the interaction are systematically manipulated to allow users to explore different social compositions. The avatars can make different eye contact and stand at varying distances from the participant in virtual environment. They can converse by lip-synching with the recorded sound files. The participant responds to the avatars using a keypad to select from transparent text boxes superimposed in the corner of the VR scene.

The social parameters of interest for this preliminary work, namely eye gaze and social distance, are manipulated in a 4x2 experimental design, which makes possible eight distinct situations. These parameters are chosen because they play significant roles in social communication and interaction (Bancroft, 1995), and manipulation of these factors may elicit variations in affective reactions (Argyle, and Dean, 1965) and physiological responses (Groden et al., 2005). Each situation is represented three times, which creates 24 trials in the experiment, following a Latin Square design to balance for sequencing and order effects (Keppel, 1991). Each trial of an experiment session includes one avatar for one-on-one interaction with the participant. Participants are asked to participate in a social communication task in VR. In each trial, participants are instructed to watch and listen as the virtual peer tells a 2-min story. The stories are written in first-person. Thus, the task can be likened to having different people introduce themselves to the user, which is comparable to research on social anxiety and social conventions (Argyle, and Dean, 1965; Schneiderman, and Ewens, 1971; Sommer, 1962). Other social parameters, such as facial expression and vocal tone are kept as neutral as possible.

Detailed Specifications of the Social Parameters Studied

The two social parameters e.g., eye gaze and social distance of the virtual peers of the participants are systematically manipulated in this study.

The eye gaze parameter dictates the percentage of time a virtual peer looks at the participant (i.e., staring straight out of the computer monitor). Four types of eye gaze are examined. These are defined as "straight," "averted," "normal," and "flip of normal." Straight gaze means looking straight ahead for the duration of the story (i.e., for the entire trial). Averted gaze means the avatar never attempts to make direct eye contact with the participant, but instead alternates between looking to the left, right, and up. Research represents averted gaze as looking more than 10° away from center in evenly-distributed, randomly-selected directions (Garau et al., 2001; Jenkins, Beaver, and Calder, 2006). Therefore, our averted gaze is an even distribution (33.3% each) of gazing left, right, and up more than 10° from the center. Based on social psychology literature from experimental observations of typical humans (Argyle, and Cook, 1976) and algorithms adopted by the artificial intelligence community to create realistic virtual characters (Colburn, Drucker, and Cohen, 2000; Garau et al., 2001), normal eye gaze is defined as a mix of straight and averted gaze. A person displays varying mixes of direct and averted eye contact depending on if the person is speaking or listening during face-to-face conversations. Since the virtual peer in the VR environment is speaking, we use the gaze definitions for a person speaking, which is approximately 30% straight gaze and 70% averted gaze (Argyle, and Cook, 1976; Colburn, Drucker, and Cohen, 2000). Flip gaze is defined as the flip of normal, which means looking straight approximately 70% of the time and averted 30% of the time, which is indicative of a person's gaze while listening.

The social distance parameter is characterized by the distance between the virtual peer and the participant. Two types of social distance, termed "invasive" and "decorum," are examined. In the VR environment, distance is simulated but can be appropriately represented to the view of the participant. For invasive distance, the virtual peer stands approximately 1.5 ft. from the main view of the scene. This social distance has been characterized as intimate space not used for meeting people for the first time or for having casual conversations with friends (Hall, 1955). A distance of 1.5 ft. apart has been investigated by several research groups in experiments with similar experimental setups to ours in which two people are specifically positioned while one introduces himself/herself to the other and discusses a personal topic for approximately 2 min (Argyle, and Dean, 1965; Schneiderman, and Ewens, 1971; Sommer, 1962), and this invasive distance is characterized by eliciting uncomfortable feelings and attempts to increase the distance to achieve a social equilibrium consistent with comfortable social interaction (Argyle, and Dean, 1965). Decorum distance means the avatar stands approximately 4.5 ft. from the main view of the scene. This social distance is consistent with conversations when meeting a new person or a casual friend (Hall, 1966), and research indicates this distance results in a more comfortable conversation experience than the invasive distance (Argyle, and Dean, 1965). Using Vizard software we project virtual social peers who display different eye gaze patterns at different distances; two examples are shown in Figure III-1.

Figure III-1. Avatar displays direct gaze at invasive distance (top); and averted gaze at decorum distance (bottom).

Design Specifications of the Humanoid Avatars

The virtual peers i.e., the avatars have fixed male or female body (supplied by Worldviz), but Dr. Jeremy Bailenson, director of the Virtual Human Interaction Lab at Stanford University, provided a set of distinct humanoid avatar heads for use in this work. These avatar heads are created from front and side 2D photographs of college-age students. Using 3DMeNow software (biovirtual.com), the photos are then converted to 3D heads for compatibility with Vizard. These avatar heads are chosen because of the following advantages:

(i) open accessibility, (ii) age range close to our participant pool's peers, (iii) and the authentic facial features (e.g., variations in skin complexion, brow line, nose dimensions, etc.) allow the interaction to be interpreted as realistically as possible.

Design Specifications of Audio Files Used

The personal stories that the virtual peers share with the participants are adapted from Dynamic Indicators of Basic Early Literacy Skills (DIBELS, 2007) reading assessments. The assessments are written on topics such as geographical locations, weather phenomena, and intriguing occupations. In each trial of the experiment, an avatar narrates one of these first-person stories to the user. The voices are gathered from teenagers and college-age students from the regional area. Their ages (range = 13-22 years, mean = 18.5 yrs, SD = 2.3 yrs) are similar to the age of people used for the avatar heads and our participant pool.

Design of Menu-Driven Social Interactions

The interaction involves the virtual peer telling a personal story while a participant listens. At the end of the story, the virtual peer asks the participant a question based on some basic facts narrated in the story. The questions are designed to facilitate interaction and to serve as a possible objective measure of engagement. The participant is not aware of the exact question before the story begins so that he/she engages in the task and is not focused on listening to one specific part of the discourse. The questions are intended to be easy to answer correctly if the participant listened to the story. Near the beginning of the first experiment session, the participant takes part in two demonstrations of the process of the VR task; therefore, any difficulty over correctly answering the questions that could be related to not understanding the process of the task is dealt with prior to starting the experiment and collecting data. Each question is accompanied by three possible answer choices (Figure III-2). The correct choice is spoken at least five times during the story, which is sufficient for the information to be relayed (Jonides et al., 2008), and the incorrect choices are never spoken in the story. We expect that a participant who engages in the task would achieve near to or complete 100% accuracy on the questions; and consequently, a severely low percentage of correct answers would indicate a lack of engagement with the task.

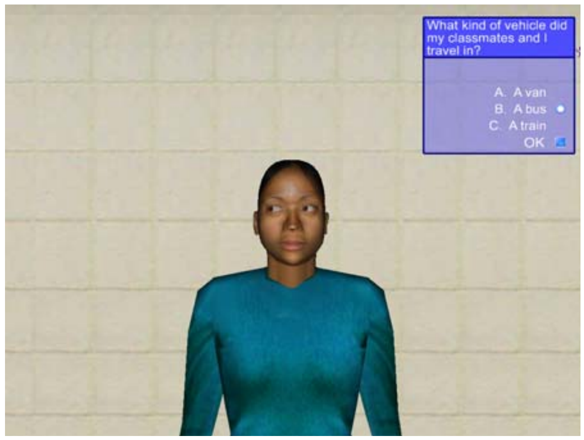


Figure III-2. Example of question asked at the end of a story.

Mapping of Physiological Indices to Affective States

Literature review indicates evidence of the association of physiological activity with the underlying affective states to be differentiated (Bradley, 2000). Studies in the human factors and psychophysiology fields provide a rich history in support of physiology-based methodologies for studying stress (Groden et al., 2005; Zhai et al., 2005), engagement (Pecchinenda, and Smith, 1996), operator workload (Kramer, Sirevaag, and Braune, 1987), mental effort (Vicente, Thornton, and Moray, 1987), and other similar mental states based on physiological measures such as those derived from electromyogram (EMG), galvanic skin response (GSR; i.e., skin conductance), and heart rate variability (HRV). Meehan et al. (Meehan et al., 2005) reported that changes in physiological activity are evoked by different amounts of presence in stressful VR environments.

Prendinger et al. (Prendinger, Mori, and Ishizuka, 2005) demonstrated that the measurement of GSR and EMG can be used to discriminate a user's instantaneous change in levels of anxiety due to sympathetic vs. unconcerned reactions from a life-like virtual teacher. Cardiovascular and EMG activities have been used to examine positive and negative affective states of people (Cacioppo et al., 2000; Papillo, and Shapiro, 1990). Also, Electrodermal activities (EDA) have been shown to be associated with task engagement (Pecchinenda, and Smith, 1996). Different studies have investigated the relationships between both EDA and cardiovascular activities with anxiety (Dawson, Schell, and Filion, 1990; Pecchinenda, and Smith, 1996). Further, variation of peripheral temperature due to emotional stimuli was studied by Kataoka et al. (Kataoka et al., 1998).

In our study presented here, the peripheral physiological signals, such as cardiovascular, electrodermal, electromyographic, etc. of the participants were acquired while they interacted with the VR-based social tasks. At the same time, a clinical observer/therapist and the participant's caregiver/parent rated the participant as to what they thought the level of the affective state (e.g., engagement, enjoyment/liking, and anxiety) was for the participant during the finished trial. Then the physiological signals were mapped to the affective states of the participants.

Experimental Investigation

Participant Characteristics

Thirteen pairs of ASD and typically-developing (TD) participants were recruited through existing clinical and research programs of the Vanderbilt Kennedy Center's Treatment and Research Institute for Autism Spectrum Disorders and Vanderbilt

University Medical Center. Our protocol calls for enlisting participants with ASD age 13-18 years old and an age- and verbal-ability-matched control group of TD participants. ASD participants must have documentation of their diagnosis on the autism spectrum, either Autism Spectrum Disorder, Autistic Disorder, or Asperger's Syndrome, according to their medical records. For all participants, the Social Responsiveness Scale (SRS; Constantino, 2002) profile sheet and Social Communication Questionnaire (SCQ; Rutter, et al., 2003a) are completed by a participant's parent/caregiver before the first session to provide an index of current functioning and ASD symptom profiles. Selection is also based on a receptive vocabulary standard score of 80 or above on the Peabody Picture Vocabulary Test – 3rd Edition (PPVT-III; Dunn and Dunn, 1997) to ensure that language understanding is adequate for participating in the current protocol. Table III-1 presents summary of participant characteristics.

Table III-1 Characteristics of Participants.

Participant (Gender)	**Age (years)**	**PPVT[a] Standard score**	**SRS[b] Total *T*-score**	**SCQ[c] Total score**	**ADOS-G[d] Total score**	**ADI-R[e] Total score**
ASD (N=13)						
Group Mean	16.0	105.9	79.5	21.9	10.7	50.8
TD (N=13)						
Group Mean	15.6	113.7	41.9	3.3	–	–
t-value	0.66	1.50	11.84	9.62		
Exact *p*-value	0.5175	0.1468	1.6500e-11**	1.0341e-9**		

[a]Peabody Picture Vocabulary Test-3rd edition (Dunn, and Dunn, 1997)
[b]Social Responsiveness Scale (Constantino, 2002)
[c]Social Communication Questionnaire (Rutter, et al., 2003a)
[d]Autism Diagnostic Observation Scale-Generic: Module 3 or 4 depending upon subject’s developmental level (Lord, et al., 2000)
[e]Autism Diagnostic Interview-Revised (Rutter et al., 2003b)
Significant group differences, **p<0.001.
No significant group differences were found for the age or PPVT standard score variables (p>0.05 for all).

Procedure

Each participant participated in a total of two sessions lasting for approximately 2.5 hrs. The first session ran approximately 1.5 hrs, due to gathering consent and assent, administering the PPVT-III, and running demonstrations of the social task. The second session lasted about 1 hr. For each completed session, a participant received compensation in the form of gift cards. The equipment setup included a computer dedicated to the social interaction tasks where the participants interacted with the VR environment, biological feedback equipment (www.biopac.com) that collected physiological signals of the participant, and another PC dedicated to acquiring signals from the Biopac system (see Figure III-3). The Vizard Virtual Reality Toolkit ran on a computer (C1) connected to the Biopac system via a parallel port to transmit task-related event-markers (e.g., start/stop of a trial, participant's response to question asked at the end of each trial, etc.). The physiological signals along with the task-related event markers were acquired by the Biopac system and sent over an Ethernet link to the Biopac computer (C2). We also video recorded the sessions to cross-reference observations made during the experiment. The clinical observer/therapist and a participant's parent/caregiver watched the participant from the view of the video camera, whose signal was routed to a television hidden from the view of the participant. The signal from the participant's computer screen where the task was presented was routed to a separate computer monitor (M2) so that the clinical observer and the caregiver could view how the task progressed. Each participant was engaged in two VR-based social interaction sessions on two different days. During the first session, the participants were told about the experiment purpose, the sensors, and the VR tasks. After the physiological sensors were placed, the participants were asked to relax quietly for three minutes while a resting/baseline recording of physiological signals was taken. The first session included two demonstrations of the VR task, the resting/baseline physiological measurement, and a set of eight 2-min trials with different virtual social peers. The second session consisted of the resting/baseline physiological measurement and the remaining 16 trials of social interaction tasks. After each trial, the participant answered a story-related question and self report questions on affective states. The clinical observer and parent/caregiver also rated as to what they thought the level of the affective states of anxiety, engagement, and enjoyment/liking was for the participant during the finished trial.

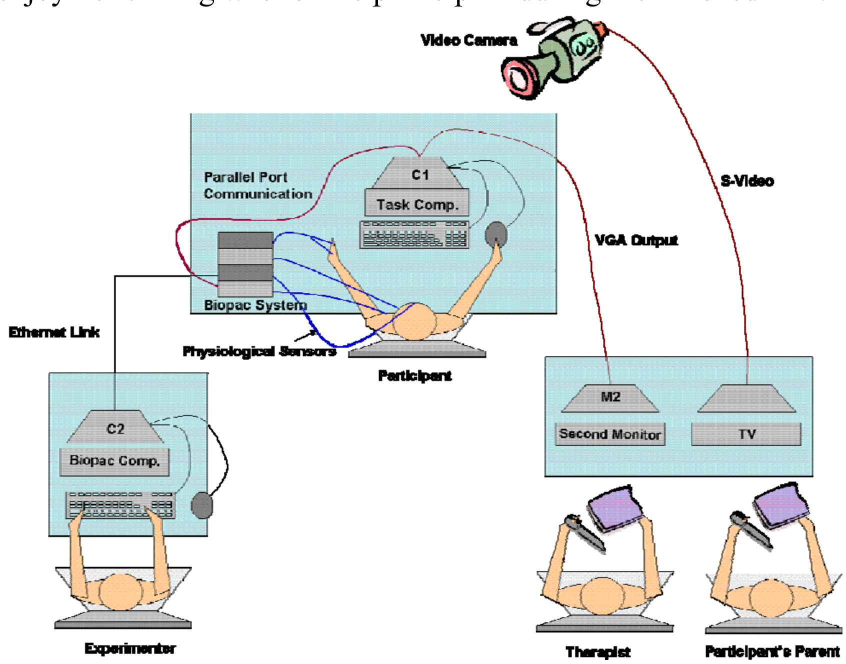


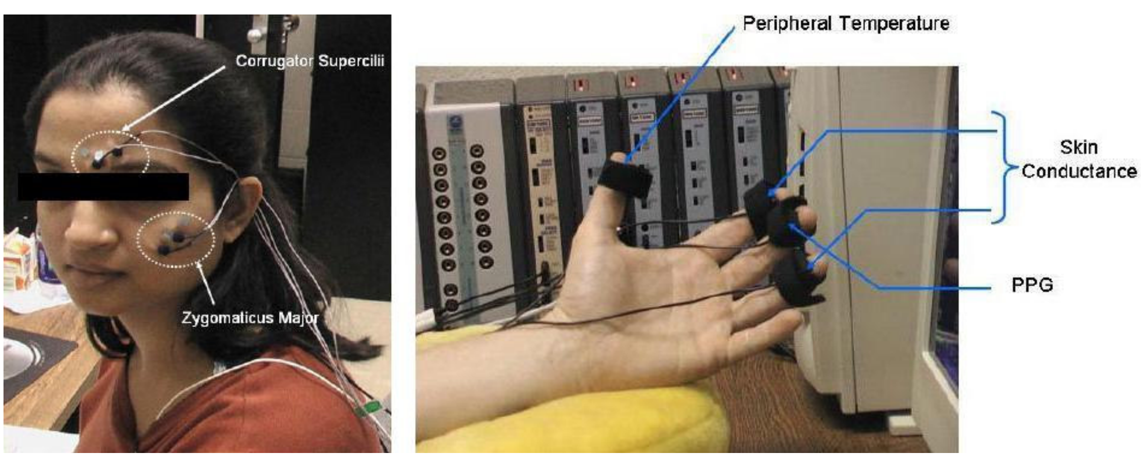
Figure III-3. Experimental setup.

Acquisition of Physiological Signals and Extraction of Physiological Indices

In this work, the physiological signals were acquired using the Biopac MP150 physiological data acquisition system (www.biopac.com). Various physiological signals, broadly classified as Cardiovascular activities including electrocardiogram (ECG), impedance cardiogram (ICG), photoplethysmogram (PPG), and phonocardiogram (PCG)/heart sound; Electrodermal activity (EDA) including tonic and phasic responses from skin conductance; Electromyographic activities from corrugator supercilii, zygomaticus major, and upper trapezius muscles; and Peripheral Temperature were examined. ECG was measured from the chest using the standard two-electrode configuration. ICG describes the changes of thorax impedance due to cardiac contractility and was measured by four pairs of surface electrodes that were longitudinally configured on both sides of the body. A microphone specially designed to detect heart sound waves was placed on the chest to measure PCG. PPG, peripheral temperature, and EDA were measured from the middle finger, the thumb, and the index and ring fingers of the non-dominant hand, respectively. EMG was measured by placing surface electrodes on two facial muscles (corrugator supercilii and zygomaticus major) and an upper back muscle (upper trapezius). Figure III-4a and III-4b show the sensor setup. The sampling rate was fixed at 1000 Hz for all the channels. Appropriate amplification and band-pass filtering were performed.

These signals were selected because they are likely to demonstrate variability as a function of the target affective states, as well as they can be measured non-invasively, and are relatively resistant to movement artifacts (Dawson, Schell, and Filion, 1990; Lacey, and Lacey, 1958). The peripheral physiological signals examined in this work along with the large set of features derived from each signal are described in Appendix A (Table A-1). Signal processing techniques such as Fourier transform, wavelet transform, thresholding, and peak detection were used to derive the relevant features from the physiological signals. For example, inter beat interval (IBI) is the time interval between two "R" waves in the ECG waveform. Power spectral analysis is performed on the IBI data to localize the sympathetic and parasympathetic nervous system activities associated with the different frequency bands. The high-frequency component (0.15–0.4 Hz; which corresponds to the rate of normal respiration) measures the influence of the vagus nerve in modulating the sinoatrial node and is associated with parasympathetic nervous system activity. The low-frequency component (0.04–0.15 Hz) provides an index of sympathetic effects on the heart. The very low-frequency is associated with the frequency band <0.04Hz. The ratios of the power at these frequency components are also computed. PPG signal measures changes in the volume of blood in the finger tip associated with the pulse cycle and provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse Transit Time (PTT) is estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. The heart sound signal measures sounds generated during each heartbeat. The features extracted from the heart sound signal consist of the mean and standard deviation of the third-, fourth-, and fifth-level coefficients of the Daubechies wavelet transform. Bioelectrical impedance analysis (BIA) measures the impedance or opposition to the flow of an electric current through the body fluids contained mainly in the lean and fat tissue. A common variable in recent psychophysiology research, pre-ejection period (PEP) is derived from ICG and ECG and is most heavily influenced by sympathetic innervation of the heart. EDA consists of two main components - tonic response and phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event-related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The EMG signal from corrugator supercilii muscle (eyebrow) captures a person's frown and detects the tension in that region. This EMG signal is also a valuable source of blink information. The EMG signal from the zygomaticus major muscle captures the muscle movements while smiling. Upper trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. Variations in the peripheral temperature mainly come from localized changes in blood flow caused by vascular resistance or arterial blood pressure and reflect the autonomic nervous system activity.

In the work presented in this chapter, we examined the physiological signals collected from the participants when they interacted with their virtual peers during each trial. We investigated the different physiological signals to understand the mapping of physiology with the underlying affective states of anxiety, engagement, and liking.



(a) (b)

Figure III-4. Sensor setup showing the position of facial EMG sensors (a) and the placement of sensors on the non-dominant hand (b).

Results

While our participants interacted with the virtual peers, their affective states of anxiety, engagement and liking were labeled by a clinical observer. In addition, we tried to capture the subtle variations in the physiological signals of the participants and thereby correlate these physiological signals with the affective states as labeled by the clinical observer.

Group Analysis of Physiological Features with Affective States

Our hypothesis was that manipulation of the social parameters in a VR environment may elicit variations in affective reactions (Argyle, and Dean, 1965; Bancroft, 1995) and physiological responses (Farroni et al., 2002; Groden et al., 2005). A participant is likely to experience a range of short-lived affective states (such as, anxiety, interest, etc.) as he/she interacts with the VR system. However, these feelings should not be more intense than the levels of these affective states that are commonly experienced in daily life and should not carry over when the participant leaves the laboratory.

In this work, we studied how the affective states of anxiety, engagement, and enjoyment/liking, measured by ratings from a clinical observer and a participant's physiological signals, vary with respect to the variation of specific communication factors (e.g., social distance and eye contact) presented in the virtual environment. Here we present results of the similarities and differences in physiological responses within the two groups of participants (ASD and TD) during the interaction with the VR avatars associated with manipulation of the two communication factors.

A group of 13 (10 male) adolescents with ASD and a matched group of TD adolescents, age 13-18 years old participated in the VR experiment. Their characteristics are summarized in Table III-1. Physiological signals from the participants and ratings of affective states from a clinical observer, a participant's parent (or caregiver), and self-reports from the participant were recorded during the 2-min. experiment trials. The clinical observer rated what she thought the level of the affective state was for the participant during the finished trial using a binary scale (e.g., Low Engagement or High Engagement). This binary scale was used to label trials as "high" or "low" for data analysis.

Here we present the results of our investigation to evaluate the potential of VR-based social interaction system capable of objectively identifying specific communication aspects to induce affective response in the group of ASD and TD individuals by using a physiology-based approach. The results indicate significant within-group differences in responses to elements of social interaction and this can help to enhance our ability to understand and tailor interventions to the specific vulnerabilities in social communication of participants with ASD. Thus this study can provide valuable information to caregivers and clinicians about the specific affect-eliciting aspects of social communication for this target population. Further, the ability to detect the physiological processes that are a part of impairments in social communication may also prove important for understanding the physiological mechanisms that underlie the presumed core impairments associated with ASD themselves.

Table III-2 presents the reactions in the physiological signals of the participants for trials rated as eliciting “Low Anxiety” (LA) and “High Anxiety” (HA) by the clinical observer. Our preliminary investigation identified certain physiological features of the participants that were statistically different between the two groups. Additionally, we also found certain physiological features that varied similarly between these two groups and those that varied for each of the two groups.

Table III-2. Variations in Physiological Signals of participants during trials rated as eliciting Low Anxiety (LA) and High Anxiety (HA) states.

Physiological Feature	Within ASD (Exact p-value)	Within TD (Exact p-value)
IbiMean (bpm)	0.0223*	0.1311
PEPMean (ms)	0.0367*	0.8044
ZFreqMed (Hz)	0.0441*	0.4304
CBlinkPeakMean (µV)	0.0127*	0.9966
CBlinkStd (µV)	0.0451*	0.4250
PPGIbiMean (ms)	0.2917	0.0399*
PhasicMax (µS)	0.9360	0.0473*
ZMean (µV)	0.6485	0.0070**
ZSlope (µV/s)	0.3130	0.0353*
TStd (µV)	0.9289	0.0471*
PPGPeakMax (µV)	0.1329	0.0012**
PPGPeakMean (µV)	0.6001	0.0311*
PhasicRate (peaks/min)	0.0311*	0.0211*

* : p<0.05; ** : p<0.01

As reports on enjoyment/liking varied from "low liking" (LL) to "high liking" (HL), physiological signals also varied significantly. Table III-3 presents the reactions in the physiological signals of the participants for trials rated as LL and HL by the clinical observer.

Table III-3. Variations in Physiological Signals of participants during trials rated as eliciting Low Liking (LL) and High Liking (HL) states.

Physiological Feature	Within ASD (Exact p-value)	Within TD (Exact p-value)
PEPMean (ms)	0.0004**	0.3002
ZFreqMed (Hz)	0.0216*	0.6309
IbiStd (ms)	0.0060**	0.8758
PowerPara (unit/s2)	0.0061**	0.9333
TonicMean (µS)	0.0146*	0.8791
CMean (µV)	0.0235*	0.3185
HSStdD5	0.0532	0.0116*
PTTStd (ms)	0.4111	0.0097**
ZSlope (µV/s)	0.5082	0.0289*
PPGPeakMax (µV)	0.0984	0.0241*
CStd (µV)	0.0275*	0.0411*

* : p<0.05; ** : p<0.01

The result for significant changes in physiological signals to trials rated as eliciting "low engagement" (LE) versus "high engagement" (HE) for the ASD and TD groups is shown in Table III-4.

Table III-4. Variations in Physiological Signals of participants during trials rated as eliciting Low Engagement (LE) and High Engagement (HE) states.

Physiological Feature	Within ASD (Exact p-value)	Within TD (Exact p-value)
PEPMean (ms)	0.0248*	0.7606
TonicMean(μS)	0.0410*	0.7135
CStd (μV)	0.0497*	0.1312
ZFreqMed (Hz)	0.0079**	0.8917
TFreqMed (Hz)	0.0090**	0.1537
TFreqMean (Hz)	0.0048**	0.2044
PowerVLF (units/s2)	0.0310*	0.3832
TonicMean (μS)	0.0263*	0.74
CBlinkPeakMean (μV)	0.0178*	0.1224
CBlinkStd (μV)	0.0098**	0.3343
PTTStd (ms)	0.7695	0.0393*
PPGPeakMean (μV)	0.6015	0.0487*
PPGPeakMax (μV)	0.36	0.0136*
PowerSym (units/s2)	0.2947	0.0199*
PowerPara/VLF	0.86	0.0364*
CIbiBlinkMean (s)	0.5914	0.0217 *
ZSlope (μV/s)	0.0392*	0.0106*

* : p<0.05; ** : p<0.01

Discussion

In this work, a number of peripheral physiological features, broadly categorized as cardiovascular, electrodermal, electromyographic, etc., were examined for a group of ASD and TD adolescents during social communication task presented on a VR platform for elicitation of multiple affective states. The results show that the VR system provokes variations in both affective ratings and physiological signals to changes in social experimental stimuli for participants with ASD and TD participants. This work used virtual peers and systematically manipulated specific aspects of social communication and thereby provides a vital step towards development of future social interventions using technologies such as VR for the ASD population. Since physiological signals have been shown to be differentiated during social interaction with a virtual environment, the signals could be a useful measure in real-time VR-assisted social skill intervention, an important therapeutic instrument for addressing the core deficits in the ASD population.

References

Argyle, M., and Cook, M. "Gaze and Mutual Gaze," Cambridge, MA: Cambridge University Press, 1976.

Argyle, M., and Dean, J. "Eye-contact, distance and Affiliation," Sociometry, vol. 28, no. 3, pp. 289-304, 1965.

Bancroft, W. J. "Research in Nonverbal Communication and Its Relationship to Pedagogy and Suggestopedia," ERIC, 1995.

Bradley, M. M. "Emotion and motivation," in Handbook of Psychophysiology, J. T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds. New York: Cambridge University Press, pp. 602-642, 2000.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., and Ito, T. A. "The psychophysiology of emotion," in Handbook of Emotions, M. Lewis and J. M. Haviland-Jones, Eds. New York: Guilford, pp. 173-191, 2000.

Colburn, A., Drucker, S., and Cohen, M. "The role of eye-gaze in avatar mediated conversational interfaces," in SIGGRAPH Sketches and Applications, New Orleans, Louisiana, USA, 2000.

Constantino, J. N. "The Social Responsiveness Scale," Los Angeles, California: Western Psychological Services, 2002.

Dawson, M. E., Schell, A. M., and Filion, D. L. "The electrodermal system," in Principles of Psychophysiology: Physical, Social, and Inferential Elements, J. T. Cacioppo and L. G. Tassinary, Eds. New York: Cambridge University Press, pp. 295-324, 1990.

DIBELS, Dynamic Indicators of Basic Early Literacy Skills (dibels.uoregon.edu/measures/), 2007.

Dunn, L. M., and Dunn, L. M. "PPVT-III: Peabody Picture Vocabulary Test," 3rd ed. Circle Pines, Minnesota: American Guidance Service, 1997.

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. "Eye contact detection in humans from birth," in Proceedings of the National Academy of Sciences, vol. 99 no. 14, 2002.

Garau, M., Slater, M., Bee, S., and Sasse, M. A. "The impact of eye gaze on communication using humanoid avatars," in Proceedings of the SIGCHI conference on human factors in computing systems, pp. 309-316, 2001.

Groden, J., Goodwin, M. S., Baron, M. G., Groden, G., Velicer, W. F., Lipsitt, L. P., Hofmann, S. G., and Plummer, B. "Assessing cardiovascular responses to stressors in individuals with autism spectrum disorders," Focus on Autism and Other Developmental Disabilities, vol. 20, no. 4, pp. 244-252, 2005.

Hall, E. T. "The Anthropology of Manners," Scientific American, vol. 192, pp. 84-90, 1955.

Hall, E. T. "The Hidden Dimension," New York: Doubleday & Company, Inc., 1966.

Jenkins, R., Beaver, J. D., and Calder, A. J. "I thought you were looking at me: Direction-specific aftereffects in gaze perception," Psychological Science, vol. 17, pp. 506-513, 2006.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., and Moore, K. S. "The mind and brain of short-term memory," Annual Review of Psychology, vol. 59, pp. 193-224, 2008.

Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M., and Osumi, M. "Development of a skin temperature measuring system for non-contact stress evaluation," in Proceedings of the IEEE Annual Conference on Engineering Medicine Biology Society, pp. 940-943, 1998.

Keppel, G. "Design and Analysis: A Researcher's Handbook," 3rd ed. Prentice Hall, 1991.

Kramer, A., Sirevaag, E., and Braune, R. "A Psychophysiological Assessment of Operator Workload during Simulated Flight Missions," Human Factors, vol. 29, no. 2, pp. 145-160, 1987.

Lacey, J. I., and Lacey, B. C. "Verification and extension of the principle of autonomic response-stereotypy," American Journal of Psychology, vol. 71, no. 1, pp. 50-73, 1958.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. "The Autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism," Journal of Autism and Developmental Disorders, vol. 30, no. 3, pp. 205-223, 2000.

Meehan, M., Razzaque, S., Insko, B., Whitton, M., and Brooks, Jr., F. P. "Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments," Applied Psychophysiology and Biofeedback, vol. 30, no. 3, pp. 239-258, 2005.

Papillo, J. F., and Shapiro, D. "The cardiovascular system," in Principles of Psychophysiology: Physical, Social, and Inferential Elements, J. T. Cacioppo and L. G. Tassinary, Eds. New York: Cambridge University Press, 1990.

Parsons, S., and Mitchell, P. "The potential of virtual reality in social skills training for people with autistic spectrum disorders," Journal of Intellectual Disability Research, vol. 46, pp. 430-443, 2002.

Pecchinenda, A., and Smith, C. A. "The affective significance of skin conductance activity during a difficult problem-solving task," Cognition & Emotion, vol. 10, no. 5, pp. 481-504, 1996.

Pereira, A.F., Yu, C., Smith, L.B., and Shen, H. "A First-Person Perspective on a Parent-Child Social Interaction During Object Play," Proc. 31st Annual Meeting of Cog. Sc. Society, Amsterdam, 2009.

Prendinger, H., Mori, J., and Ishizuka, M. "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game," International Journal of Human-Computer Studies, vol. 62, no. 2, pp. 231-245, 2005.

Rutter, M., Bailey, A., Berument, S., Lord, C., and Pickles, A. "Social Communication Questionnaire," Los Angeles, California: Western Psychological Services, 2003a.

Rutter, M., Le Couteur, A., and Lord, C. "Autism diagnostic interview revised WPS edition manual," Los Angeles, California: Western Psychological Services, 2003b.

Schneiderman, M. H., and Ewens, W. L. “The Cognitive Effects of Spatial Invasion,” The Pacific Sociological Review, vol. 14, no. 4, pp. 469-486, 1971.

Sommer, R. “The Distance for Comfortable Conversations: A Further Study,” Sociometry, vol. 25, no. 1, pp. 111-116, 1962.

Vicente, K., Thornton, D., and Moray, N. “Spectral-Analysis of Sinus Arrhythmia - a Measure of Mental Effort,” Human Factors, vol. 29, no. 2, pp. 171-182, 1987.

Zhai, J., Barreto, A., Chin, C., and Li, C. “User Stress Detection in Human-Computer Interactions,” Biomedical Sciences Instrumentation, vol. 41, pp. 277-286, 2005.

CHAPTER IV

VIRTUAL REALITY SYSTEM FOR SOCIAL COMMUNICATION WITH GAZE-SENSITIVE INDIVIDUALIZED FEEDBACK

Introduction

The objective of this chapter is to describe the design, development and a usability study of a VR-based system seamlessly integrated to technologically enhanced eye-tracking technology to provide individualized feedback. In recent years, several assistive technologies, particularly VR, have been investigated to promote social interactions in children with ASD. Also, it is well-known that these children demonstrate atypical viewing patterns during social interactions and monitoring eye-gaze can be valuable to design intervention strategies. There are several studies that have used eye-tracking technology to monitor eye-gaze with static stimuli along with off-line analysis (Joseph and Tanaka, 2003; Trepagnier, Sebrechts, and Peterson, 2002; Klin, et. al., 2002). Also a recent study has shown that eye-tracking can be used to drive changes in visual behavior of a virtual character in a gaze-contingent individualized manner while following joint-attention task (Wilms et al., 2010). However, there exists no system that monitors eye-gaze dynamically and use this information to provide individualized feedback to investigate the effect of the feedback on the participants' viewing pattern. Given the promise of VR-based social interaction and the usefulness of monitoring eye-gaze in real-time, a novel VR-based dynamic eye-tracking system is developed in this work. This system is capable of delivering individualized feedback based on a child's instantaneous gaze patterns during VR-based social communication task. Results from a usability study with 6 adolescents with ASD are presented that examine the acceptability of this system and investigate how these participants interact with such a system. The results in terms of improvement in behavioral viewing and changes in relevant eye physiological indices of the participants while interacting with the system indicate the potential of this novel technology.

Design of VR-based Gaze-Sensitive Social Communication System with Individualized Feedback Capability

The dynamic closed-loop interaction provided by VR-based Gaze-Sensitive Social Communication System has three main subsystems: (i) a VR-platform that can present social tasks; (ii) a real-time eye-gaze monitoring mechanism; and (iii) an integration module that establishes communication between the VR-based task presentation module and the real-time eye-gaze monitoring module.

VR-based Task Presentation

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

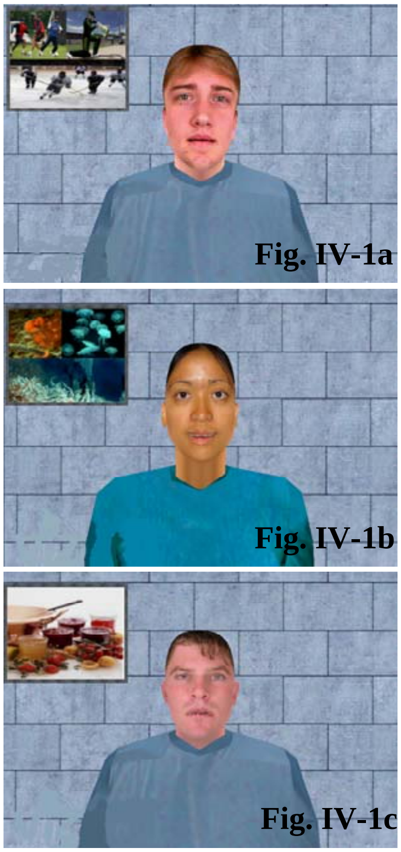


Figure IV-1 Screenshots of avatars demonstrating neutral (*top*), happy (*middle*) and angry (*bottom*) facial expression.

Thus new avatar heads are created from 2D photographs of teenagers, which are then converted to 3D heads by '3DmeNow' software for compatibility with Vizard. These new avatar heads are used to create avatars: (i) with age range close to our participant pool's peers, and (ii) with more authentic facial features (e.g., realistic brow line, nose dimensions, etc.) allowing the interaction to be interpreted as realistically as possible. Facial expressions (e.g., 'neutral', 'happy', 'angry') (Figure IV-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. (Itti, Dhavale, and Pighin, 2003). One can view the avatars within the system from first-person perspective while the avatars narrate personal stories, which is comparable to research on social anxiety and social conventions (Pereira et. al., 2009). In the present study, the first-person stories shared by avatars are adapted from Dynamic Indicators of Basic Early Literacy Skills (Dibels, 2007) reading assessments and includes content thought to be related to potential topics of school presentations (e.g., reports on experiences, trips, favorite activities, etc.). Audio files are developed first by using text-to-speech 'NaturalReader' converter and then recorded using 'Audacity' software. In order for the avatars to speak the content of the story, these audio files are lip-synched with the avatars using a Vizard-based speak module. Additionally, where a participant is looking inside the VR-based visual stimuli (e.g., avatar's face, objects of interest, etc.) is characterized by a set of Regions of Interest (ROIs) that have been programmed such that the dynamic eye-tracking algorithm we develop would keep track of the eye-gaze of the participant as they interact with the VR-based tasks.

Real-time Eye-Gaze Monitoring Mechanism

The system captures eye data of a participant interacting with a virtual peer (i.e., an avatar) using an Eye-Tracker goggles from Arrington Research (http://www.arringtonresearch.com/). This eye-tracker comes with some basic features (e.g., acquiring raw pupil diameter (PD), raw pupil aspect ratio (PAR), etc.) acquiring capability for offline analysis.

One of the key research issues is acquiring the raw eye-gaze data, performing signal processing on this data, and extracting relevant features that can be correlated with engagement and emotion recognition, all in real-time. In this study, we correlate the extracted features reflecting the behavioral viewing patterns of a participant with ASD with his/her engagement level because, engagement, defined as ‘‘sustained attention to an activity or person’’ (NRC, 2001), is one of the key factors for these children to make substantial gains in communication and social domains (Ruble and Robson, 2006). In addition, we correlate the extracted features reflecting the eye physiological indices with emotion recognition capability of the participants because it is characterized as one of the core deficits indicating ASD (Baron-Cohen, 1997; Frith and Frith, 1999) as well as its importance in social communication (Buchnan, Pare´, and Munhall, 2007; Hsiao and

Cottrell, 2008; Williams, et al., 2001).

- *Data Acquisition*

The Eye-Tracker that we use comes with a Video Capture Module with a refresh rate of 30Hz to acquire a participant's eye-gaze data using software called Viewpoint. We designed Viewpoint-Vizard Handshake module (Figure IV-2) for communication between the Viewpoint Interface (Eye-Tracker) and the Vizard Interface (VR platform) modules. Subsequently, we design a new database that captures the task-related event markers (e.g., trial start/stop, amount of viewing of different ROIs, etc.), raw eye physiological signal data (e.g., pupil diameter (PD), pupil aspect ratio (PAR)), raw behavioral viewing data (e.g., fixation duration (FD), 2D gaze coordinates) and performance measures (e.g., a participant's responses to questions asked by the system) with a refresh rate of 30 Hz in a time-synchronized manner. Signal processing techniques such as windowing, noise elimination, and thresholding are used to filter these data to eliminate noise and subsequently extract the relevant features.

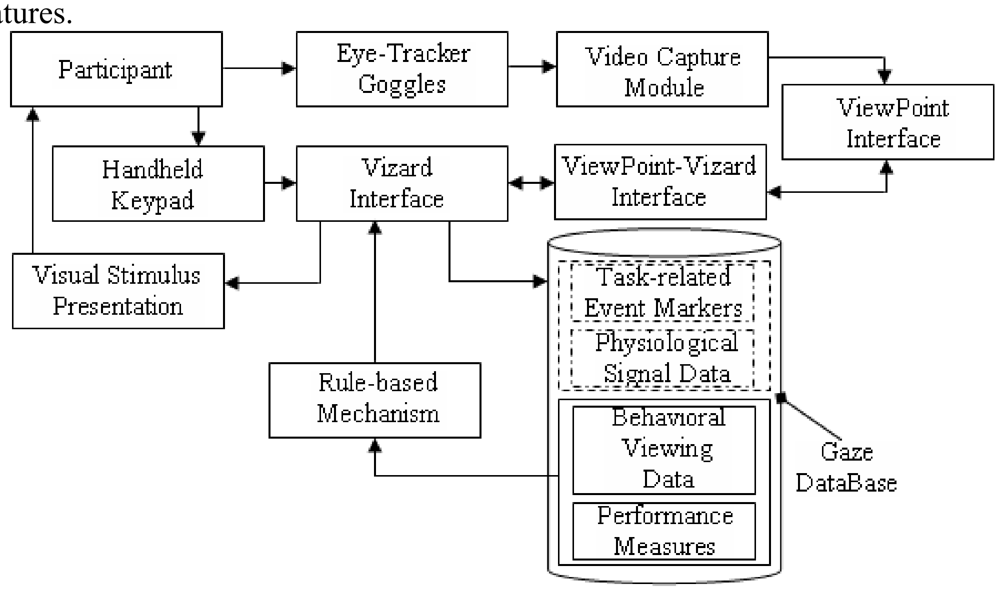


Figure IV-2. Schematic of Data Acquisition and the Control Mechanism used.

- ***Feature Extraction***

The Gaze DataBase (Figure IV-2) is processed to extract 6 features, which are: mean PD (PD_{MEAN}), mean BR (BR_{MEAN}), Sum of Fixation Counts (SFC), Total FD (FD_{TOTAL}), Face-to-nonFace Ratio (FNFR), and Object-to-Face Ratio (OFR) for each ROI from each segment of the signals monitored (Figure IV-3).

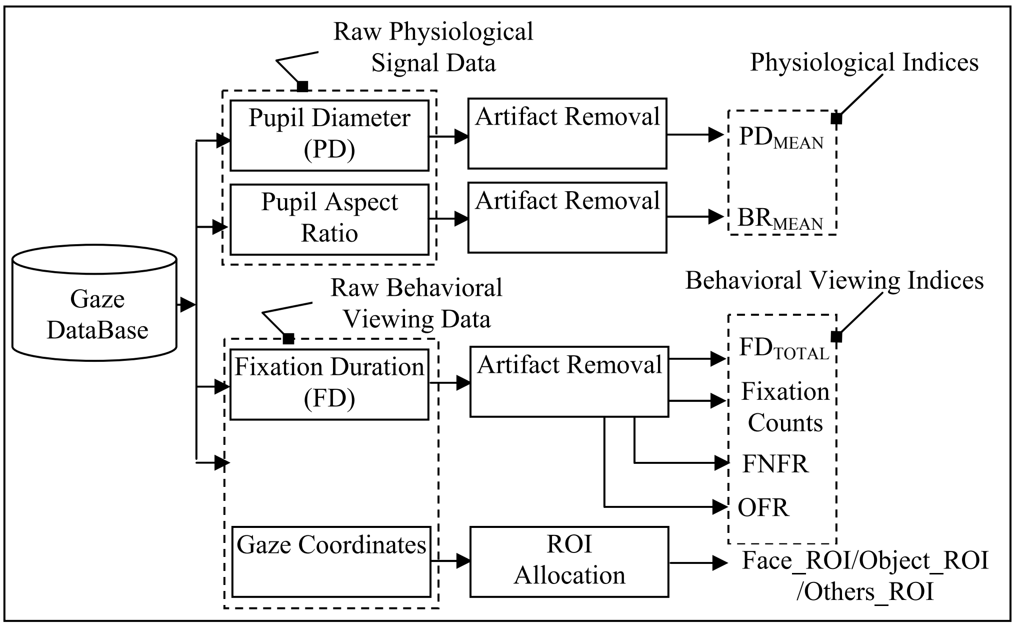


Figure IV-3. Schematic of Feature Extraction.

Computation of PD_{MEAN}: The raw Pupil Diameter (PD) is recorded by Viewpoint software in terms of normalized value (0-1) with respect to the EyeCamera window of the eye-tracker. However, this data does not reflect actual PD. Literature review indicates use of actual PD of typical (Partala, Jokiniemi, and Surakka, 2000; Kahneman, 1973), autistic (Anderson and Colombo, 2009) or schizophrenic (Bar et al., 2008) participants in different studies showing the importance of evaluating the true PD. Again, Anderson et al. (Anderson, and

Colombo, 2009) reports larger tonic pupil size in children with ASD than their typically developing counterparts. In the work presented in this chapter, in order to extract the actual PD at each instant, we use the recorded data on PAR (i.e., the ratio of the major and minor axes of the pupil image) defining the eye image (with 1 indicating a perfect circle). The raw PD value corresponding to PAR closest to 1 defines the optimal PD for a participant. Artifact removal incorporates elimination of the discontinuities in the raw PD due to blinking effects and other minor artifacts as detected by PAR value. Then using the actual EyeCamera window dimensions [Arrington Research Inc. (http://www.arringtonresearch.com/)] (640x480 pixel with each pixel equivalent to 0.13 mm approx. at the high precision setting of 30 Hz.), the PD (in mm) is computed. This is the true PD. We also record the ROIs visited by the eye at each instant. Subsequently, the PD_{MEAN} corresponding to each ROI is computed.

Computation of BR_{MEAN}: The Blink Rate (BR) is determined using the PAR data which is recorded by the Viewpoint software. Although Arrington [Arrington Research, Inc. (2002). Data Collection. In ViewPoint EyeTracker®: PC-60 Software User Guide (pp.47). Scottsdale, Arizona: Arrington Research, Inc.] mentions that blinks can be computed by monitoring the PAR data, Viewpoint software does not provide direct measurement of BR. In the present work, we computed the BR by considering the number of times the PAR value falls below the lower threshold of 0.5 within a window width of 1 minute. This threshold value for PAR was chosen after several trial test runs detecting BR with an accuracy of ±0.05%. Subsequently, the BR_{MEAN} corresponding to each ROI is computed.

Computation of Fixation Counts, and FD_{TOTAL} : The recorded data on FD corresponding to each ROI is first filtered to remove the artifacts due to blinking and noise spikes are eliminated by thresholding. This incorporates filtering the raw data by a moving window having the lower and upper amplitude thresholds of 200 and 450 msec. respectively. There are different views on fixation durations with respect to visual stimuli. In one study (Jacob, 1994), fixations have been stated to typically last between 200-600 ms, where blinks of up to 200 ms may occur during a fixation without terminating it and a window of 50 ms lying outside 1^0 of the current fixation has been considered to terminate a fixation. Some researchers have advised to set the lower threshold for fixation as 100 ms (Inhoff, and Radach, 1998). Still others have classified short fixations with FD < 240 ms and long fixations with FD > 320 ms (Graf and Kruger, 1989). In the present study, we compute the FD by using a thresholding window of 200 ms as the lower limit to eliminate the blinking effects and 450 ms as the upper threshold (i.e., up to 1.5 standard deviations from the lower threshold), the reliable data range restricted by noise due to glare effects of cameras of the eye-tracker that we use. Subsequently, the sum of fixation counts (SFC), and total fixation duration (FD_{TOTAL}) are computed for each ROI.

Computation of ROIs viewed: The 2D gaze coordinates (*x,y*) of the participant's viewing of the presented visual stimulus are recorded. Our computational algorithm, i.e., the Real-time Gaze-based Feedback Algorithm (RGFA) then 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 work presented in this chapter, we segmented the VR-based visual stimulus into 3 ROIs: avatar's face (Face_ROI), a context-relevant object (Object_ROI), and rest of the VR environment (Others_ROI) (Figure IV-4). 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).

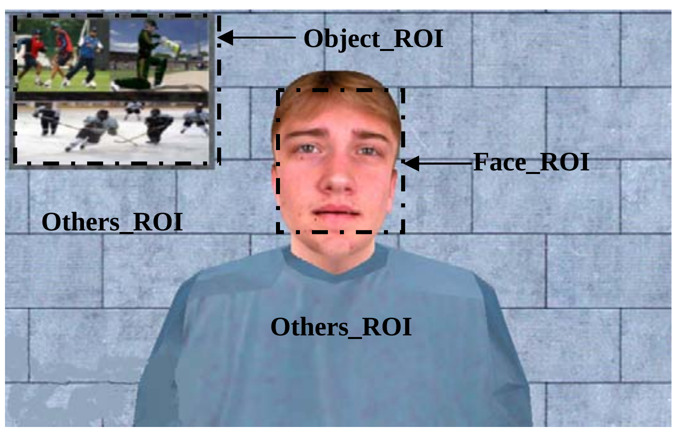


Figure IV-4. Allocation of ROIs (Face_ROI, Object_ROI, and Others_ROI).

Computation of FNFR: A new behavioral index that is analyzed in this work is the Face-to-NonFace Ratio (FNFR). Previous research has indicated an atypical visual scanning pattern of children with ASD while viewing the face and the nonface stimuli, in which they tend to look less towards the face (Anderson, Colombo, and Shaddy, 2006). Thus a computation of FNFR will capture the amount and trend of a participant's viewing patterns towards the face of an avatar. In this work, the visual stimulus presented to our participants is segmented into the Face_ROI, Object_ROI, and the Others_ROI (Figure IV-4). We compute the FD_{TOTAL} for Face_ROI which indicates the total time spent by a child in looking towards the face region of the visual stimulus. Also, we compute the sum of FD_{TOTAL} for Object_ROI and Others_ROI which represents the total time spent by a child in viewing the nonface region of the visual stimulus. Subsequently, the FNFR is computed from the ratio of the total time spent by a child in looking towards the face and nonface regions of the presented visual stimulus. The effect of the gaze-based dynamic feedback on the FNFR is investigated here as the participants view the different ROIs of the visual stimulus during the VR-based social interaction.

Computation of OFR: Another behavioral index that we introduce in this work is the Object-to-Face Ratio (OFR). Children with ASD tend to fixate less towards faces and more to other objects (Jones, Carr, and Klin, 2008; Dawson, et al., 1998) in the environment. Study reveals that children with ASD exhibit reduced FD while viewing faces with fewer shifts from object to face (Swettenham, et al., 1998). We compute the FD_{TOTAL} for Face_ROI and FD_{TOTAL} for Object_ROI indicating the total time spent by a child in looking towards the face region and the object region (Figure IV-4) respectively of the visual stimulus. Subsequently, the OFR is computed from the ratio of these two durations. Thus OFR will indicate how much time a participant spends in viewing the face of the avatar and how much time he/she spends in viewing a context-relevant object. Further, OFR will also quantify the behavioral viewing patterns.

The task-related event markers along with the ROI tags are then used by RGFA to segregate the derived filtered physiological and behavioral indices (as discussed above) during viewing of different ROIs by a participant.

Design of the Integration Module

Unlike the currently available VR environments (Parsons, Mitchell, and Leonard, 2004; Tartaro and Cassell, 2007) as applied to assistive intervention for children with ASD which are designed with an ability to chain learning via aspects of performance alone, the present system uses VR-based social situation as a platform for delivering feedback based on one's performance and real-time gaze patterns, thereby offering a high degree of individualization.

Rationale behind Gaze-based Individualized Feedback Mechanism

The presented system is capable of providing a participant with gaze-based individualized feedback based on the behavioral viewing patterns so as to capture his/her attention. In dyadic communication, eye-gaze information underlying one's expressive behavior (i.e., amount of time a speaker and a listener look at each other) plays a vital role in regulating conversation flow, providing feedback, communicating emotional information, and avoiding distraction by restricting visual input (Argyle, and Cook, 1976). For example, a listener looking at the speaker 70% of the time during an interaction has been identified as '*normal while listening*' (Colburn, Drucker, and Cohen, 2000; Argyle and Cook, 1976).

In the present study, a participant can serve as a listener while interacting with the avatars narrating personal stories and displaying context-relevant facial expressions (Figure IV-1) to capture the mood inherent in the story content. Thus, we chose the '*normal while listening*' criterion for our participants while looking at the avatar during VR-based social interaction. Subsequently, the participant's Fixation Duration (FD) extracted from the behavioral viewing data (FD for Face_ROI viewing as a percentage of total FD) and the performance measure (the participant's response to question asked by the system) initiates a rule-based mechanism (Figure IV-2) to trigger the system to provide feedback (Table IV-1) to the participant using the individualized real-time gaze-based feedback algorithm (RGFA).

Table IV-1. Rationale behind Attention-based Real-time Motivational Feedback.

Response to Q1	$t \geq 70\%$	System Response [Label]
Right	Yes	Your classmate really enjoyed having you in the audience. You have paid attention to her and also made her feel comfortable. Keep it up! [S1]
Right	No	Your classmate did not know if you were interested in the presentation. Perhaps, if you had paid more attention to her, she would have felt more comfortable. Try next time. [S2]
Wrong	Yes	Your classmate felt comfortable in having you in the audience. But, try to pay some more attention to her as she makes the presentation so that you can correctly understand her emotion. [S3]
Wrong	No	Your classmate would have felt more comfortable if you had paid more attention to her. You paid little attention to the presenter. If you had paid more attention to the presenter, then you would have correctly understood her emotion as well as made her feel comfortable. Try next time. [S4]

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

Overview of the Individualized Real-time Gaze-based Feedback Algorithm (RGFA)

The Data Flow Diagram for the RGFA (Figure IV-5) presents a brief overview of the logic used by the present system. 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 Vizard-Viewpoint handshake module (Figure IV-2). A Computer (where the VR-based tasks are presented) runs Viewpoint Software at the background and Vizard software at the foreground and 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 respective buffers which are added up at each instant during participant-avatar interaction. This determines the Face_ROI$_{Time}$,

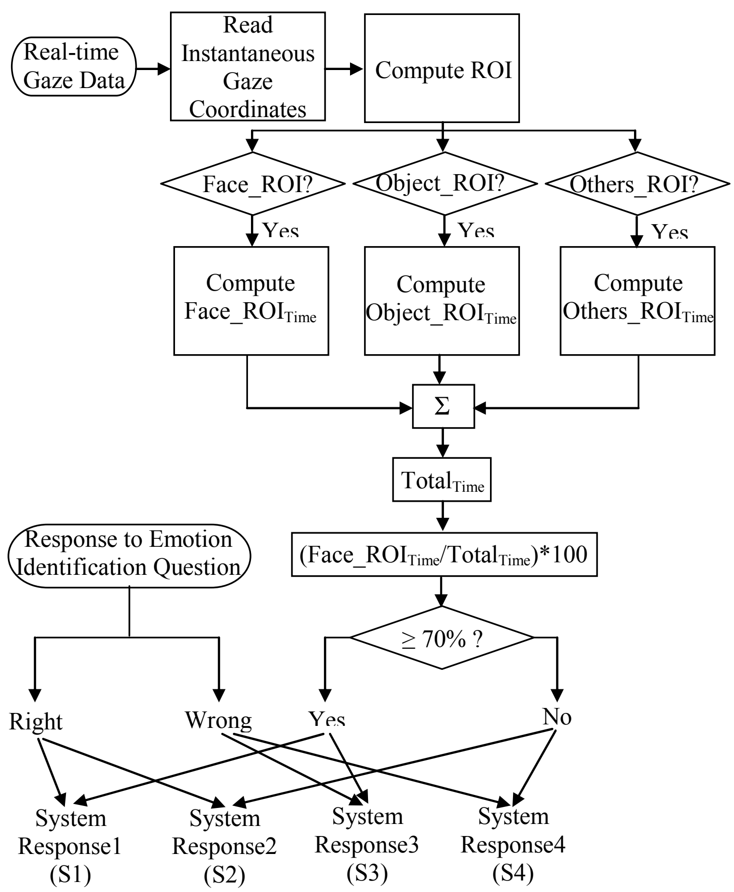


Figure IV-5. Data Flow Diagram for Individualized Real-time Gaze-based Feedback Algorithm (RGFA).

Object_ROI$_{Time}$, and Others_ROI$_{Time}$. Then these times are summed up to get the Total$_{Time}$. Then, the RGFA computes the percentage of time spent by a participant in looking at Face_ROI. Subsequently, based on the participant's percentage of time spent for Face_ROI viewing and response to question asked by the system, 4 different responses (S1-S4) (Table IV-1) are generated by the system.

In short, the algorithm (RGFA) seamlessly integrates the VR-based platform where social tasks are presented with a participant's behavioral viewing patterns as captured by monitoring his/her dynamic gaze data in real-time. This is then used to provide individualized feedback in an attempt to improve the participant's involvement in the social tasks.

Experimental Investigation

Experimental Setup

In this work a pilot study was designed and tested with 6 children with ASD while interacting with the system. We wanted to investigate whether the system was acceptable to the target population, and how the children responded to the system in a virtual social communication task. The experiment was created using the VR design package described in Section '*VR-based Task Presentation*'. The participant's eye movements were tracked by the eye frame Eye-Tracker (discussed in Section '*Real-time Eye-Gaze Monitoring Mechanism*'). Stimuli were presented on a 17" computer monitor (C1) (Figure IV-6a). A chin rest (with height-adjustable telescopic shaft) was designed and used to stabilize the participant's head (Figure IV-6b) and maintain participant-monitor distance of 50cm, considered as an appropriate distance in social gaze-based experiments (Wieser et al., 2009). Uniform room illumination was maintained. The computer (C1) was customized to present the VR-based social tasks in the foreground and compute dynamic gaze information in the background using the eye-tracking data. Gaze data along with task-related event markers (e.g., trial start and trial stop, participant feedback etc.) were logged in a time-synchronized manner. The participant's caregiver (i.e., the observer) watched the participant from a video camera view, whose signal was routed to a television, hidden from the participant's view. Signal from C1 was routed to a separate monitor (M1) for the caregiver to view how the task progressed. Based on these two observations, the observer rated the participant’s engagement level.

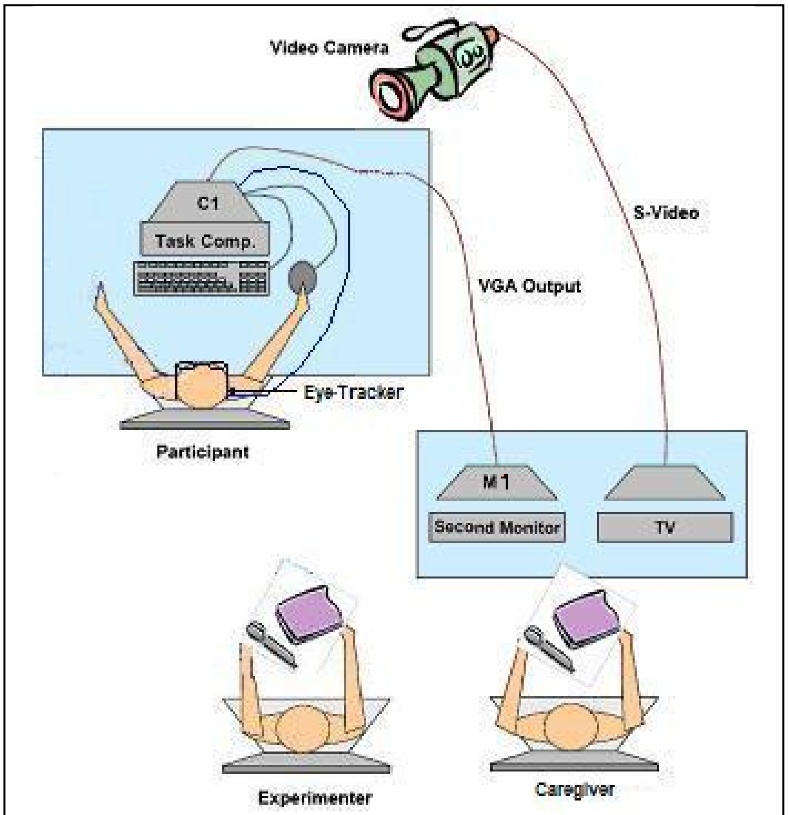


Figure IV-6a. Experimental setup.

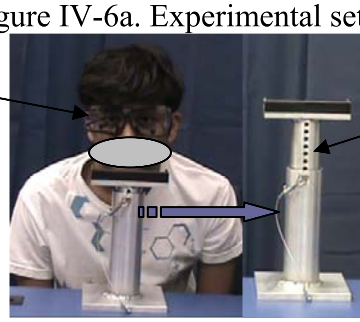


Figure IV-6b. Participant with Eye-Tracker and Chin Rest.

Participant Characteristics

Six adolescents (Male: n = 5, Female: n = 1) with ASD, ages 13-17y (M=15.60y, SD=1.27y) participated in this study. All participants were recruited through existing clinical research programs at Vanderbilt University (VU) and had established clinical diagnoses of ASD. Participants were also required to score ≥ 80 on the Peabody Picture Vocabulary Test-3rd Edition (PPVT-III: Dunn and Dunn, 1997) to ensure that language understanding was adequate for participating in the current protocol. Data on core ASD related symptoms and functioning was obtained through parents' report on the Social Responsiveness Scale (SRS) (Constantino, 2002) profile sheet and the Social Communication Questionnaire (SCQ) (Rutter, et al., 2003) with all participants falling above clinical thresholds. Autism Diagnostic Observation Schedule (ADOS) scores were also available for 5 of the 6 participants from prior evaluation (Table IV-2 provides individual participant characteristics). All research procedures were approved by the VU Institutional Review Board.

Table IV-2. Individual Participant Characteristics.

Participant (Gender)	**Age (years)**	**PPVT[a] Standard score**	**SRS[b] Total *T*-score**	**SCQ[c] Total score**	**ADOS-G[d] Total score**
ASD1 (Male)	13.83	126	69	23	11
ASD2 (Male)	15.5	110	73	13	7
ASD3 (Female)	15.17	83	90	28	10
ASD4 (Male)	16.5	97	63	17	9
ASD5 (Male)	15.08	92	87	20	Not Available
ASD6 (Male)	17.5	103	83	31	20
Mean (SD)	15.60 (1.27)	102 (15)	78 (11)	22 (7)	11 (5)

[a]Peabody Picture Vocabulary Test-3rd edition (Dunn and Dunn, 1997)
[b]Social Responsiveness Scale (Constantino, 2002)
[c]Social Communication Questionnaire (Rutter et al., 2003)
[d]Autism Diagnostic Observation Scale-Generic: Module 3 or 4 depending upon subject's developmental level (Lord et al., 2000)

Procedure

In the present study, we constructed five VR-based social communication scenarios (Trial 1 – Trial 5) in which the virtual peers (i.e., the avatars) narrated personal stories on diverse topics such as, outdoor sports, travel, favorite food, etc. The participants listened and viewed their virtual peers from the first person perspective.

Each participant participated in an approximately 50 min. laboratory visit. During the visit, the participant sat comfortably on a height-adjustable chair and was asked to wear the eye-tracker goggles and the chair was adjusted so that his/her eyes were collinear with center of C1 (Figure IV-6a). The experimenter briefed the participant about the experiment and told him/her that he/she could choose anytime to withdraw from the experiments for any reason, especially if he/she was not comfortable interacting with the system. 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 C1. We achieved a gaze coordinate accuracy of 0.4^0 (or, approx. 0.366 cm on the visual stimulus screen C1 at a 50 cm viewing distance). The task began with the participant resting for 3 min to acclimate him/her to the experimental set-up. The participants viewed an initial instruction screen followed by an interaction with 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 presentations. They were also asked to try and make their classmate feel as comfortable as possible while listening to the presentation. While 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 participant's virtual peer always maintained 'direct' eye-contact (staring straight out of C1) with the participant. The experiment began with trial1 with the virtual classmate exhibiting a 'neutral' facial expression (Figure IV-1a) and narrating a personal story. This trial was followed by 4 other trials that were similar to the trial1 except that in these subsequent trials the virtual peer displayed 'happy' (Figure IV-1b) or 'angry' (Figure IV-1c) facial expressions to capture the mood inherent in the content of the story. After each trial, the participant was asked an emotion-identification question (Q1) and a story-related question (Q2). The Q1 was about the virtual peer's emotion which had 3 answer choices (A. Happy, B. Angry, C. Not Sure). The Q2 was about some basic facts as narrated in the story. It also had 3 answer choices. The correct choice was spoken at least 5 times during the narration, considered sufficient for information relay (Jonides et al., 2008). The incorrect choices were never spoken. The participant responded with a keypad. Q2 was 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 peer, as measured by the real-time computation of the percentage of time spent in looking at the avatar's face, the system encouraged the participant to either pay more or keep the same attention towards the presentation (Table IV-1). After each trial, the observer (e.g., the caregiver) rated about what he/she thought about how engaged the participant was during the VR-based social interaction using a 1-9 scale (1 - least engagement, 9 - most engagement). Each participant was compensated in the form of $15 gift card for completing a session.

In our study, the participant served as a listener while interacting with the avatars. After the participant's reply, an audio-visual feedback, which was computed based on the real-time gaze data to determine the actual time the participant spent looking at the face of the avatar during the presentation, was provided to the participant. The feedback had two parts. First, it informed the participant whether their answers to Q1 and Q2 were correct, and how much attention they paid to the presenter (i.e., the avatar). Second, based on how they responded to Q1 and how much attention they paid to the presenter, the system encouraged them to either pay more or keep the same attention towards the presentation (using RGFA). Since our objective was to encourage a participant to look more towards avatar's face during the social interactions, we used the response to Q1 and amount of attention on the face as the basis for providing feedback. However, Q2 was asked to determine whether the participant was actually paying attention to the story content. Table IV-1 shows the system's responses for providing feedback to the participant.

Results

Here we present the results of our pilot study with 6 adolescents with ASD to (i) examine the acceptability of the system by the target population, and (ii) investigate the effectiveness of the system to elicit variation in participant's engagement level (based on the observers'/caregivers' rating on participants' engagement level) as a result of the individualized feedback. Subsequently, we (iii) analyze the impact of gaze-based dynamic feedback on the behavioral viewing patterns of the participants while scanning the faces of the avatars, and (iv) scanning of the total visual stimulus presented to the participants. These are studied by using the set of quantitative indices e.g., Sum of Fixation Counts (SFC), Total Fixation Duration (FD_{Total}), Face-to-non Face Ratio (FNFR) and Object-to-Face Ratio (OFR), and the scan paths between their gaze fixation points distributed over the different ROIs of the visual stimulus. These behavioral viewing indices interpret the participant's performance from pre-training (PT) (i.e., Trial 1) to post-training (PoT) (Trial 5) trial. In addition, we (v) also present our results that show the ability of the system to influence the eye physiology of the participants during emotion recognition, although our experiment in this usability study was not designed to improve the emotion-recognition capability of our participants.

System Acceptability

In this usability study, we wanted to investigate whether the system, presenting gaze-sensitive VR-based social communication tasks and capable of providing individualized feedback, was acceptable to the children with ASD. In order to achieve this, we tested our system with a small sample of 6 participants with ASD. In spite of being given the option of withdrawing from the experiment at any time during their interaction with the system, all the participants completed the session. An exit survey carried out at the end of the experiment revealed that all 6 participants liked interacting with the system, and had no problem in either wearing the eye-tracker goggles, understanding the stories narrated by their virtual peers, or responding to questions asked by the system. In fact 5 of them inquired whether there would be any future participation possibilities with this new system. Thus it is reasonable to infer from this small usability study that the system has a potential to be accepted by the target population.

Impact of gaze-based dynamic feedback on Participants' Engagement (based on Caregivers' rating)

We wanted to assess whether the presented system can be used in virtual social communication task to create improved engagement levels among the participants so that engagement manipulation using individualized feedback could be potentially feasible in the future as a part of intervention. In our usability study with the system, the participants' caregivers rated as to what they thought regarding the participants' engagement level while interacting socially with their virtual peers. We asked the caregivers to rate the participants using a 1-9 scale (1 - least engagement, 9 - most engagement). With dynamic feedback during VR-based social interaction, the reported group engagement mean (as evident from the caregivers' rating on participants' engagement level) (Table IV-3) improved during Post-Training (PoT) trial from Pre-Training (PT) trial. For all participants (except ASD2) the engagement rating improved from PT to PoT. Further analysis revealed that ASD2 was incorrect in responding to story-related question in PoT (i.e., Trial5) which may be due to his lower engagement. The caregiver of ASD2 reported that he liked the story in PT (i.e., Trial1) the most and the PoT (i.e., Trial5) the least. Also the range (1-9 scale) of engagement rating shows that group engagement increased during PoT.

Table IV-3. Impact of gaze-based dynamic feedback on Participants' Engagement.

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

PT : Pre-Training (i.e., Trial1); PoT: Post-Training (i.e., Trial5).

Impact of gaze-based dynamic feedback on Behavioral Viewing Patterns in terms of Attention to the Faces (Face_ROI) of the Avatars

We chose to use certain primary behavioral viewing indices (e.g., SFC, and FD_{Total}) of the participants, as they viewed the Face_ROI of the avatars while attending to the avatars' presentations to gauge attention towards social stimuli in the VR environment. Results indicate that the participants looked more frequently towards the face region (Face_ROI) of the avatars from the pre-to-post measurement. This is reflected from the improvement in the SFC for each participant from pre-training (PT) to post-training (PoT) measurement for Face_ROI viewing with dynamic feedback (Table IV-4) with SFC for Face_ROI viewing during PT trial being statistically different ($t = 3.464$; $p = 0.0180$) from that during PoT trial by using a dependent sample t-test between these two groups.

Also, in this work, the FD of the participants was analyzed while viewing Face_ROI due to its importance as an indicator of social engagement (Jones, Carr, and Klin, 2008). The FD_{Total} of the participants was computed during Face_ROI viewing and the results indicate increase in this index for all of the participants from pre-to-post measurement and in statistically different ways ($t = 8.068$; $p = 0.0005$) by using a dependent sample t-test between these two groups (Table IV-4). Overall, the results reflect a trend for participants to not only fixate on the Face_ROI more frequently, but also for a longer duration with dynamic feedback.

Table IV-4. Improvement in Viewing Pattern in terms of Sum of Fixation Counts (SFC), and Total Fixation Duration (FD_{Total}) while viewing Face_ROI.

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

Impact of gaze-based dynamic feedback on Behavioral Viewing Patterns in terms of Scanning of the total Visual Stimulus (i.e., Face_ROI, Object_ROI, and Others_ROI)

Furthermore, both FNFR and OFR were computed from the above primary viewing indices. Results of FNFR based on FD_{Total} of the participants indicate a non-statistically significant trend ($t = 1.3332$; $p = 0.2400$) toward improvement in viewing patterns (Table IV-5). The percent of total fixation duration towards Face_ROI, as compared to the Object_ROI and Others_ROI improved (Figure IV-7) as well for all participants implying that each participant looked at avatar's face for a longer duration of time during the PoT than the PT trial. Thus, with the gaze-based feedback, the participants attended to the Face_ROI of the avatars more than the non-face regions (i.e., the Object_ROI and the Others_ROI).

Table IV-5. Improvement in Viewing Pattern in terms of Face-To-NonFace Ratio (FNFR).

Participant	**FNFR based on FD_{Total}**		
	PT	**PoT**	**%Improvement**
ASD1	1.2805	2.9336	129.10
ASD2	8.3477	403.3562	4731.95
ASD3	3.4581	19.9005	475.47
ASD4	8.2131	40.5131	393.27
ASD5	3.8033	48.4755	1174.55
ASD6	2.2628	13.1104	479.39
M(SD)	4.5609(3.02)	88.0482(155.40)	1230.62 (1749.95)

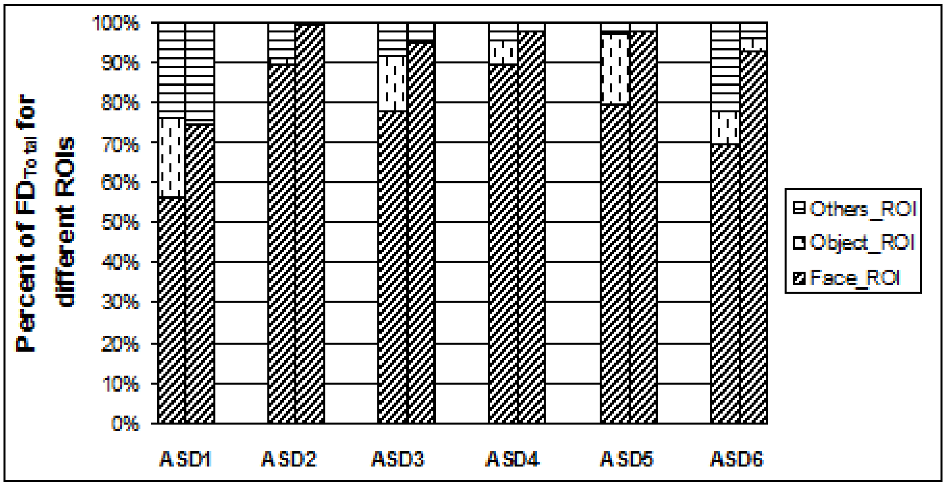


Figure IV-7. Comparative Analysis of FD_{Total} for Face_ROI, Object_ROI, and Others_ROI Viewing for each Participant (Left bars indicate PT trial and Right bars indicate PoT trial).

Subsequently, the Object-to-Face Ratio (OFR) was computed based on the FD_{Total} while viewing the face_ROI and object_ROI. From Table IV-6, it can be seen that the OFR decreased from PT to PoT trial for each participant, implying that the participants fixated more on the face region than the context-relevant object of the visual stimulus, during PoT trial of VR-based tasks. The group mean FD_{Total} during Face_ROI viewing increased by 52.91% from PT to PoT ($p = 0.0005$). Significant group difference ($p = 0.0351$) existed for FD_{Total} during Object_ROI viewing, which decreased by 91.20% and also for OFR ($t = 3.1722$; $p = 0.0248$) which decreased by 95.16% between the PT and PoT trials. Thus, with gaze-based dynamic feedback, the participants demonstrated increased attention to the faces of the avatars and reduced distraction by objects.

Table IV-6. Improvement in Viewing Pattern in terms of Object-to-Face Ratio (OFR) based on Total Fixation Duration (FD_{Total}).

Participant	OFR		
	PT	PoT	%Reduction
ASD1	0.3622	0.0049	98.64
ASD2	0.0242	0.0000	100.00
ASD3	0.1802	0.0041	97.72
ASD4	0.0707	0.0000	100.00
ASD5	0.2269	0.0000	100.00
ASD6	0.1250	0.0317	74.62
M(SD)	0.1649(0.12)	0.0068(0.01)	95.16 (10.11)

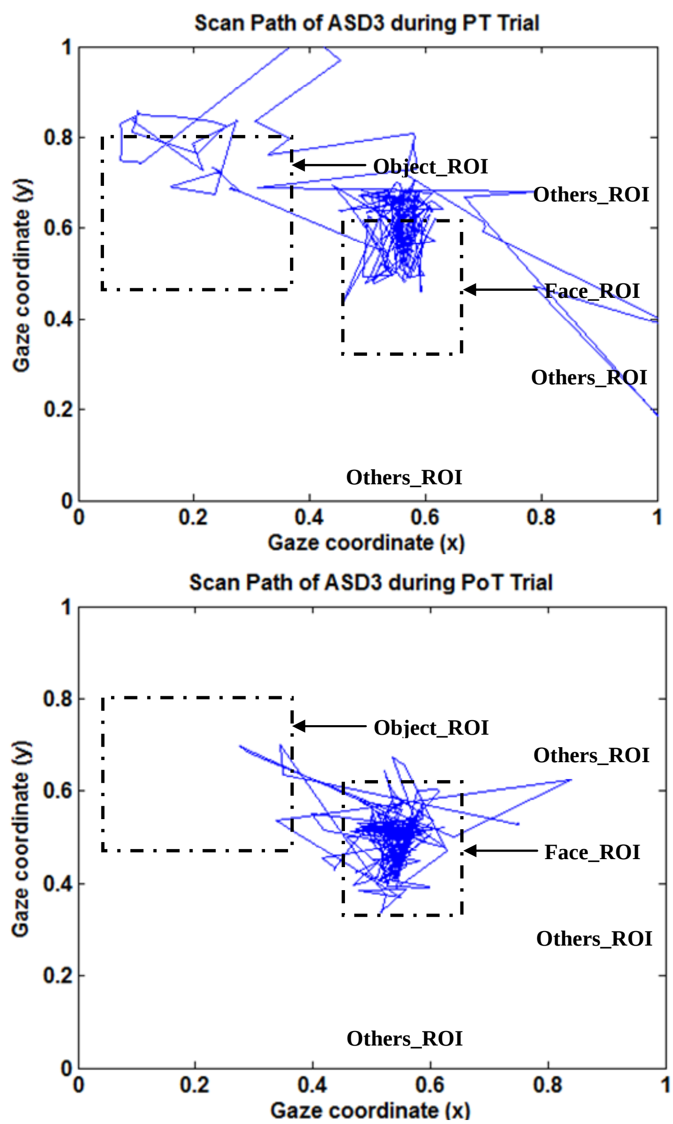


Figure IV-8. Improvement in Behavioral Viewing pattern of ASD3 in terms of Scan Path distributed over different ROIs (Face_ROI, Object_ROI, and Others_ROI).

We also studied the impact of dynamic feedback on the scan paths of the participants as children with ASD have been shown to exhibit atypical scan paths during social interaction (Rutherford, and Towns, 2008). Our investigation revealed that all participants fixated more on the Face_ROI of the avatars, with reduced distraction by the Object_ROI and the Others_ROI, during the PoT trial as compared to the PT trial. For example, as is evident from the scan path (Figure IV-8), ASD3 fixated on different ROIs of the visual stimulus during the PT trial. However, during the PoT trial, ASD3 fixated mainly on the Face_ROI and much less on the Object_ROI and Others_ROI. Note that, these scan paths were analyzed in the background and they were not visible to the participant.

Potential of VR-based Gaze-sensitive system to influence the Eye Physiological Indices (e.g., BR and PD) as Function of the Participants' capability of Emotion Recognition

Children with ASD often experience states of emotional or cognitive stress measured as Autonomic Nervous System activation without external expression (Picard, 2009) challenging their interests in learning and communicating. Thus observation of facial expressions may not be reliable to learn whether they are able to recognize the emotional expressions of others during social communication (McIntosh, et al., 2006; Picard, 2009). In this context, eye physiological indices could be a valuable source to indicate the process of emotion recognition in these children. In fact, literature review indicates an important role of BR and PD in emotion recognition.

One study reports that children with autism exhibit normal BR on seeing static human faces displaying emotional expression (Wong et al., 2008). Other findings show startle potentiation for both positive and negative stimuli (static pictures) for children with ASD

(Wilbarger, McIntosh, and Winkielmanc, 2009) and increased BR for some emotions e.g., anger (Karson, 1983). Additionally, pupil has been considered as an indicator of emotion recognition (Bradley et. al., 2008). However, links between PD and emotional status are not yet clearly established due to diverse views. One study reports pupillary dilation for pleasant, and contraction for aversive stimuli (Hess, 1972). Again, other studies indicate pupil to have sympathetic innervation with pupillary dilation to both pleasant and unpleasant auditory stimuli (Partala, Jokiniemi, and Surakka, 2000; Partala and Surakka, 2003), being greater to unpleasant than pleasant visual stimuli (Libby, Lacey, and Lacey, 1973). Still another has reported ability of static visual displays of avatars displaying emotional expressions to create pupillary dilation to pleasant facial expression and this ability being reduced by the participants' emotional habituation beyond first two avatars (Causse et al., 2007).

Eye physiological indices, namely BR and PD can be made continuously available within the system using the feature extraction method discussed in Section '*Feature Extraction*' when a participant socially interacts with the avatars. As a result, we analyzed to see whether the eye physiological indices are also influenced by interaction with the system.

Analysis of changes in Blink Rate: In our present study, investigation results, as presented in Table IV-7, (similar to the findings of non-VR based applications studied by Wilbarger, McIntosh, and Winkielmanc, 2009; Karson, 1983) reflect a higher change in BR_{MEAN} for Neutral-to-Angry (an overall increase of 107.93% and $p = 0.0489$) than that for *Neutral*-to-*Happy* (an overall increase of 49.91% and $p = 0.0495$) for all participants, except ASD3. In addition, BR_{MEAN} for all participants (except ASD5) while viewing the *Angry* and *Happy* facial expression of their virtual peer was greater than that of 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 as *Happy*. Also, ASD5 was not able to identify the *Neutral* facial expression and misidentified this as *Happy* and he possessed a much higher BR_{MEAN} in general, as compared to the other participants.

Table IV-7. Change in Blink Rate (BR) as a measure of emotion recognition.

Participant	BR_{MEAN} (times/min)		%Increase	BR_{MEAN} (times/min)		%Increase
	(*Neutral*)	(*Happy*)	*Neutral*-To-*Happy*	(*Neutral*)	(*Angry*)	*Neutral*-To-*Angry*
ASD1	8.13	13.02	60.18	8.13	23.83	193.20
ASD2	5.42	8.13	50.00	5.42	8.63	59.23
ASD3	7.35	13.16	79.05	7.35	9.86	34.15
ASD4	12.39	14.42	16.38	12.39	15.20	22.68
ASD5	43.34	42.41	-2.16	43.34	71.50	64.96
ASD6	11.30	22.15	96.02	11.30	42.19	273.36

Analysis of changes in Pupil Diameter: In the present investigation, our findings on PD as presented in Table IV-8, are in line with some of the previous non-VR based findings (e.g., Libby, Lacey, and Lacey, 1973; Partala, Jokiniemi, and Surakka, 2000; Partala and Surakka, 2003). We found that the PD_{MEAN} of each participant was less for pleasant (*Happy*) than that for the unpleasant (*Angry*) one and both being greater than that with no emotional (*Neutral*) expression (except ASD5). Also, the change in PD_{MEAN} for *Neutral*-to-*Angry* (an overall increase of 8.84% and $p = 0.1653$) was found to be greater than that for *Neutral*-to-*Happy* (an overall increase of 3.43% and $p = 0.0721$) for all participants (except ASD5) (Table IV-8). We examined the participants' responses to each of the emotion-identification questions and found that ASD5 could not identify the *Neutral* and

Angry facial expressions of his virtual peers, misidentifying them as *Happy*. We believe that this may be the reason for his percent increase in PD_{MEAN} for *Neutral*-to-*Angry* to be lower than that for *Neutral*-to-*Happy*.

Table IV-8. Change in Pupil Diameter (PD) as a measure of emotion recognition.

Participant	PD_{MEAN} (*Neutral*) (mm)	PD_{MEAN} (*Happy*) (mm)	%Increase *Neutral*-To-*Happy*	PD_{MEAN} (*Neutral*) (mm)	PD_{MEAN} (*Angry*) (mm)	%Increase *Neutral*-To-*Angry*
ASD1	8.036	8.809	9.61	8.036	10.713	33.30
ASD2	6.213	6.272	0.96	6.213	6.394	2.92
ASD3	6.730	6.966	3.51	6.730	7.215	7.21
ASD4	6.741	6.817	1.12	6.741	7.262	7.73
ASD5	7.678	8.032	4.61	7.678	7.711	0.44
ASD6	7.209	7.264	0.77	7.209	7.311	1.42

In the present usability study, with a limited sample size, we find that the BR of the participants is more sensitive to their ability of recognizing different emotional expressions exhibited by their virtual peers, than the PD. For PD, the percent change for *Neutral*-to-*Happy* and that for *Neutral*-to-*Angry* though quite small, yet, the overall trend is similar to that of other non-VR based tasks. More importantly, the above results indicate the ability of the system to correlate the eye physiological indices (BR and PD) to a participant's ability to recognize emotions while interacting socially with virtual peers.

Discussion

In the work presented in this chapter, we set out to a) develop a new technology-based system that could measure gaze information and provide dynamic feedback during social interaction tasks presented in a VR environment and to b) assess the impact of such a feedback on the viewing patterns of a small sample of adolescents with ASD. There is considerable amount of work using static faces (i.e., photographs) (Joseph and Tanaka, 2003; Trepagnier, Sebrechts, and Peterson, 2002) with published results on offline analysis of gaze information while viewing static scene (Klin, et al., 2002). However, work on VR-based systems with a capability to process eye-gaze data in real-time and communicate this individualized information (as feedback) to the participants is at its infancy. In this chapter, we describe the development of a prototyped model of VR-based social communication system for children with ASD with the ability to process eye-gaze information in real-time and communicate this to the VR environment to provide feedback to the participants based on their instantaneous interaction with the virtual social world. While our feedback mechanism was limited to providing systematic information about performance at the end of a several minute interval, we actually realized capability for calculating viewing indices in real-time (i.e., every 33 ms). Thus, while our technology paused and presented feedback to participants within fairly discrete training trials, ultimately the developed technology is capable of providing feedback in an on-line, continuous manner. Such capability suggests great potential for flexible intervention paradigms. For example, such feedback could be continuously monitored and conveyed to the participant when they are not paying proper attention or levels of engagement could be set and modified in an individualistic and relativistic manner (i.e., thresholds of performance based on baseline and learning trajectory).

In addition, the results of this study indicate the ability of the system to measure the eye physiological indices (blink rate and pupil diameter) and correlate these as function of a participant's ability to recognize emotions while interacting socially with virtual peers. Thus, the results suggest that a participant's eye physiological response in VR-based social communication task as presented in VR-based gaze-sensitive system indicate whether or not one is able to recognize emotion similar to that which has been observed in non-VR based tasks. Therefore, it is reasonable to believe that such a system could be used in intervention, perhaps as a supplementary tool, to allow an individual with ASD to enhance his/her social communication skills. The developed technology reported here could be integrated into a more complex and sophisticated social interaction task to achieve targeted goals if paired with appropriate reinforcement paradigms.

References

Anderson, C.J., and Colombo, J. “Larger tonic pupil size in young children with autism spectrum disorder,” Dev. Psychobiol., vol. 51, no. 2, pp. 207-211, 2009.

Anderson, C.J., Colombo, J., and Shaddy, D.J. "Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder," J. of Cli. and Exp. Neuropsy., vol. 28, pp. 1238–1256, 2006.

Argyle, M., and Cook, M. "Gaze and Mutual Gaze," Cambridge, MA: Cambridge Univ. Press, 1976.

Arrington Research Inc. (http://www.arringtonresearch.com/)

Arrington Research, Inc. “Data Collection. In *ViewPoint EyeTracker®: PC-60 Software User Guide* (pp.47). Scottsdale, Arizona: Arrington Research, Inc., 2002.

Bar, K-J., Boettger, M.K., Schultz, S., Harzendorf, C., Agelink, M.W., Yeragani, V.K., Chokka, P., and Voss, A. “The interaction between pupil function and cardiovascular regulation in patients with acute schizophrenia,” Clinical Neurophysiology, vol. 119, pp. 2209-2213, 2008.

Baron-Cohen, S. “Mindblindness: An essay on autism and theory of mind,” Cambridge: MIT, 1997.

Bradley, M.B., Miccoli, L., Escrig, M.A., and Lang, P.J. “The pupil as a measure of emotional arousal and autonomic activation,” Psychophysiology, vol. 45, no. 4, pp. 602-607, 2008.

Buchnan, J.N., Pare´, M., and Munhall, K.G. “Spatial statistics of gaze fixations during dynamic face processing,” Social Neuroscience, vol. 2, no. 1, pp. 1-13, 2007.

Causse, M., Pavard, B., Sénard, J-M., Démonet, J-F., and Pastor, J. “Emotion Induction through Virtual Avatars and its Impact on Reasoning: Evidence from Autonomous Nervous System Measurements and Cognitive Assessment,” Actes de la Conférence VRIC’06, Laval, France, 2007.

Colburn, A., Drucker, S., and Cohen, M. “The role of eye-gaze in avatar-mediated conversational interfaces,” In SIGGRAPH Sketches and Applications, New Orleans, Louisiana, USA, 2000.

Constantino, J.N. “The Social Responsiveness Scale,” California: Western Psych. Serv., 2002.

Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., and Brown, E. “Children with Autism fail to orient to naturally occuring social stimuli,” Journal of Autism and Developmental Disorders, vol. 28, pp. 479-485, 1998.

Dunn, L.M., and Dunn, L.M. „PPVT-III: Peabody Picture Vocabulary Test," 3rd ed. Circle Pines, Minnesota: American Guidance Service, 1997.

DIBELS, Dynamic Indicators of Basic Early Literacy Skills (dibels.uoregon.edu/measures/), 2007.

Frith, C. D., and Frith, U. "Interacting minds—a biological basis," Science, vol. 286, no. 5445, pp. 1692–1695, 1999.

Graf, W., and Krueger, H. "Ergonomic evaluation of user-interfaces by means of eye-movement data," In M. J. Smith, and G. Salvendy (eds.) Work with Computers: Organizational, Management, Stress and Health Aspects, Elsevier Science Publishers, B.V., Amsterdam, pp. 659-665, 1989.

Hess, E.H. "Pupillometrics," In Greenfield, N.S. and Sternbach, R.A. (Eds.), Handbook of Psychophysiology, pp. 491-5310, New York: Holt, Rinehart and Winston, 1972.

Hsiao, J.H., and Cottrell, G. "Two fixations suffice in face recognition," Psychological Science, vol. 19, no. 10, pp. 998-1006, 2008.

Inhoff, A. W., and Radach, R. "Definition and computation of oculomotor measures in the study of cognitive processes," In G. Underwood (Ed.), Eye guidance in reading, driving and scene perception. New York: Elsevier, pp. 29-53, 1998.

Itti, L., Dhavale, N., and Pighin, F. "Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention," In: Proc. Intl. Sym. on Opt. Science and Tech., vol. 5200, pp. 64-78, 2003.

Jacob, R. J. K. "Eye Tracking in Advanced Interface Design," in Advanced Interface Design and Virtual Environments, ed. W. Barfield and T. Furness, Oxford University Press, Oxford, 1994.

Jones, W., Carr, K., and Klin, A. "Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder," Arch. Gen. Psychiatry, vol. 65, no. 8, pp. 946-954, 2008.

Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G., and Moore, K.S. "The mind and brain of short-term memory," Annual Review of Psychology, vol. 59, pp. 193-224, 2008.

Joseph, R.M., and Tanaka, J. "Holistic and part-based face recognition in children with autism," Journal of Child Psychology and Psychiatry, vol. 44, no. 4, pp. 529–542, 2003.

Kahneman, D. "Attention and effort," Englewood Cliffs, New Jersey: Prentice Hall, Inc. http://www.princeton.edu/~kahneman/publications.html., 1973.

Karson, C. N. "Spontaneous eye-blink rates and dopaminergic systems," Brain, vol. 106, pp. 643-653, 1983.

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," Archives of General Psychiatry, vol. 59, no. 9, pp. 809-816, 2002.

Libby, W. L., Lacey, B. C., and Lacey, J. I. "Pupillary and cardiac activity during visual attention," Psychophysiology, vol. 10, no. 3, pp. 270-294, 1973.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. Jr, Leventhal, B. L., DiLavore, P. C., Pickles, A., Rutter, M. "The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism," J. Autism Dev Disord., vol. 30, no. 3, pp. 205-23, 2000.

McIntosh, D.N., Reichmann-Decker, A., Winkielman, P., and Wilbarger, J.L. “When the social mirror breaks: Deficits in automatic, but not voluntary mimicry of emotional facial expressions in autism,” Developmental Science, vol. 9, pp. 295–302, 2006.

NRC, “Educating children with autism,” National Academy Press, Washington, DC, 2001.

Parsons, S., Mitchell, P., and Leonard, A. “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.

Partala, T., Jokiniemi, M., and Surakka, V. “Pupillary Responses to Emotionally Provocative Stimuli,” Proceedings of the 2000 symposium on Eye tracking research & applications, pp. 123-129, 2000.

Partala, T., and Surakka, V. “Pupil size variation as an indication of affective processing,” Int. J. Human-Computer Studies, vol. 59, pp. 185-198, 2003.

Pereira, A.F., Yu, C., Smith, L.B., and H. Shen, “A First-Person Perspective on a Parent-Child Social Interaction During Object Play,” Proc. 31st Annual Meeting of Cog. Sc. Society, Amsterdam, 2009.

Picard, R.W. “Future Affective Technology for Autism and Emotion Communication,” Philosophical Transactions of the Royal Society B: Biological Sciences, 2009.

Ruble, L.A., and Robson, D.M. “Individual and environmental determinants of engagement in autism,” J. of Aut. and Dev. Dis., vol. 37, no. 8, pp. 1457–1468, 2006.

Rutherford, M.D., and Towns, M.T. “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.

Rutter, M., Bailey, A., Berument, S., Lord, C., and Pickles, A. “Social Communication Questionnaire,” Los Angeles, California: Western Psychological Services, 2003.

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., Rees, L., and Wheelwright, S. “The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants,” J Child Psychology and Psychiatry, vol. 39, no. 5, pp. 747–753, 1998.

Tartaro, A., and Cassell, J. "Using Virtual Peer Technology as an Intervention for Children with Autism," In: J. Lazar (Ed.), Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations. UK: John Wiley and Sons, Chichester, 2007.

Trepagnier, C., Sebrechts, M. M., and Peterson, R. “Atypical face gaze in autism,” Cyberpsychology and Behavior, vol. 5, no. 3, pp. 213–217, 2002.

Wieser, M. J., Pauli, P., Alpers, G. W., and Muhlberger, A. “Is eye to eye contact really threatening and avoided in social anxiety? – An eye-tracking and psychophysiology study,” J. of Anx. Dis., vol. 23, pp. 93-103, 2009.

Wilbarger, J. L., McIntosh, D. N., and Winkielmanc, P., “Startle modulation in autism: Positive affective stimuli enhance startle response,” Neuropsychologia, vol. 47, pp. 1323–1331, 2009.

Williams, L. M., Senior, C., David, A. S., Loughland, C. M., and Gordon, E. “In Search of the “Duchenne Smile”: Evidence from Eye Movements,” J of Psychophysiology, vol. 15, no. 2, pp. 122-127, 2001.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G.R., and Vogeley, K. “It’s in your eyes: using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience,” Social Cognitive and Affective Neuroscience, vol. 5, no.1, pp. 98-107, 2010.

Wong, T. K. W., Fung, P. C. W., Chua, S. E., and McAlonan, G. M. “Abnormal spatiotemporal processing of emotional facial expressions in childhood autism: dipole source analysis of event-related potentials,” European Journal of Neuroscience, vol. 28, no. 2, pp. 407 – 416, 2008.

CHAPTER V

VIRTUAL REALITY BASED GAZE-SENSITIVE SYSTEM WITH ADAPTIVE RESPONSE TECHNOLOGY: SYSTEM DEVELOPMENT

Introduction

The objective of this chapter is to present the detailed design specifications of the Virtual-Reality based gaze-sensitive system with Adaptive Response Technology for social communication for children with ASD. Our research as described in Chapter III show the capability of VR-based system to present social communication tasks to the children with ASD and systematically manipulate specific aspects of social communication. In addition, the evaluation of the design was carried out through an experiment to combine the ratings on the affective states of anxiety, engagement, and enjoyment/liking, reported from a clinical observer with the physiological responses of the participants, both being collected when the participants participate in social tasks with the avatars in the VR environment. The evaluation results demonstrate the feasibility of VR-based social communication to cause variations in both the affective states of the participants as reported by the clinical observer, and the physiological responses of the participants. Further, our research, discussed in Chapter IV, indicates the feasibility of designing a VR-based gaze-sensitive system which quantifies the gaze patterns of a child with ASD detected in real-time during virtual social interaction and utilizes this data to provide individualized feedback. In addition, our analysis reveals the ability of such a system to improve the participants' engagement level, and influence their dynamic behavioral viewing patterns as a result of this individualized feedback.

Thus given the promise of VR-based gaze-sensitive social interaction to influence one's affective states, behavioral viewing patterns, and performance in the social task, the development of a VR-based gaze-sensitive social interactive system that can integrate the objective metrics and adapt itself to promote improved social communication skills among the children with ASD is critical. Specifically, such a system must be capable of objectively identifying and quantifying the dynamic viewing patterns, subtle changes in eye physiological responses in real-time, and performance metric of a participant and adaptively responding in an individualized manner. Motivated by this need, the objective of our present research is to develop a Virtual Interactive system with Gaze-sensitive Adaptive Response Technology that can seamlessly integrate VR-based tasks with eye-tracking techniques to encourage a participant to engage in social communication tasks while maintaining the niceties of social interactions. By this we hope to foster improved social communication skills among the participants in an individualized manner, and adaptively encourage the participants to improve his/her level of engagement and performance during social interaction.

Such a system could provide valuable information to caregivers and clinicians about the specific aspects of social communication. In addition, this will provide an integrated computer and eye physiological profiling system which may serve as a tool for designing intervention strategies. In the future, such an integrated intelligent system could be effective for use in developing a more comprehensive adaptive controlled environment that can systematically manipulate various aspects of social communication and thereby help individuals to explore social interaction dynamics gradually and automatically, while improving their engagement level and performance during social interaction task.

Thus, this would serve as an adaptive technology-assisted tool to encourage social communication. In the future, an autism intervention paradigm could use this system as a tool for adaptively responding to the systematically manipulated effects of elements of social interaction that lead to struggles in social communication in children with ASD.

This chapter presents the design and development of the dynamic closed-loop VR-based gaze-sensitive adaptive response technology system. This system has five main subsystems: (i) a VR-based social communication task module, (ii) a real-time eye-gaze monitoring module, (iii) a real-time peripheral physiological signal acquisition module, (iv) a behavioral engagement prediction module, and (v) an integration module that establishes communication between the VR-based task presentation module and the real-time eye-gaze monitoring module to provide individualized adaptive response utilizing a rule-governed intelligent behavioral engagement prediction module.

VR-based Social Communication Task Module

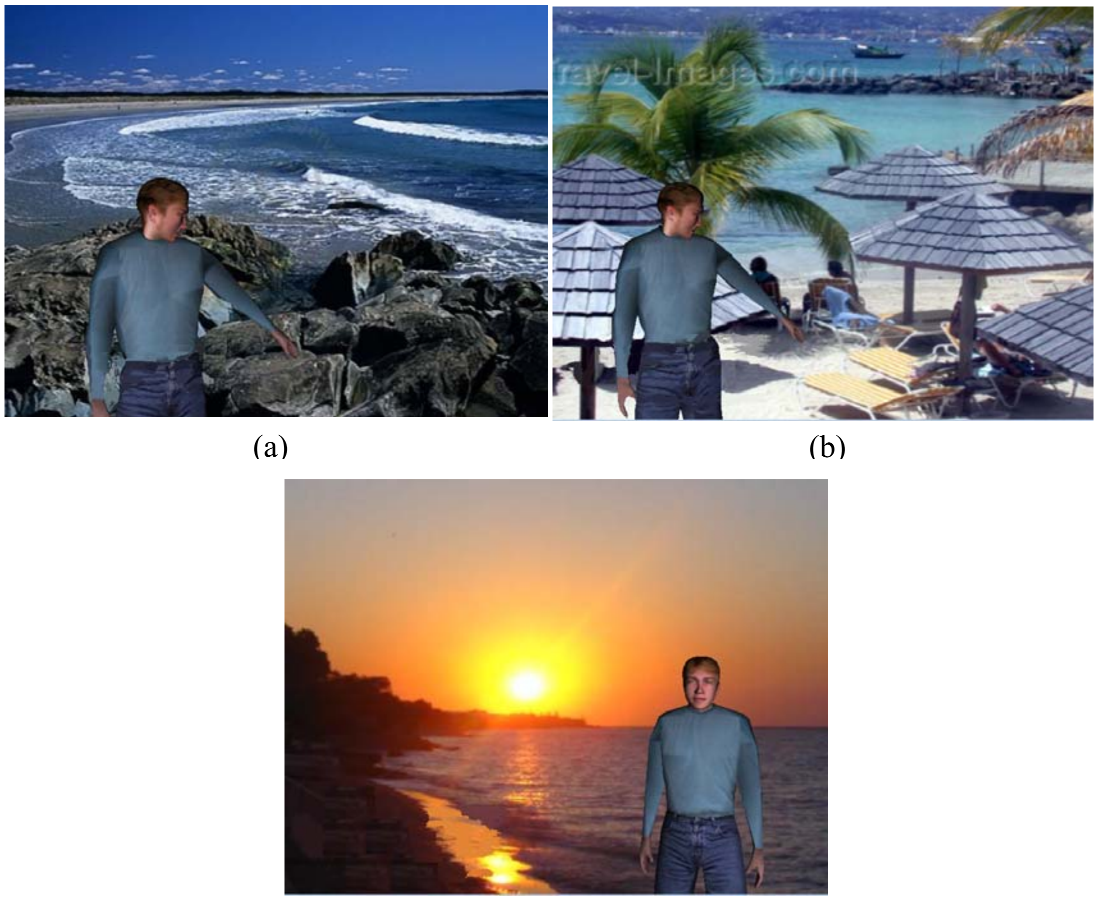
In this work, we use desktop VR applications, because it is accessible, and affordable (Cobb, et al., 1999). For ASD intervention, VR is often effectively experienced on a desktop system using standard computer input devices (Parsons and Mitchell, 2002). Vizard (www.worldviz.com), a commercially available VR design package, is used to develop the virtual environments and the assistive technology. Vizard, VR Toolkit (Enterprise edition) allows for intense access to levels of programming control such that realistic avatars, virtual social scenarios, and interactions can be designed. However, Vizard comes with a limited number of avatars, virtual objects, and scenes that can be used to create a story in VR. Thus, a number of enhancements were made on the VR-platform. In order to perform socially interactive tasks with children with ASD, we developed more extensive social situations with context-relevant backgrounds, and avatars whose age and appearance resemble those of the participants' peers without trying to achieve exact similarities. Also, for effective bidirectional social communication between the avatars and the participants, we developed conversation threads so that the participants can socially interact with the avatars while retrieving a targeted piece of information. Our social communication task module comprises of (i) a task presentation module, and (ii) a bidirectional conversation module.

(i) Design of VR-based Task Presentation Module

In the VR-based task presentation module, an avatar narrates his/her personal experience to the participant while making pointing gestures and moving dynamically in a context-relevant virtual environment.

Specifications of Social Situations with Context-Relevant Backgrounds

In this work we developed 24 social task presentation modules with avatars narrating personal stories to the participants. The personal stories that the avatars share with each participant are based on diverse topics of interest to teenagers e.g., favorite sport, best friend, memorable day in life, field trip with classmates, experience on film, and travel with family. These stories were adopted from an online database (http://www.allfreeessays.com/) of term papers which contains thousands of quality essays, book reports, and research papers written by teenagers. The voices for the avatars were gathered from teenagers from the regional area. We developed 24 different social situations where avatars narrated their personal experience, with the stories forming the context of their narrations. In each social situation, an avatar carried out one-on-one interaction with a participant. We developed three context-specific backgrounds relevant to the social situation being narrated by the avatar for each social task presentation module.



(a) (b)

(c)

Figure V-1. Snapshot of an avatar narrating his tour experience during his visit to a sea beach with (a) the VR environment displaying the view of the rocky beach in the background with the avatar pointing to the rocks, (b) the VR environment showing the view the beach where people lie down for tanning, and (c) the VR environment displaying the view of sunset on the beach.

For example, when an avatar narrates his tour experience to a sea beach in Martinique and introduces the participant to the rocky beach while narrating the rocks on the beach, the VR environment reflects the view of the beach (Figure V-1a). When the avatar narrates some of his favorite activities on the beach such as, tanning during the day, the

VR world displays such a situation to the participant (Figure V-1b). Subsequently, when the avatar narrates his experience of the remarkable view of sunset he witnessed on the beach, the VR situation changes with a smooth transition of the background image to display such a situation to the participant (Figure V-1c). To achieve this, we created a database of 72 context-relevant backgrounds for all the 24 stories which are mounted on the VR world with the avatar superimposed on the virtual environment. This helped us to create realistic social situations relevant to the topic being narrated by the avatar and thereby expose the participant to real-life social scenarios.

Avatar Selection and their Interaction with the participant while Moving Dynamically within the Virtual Environment

The humanoid avatars used in this work have fixed male or female body (supplied by Vizard). New avatar heads, as used in our previous research work (Chapters III and IV) are used in this work. We used 12 avatar heads (6 each for male and female) distributed randomly over the 24 task presentation modules with each avatar appearing twice. These heads were created from 2D photographs of teenagers, which were then converted to 3D heads by '3DmeNow' software for compatibility with Vizard. These new avatar heads were used to create avatars: (i) with age range close to our participant pool's peers, and (ii) with more authentic facial features (e.g., realistic brow line, nose dimensions, etc.) allowing the interaction to be interpreted as realistically as possible. One can view the avatars within the virtual environment from first-person perspective while the avatars narrate personal stories, which is comparable to research on social anxiety and social conventions (Pereira et al., 2009).

The participants are instructed to watch and listen as the avatar tells a story. The avatars are lip-synched with the recorded sound files by using a Vizard based ‘speak’ module. While the avatar narrates his/her personal story, the avatar’s eyes are made to blink randomly with an interval between 1-2 s to render automatic animation of a virtual face similar to the work of Itti et al. (Itti et al., 2003). Also, the avatar displays a normal eye contact which is a mix of 30% straight gaze and 70% averted gaze (Argyle and Cook, 1976; Colburn, Drucker, and Cohen, 2000). Straight gaze means looking straight ahead. Averted gaze means looking alternately to the left, right, and up more than 10° away from center in evenly-distributed, randomly-selected directions (Garau et al., 2001; Jenkins, Beaver, and Calder, 2006). Thus, to display normal eye contact, the avatar looks straight ahead 30% of the time and looks alternately to left, right and up the remaining 70% of the time. In addition, the avatar moves dynamically in the virtual world making pointing gestures such as, pointing his/her hand, rotating his/her head towards the object being narrated. For example, when an avatar narrates his tour experience to a sea beach, and describes the remarkable view of the rocky beach, the avatar turns his head and his hand to point towards the rocks on the sea beach (Figure V-1a). The avatar changes its 3D configuration in the virtual world by using Vizard based ‘walkTo’ module. Also, the avatar is programmed to demonstrate the niceties of social communication, such as, waving of hands and making friendly gestures while introducing himself/herself to the participant. With these features being added to the avatar, the interaction of the avatar with the participant, in the VR world, appear as realistic.

(ii) Design of Bidirectional Conversation Module

The VR-based task presentation is followed by a bidirectional social conversation.

This module encourages the participant to retrieve a targeted information from the avatar by interacting socially with the avatar.

Design Specifications of Bidirectional Conversation Module

The participant is asked to listen and watch the avatar, narrating personal story during the VR-based task presentation. At the end of this task presentation, the participant is asked to extract a piece of information from the avatar. The topic of the target piece of information that the participant is asked to extract from the avatar can be either ‘benign’, or ‘projected contingent’ (i.e., not directly narrated in the presentation by the avatar), or ‘sensitive’ (e.g., one’s personal feeling, or behavior, etc.) depending on the degree of interaction difficulty (discussed below). This is followed by a number of questions/statements for the participant to ask/discuss with the avatar. These appear as a menu of choices and displayed as a transparent text box on one half of the screen with the avatar at the other half of the VR screen.

For example, after an avatar narrates her experience of watching car racing during the VR-based task presentation, the participant (named as ‘Andrew’) is asked to find a target piece of information from the avatar (named as ‘Tonia’) using the bidirectional conversation module. Thus the participant is asked to find the avatar’s experience while getting her driver’s license. This is followed by a menu of 3 choices (Figure V-2) for the participant to ask the avatar.

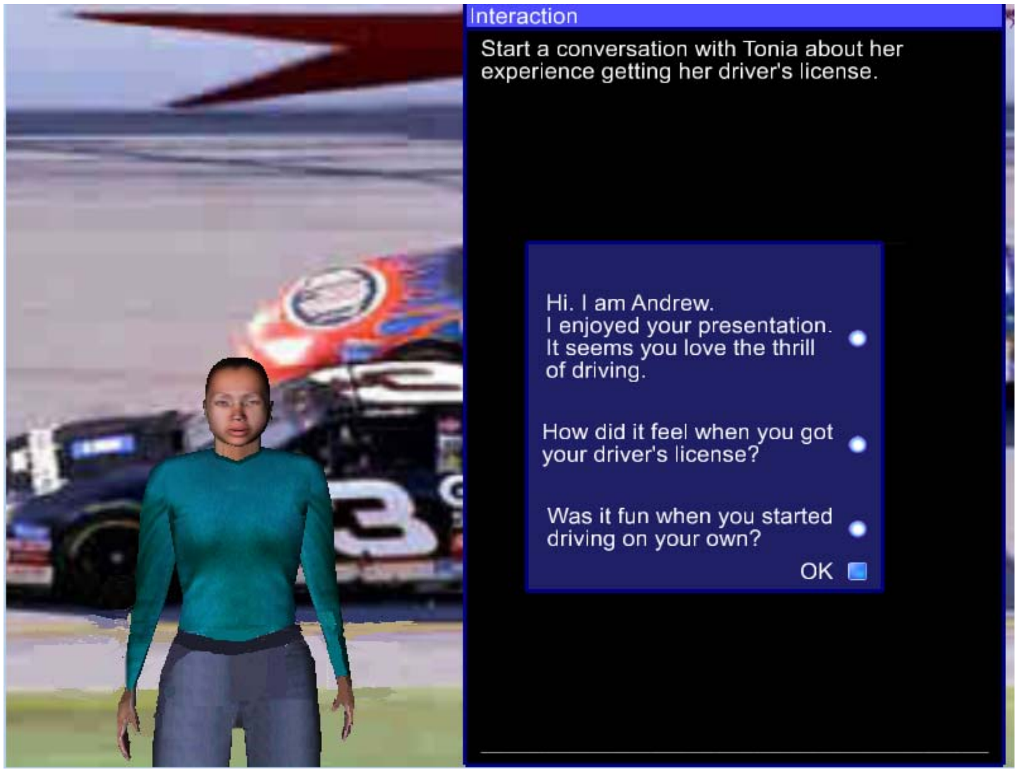


Figure V-2. Snapshot of a bidirectional conversation module with a participant (Andrew) provided a menu of choices to converse with the avatar (Tonia).

After the participant selects a choice by clicking on the radio button accompanying the choice followed by clicking the OK button to submit the selected choice, the menu of choices disappears and the avatar responds to the question asked by the participant by speaking out his/her response (Figure V-3). This continues till the end of the conversation between the avatar and the participant. The menu of choices is framed in such a way so that the participant is required to select the choices in a particular sequence to gain the target piece of information. The participant selects option choices using a mouse.

Figure V-3. Snapshot of a bidirectional conversation module with the avatar responding to the participant's selected choice.

Design Specifications of the Degree of Interaction Difficulty for the Bidirectional Conversation Module

The degree of interaction difficulty while designing the bidirectional conversation module depends on two factors: 1) the nature of the target piece of information that the participant is asked to extract from the avatar and 2) the number of option choices (e.g., questions and/or introductory statements) that the participant have to select to carry out the conversation with the avatar in order to retrieve the target piece of information. Specifically, the nature of the topic of the target piece of information that a participant is asked to extract from an avatar can be 'benign' or, 'projected contingent' or, 'sensitive'. Also, the number of option choices that the participant is required to choose can be 3 or, 5 or, 7. Each of the bidirectional conversation modules within a particular difficulty level is designed to follow a specific structure for the flow of conversation threads to ensure consistency among the bidirectional conversation modules.

- *Easy Level of Interaction Difficulty*

For an easy scenario, a participant is asked to retrieve a 'benign' piece of information from the avatar using the bidirectional conversation module which comprises of a menu of 3 choices to select from. The structure of the conversation flow is represented by the block diagram (Figure V-4).

Introduction with the presenter (avatar). (A) → Asking about the topic of the conversation. (B) → Asking about an overall feeling regarding the topic. (C)

Figure V-4. Block Diagram of the Conversation Threads for Easy Level of Interaction Difficulty.

For example, after an avatar narrates his experience of a football game during the VR-based task presentation, the participant is asked to find from the avatar the experience of his first football game that he played (Figure V-5). Thus for this Easy Level of Interaction Difficulty, the participant (named as 'Andrew') first selects the choice 3 (from the top) of the menu (Figure V-5) to introduce himself to the avatar (named as 'Tom') (represented by (A) in Figure V-4). Tom responds by saying "Hi. I am Tom. Yes. I really love football, especially when I get to play!" Then the participant selects the choice 1 (from the top) of the menu (Figure V-5) to ask the avatar about the topic of the conversation (i.e., regarding the first time Tom played a football game) (represented by (B) in Figure V-4). Tom responds by saying "Of course! I was in the second grade. Our P.E. teacher split our class into two small junior football teams." Finally, the participant selects the choice 2 (from the top) of the menu (Figure V-5) to ask Tom regarding his overall feeling of his first football game (represented by (C) in Figure V-4). Tom responds by saying “Yes, it was a lot of fun to play with my classmates.”

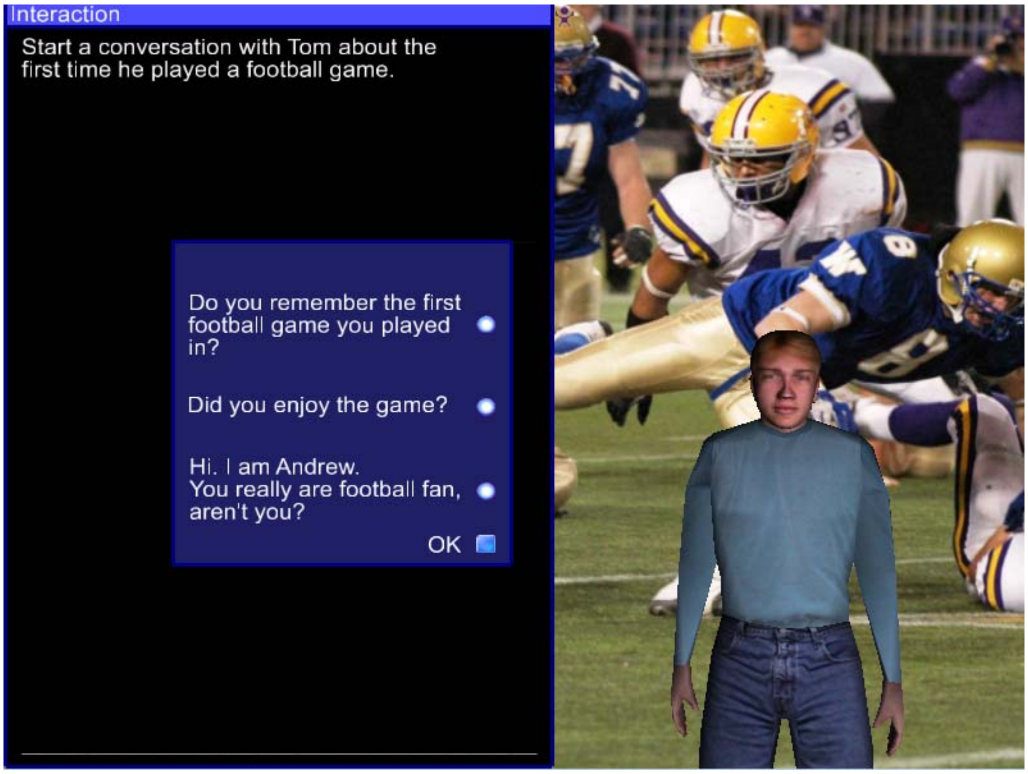


Figure V-5. Snapshot of a bidirectional conversation module for Easy Level of Interaction Difficulty.

- *Medium Level of Interaction Difficulty*

For a scenario with a medium level of interaction difficulty, a participant is asked to retrieve a ‘projected contingent’ piece of information from the avatar using the bidirectional conversation module which comprises of a menu of 5 choices to select from. The structure of the conversation flow is represented by the block diagram (Figure V-6).

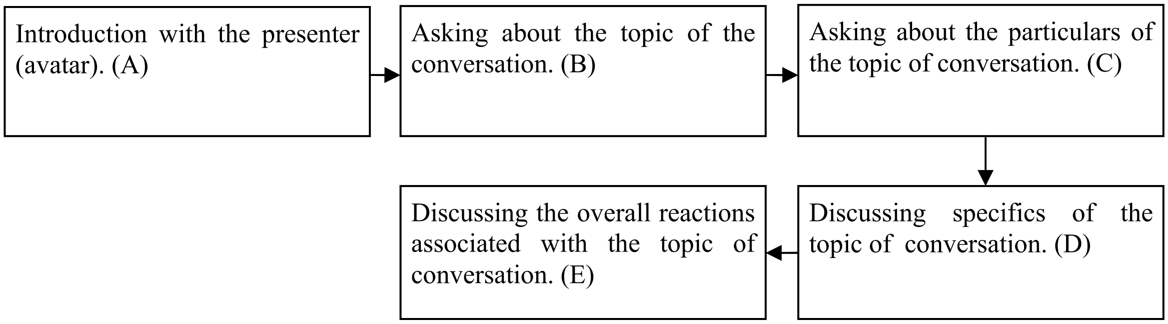


Figure V-6. Block Diagram of the Conversation Threads for Medium Level of Interaction Difficulty.

For example, after an avatar narrates her experience of her vacation while she went to a playground, ice-cream parlor, and a zoo with her friends Cindy and Tracy during the VR-based task presentation, the participant is asked to find out some more details (i.e., 'extended contingent' topic) from the avatar about her experience at the zoo (Figure V-7).

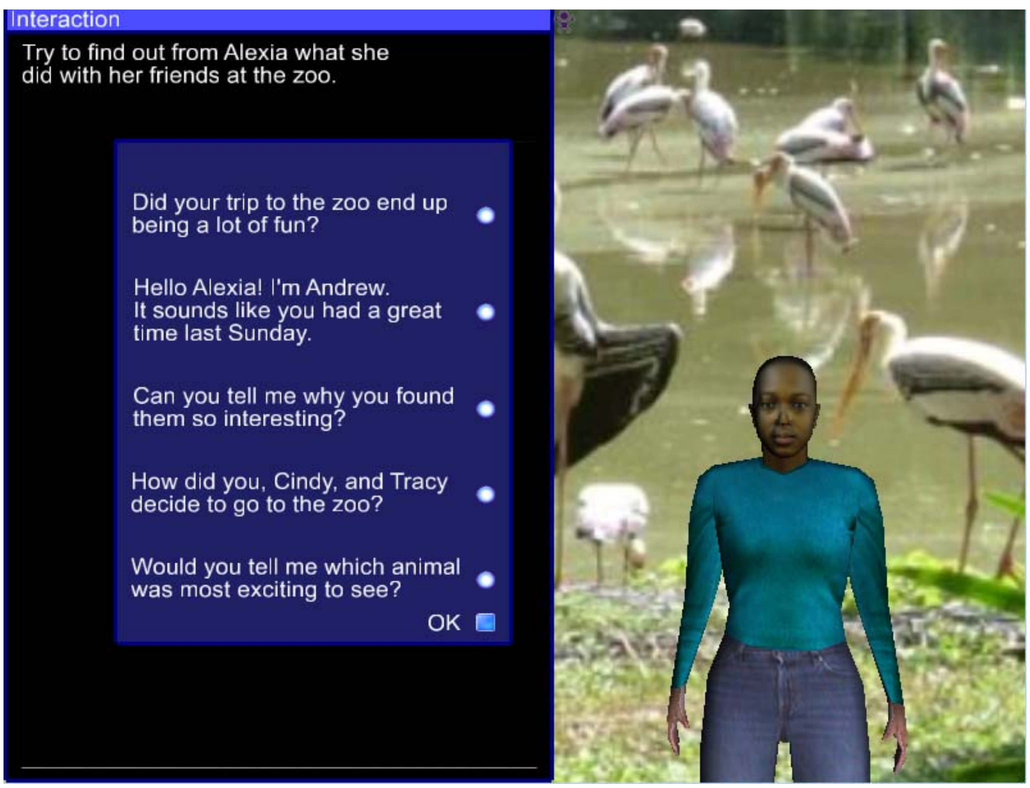


Figure V-7. Snapshot of a bidirectional conversation module for Medium Level of Interaction Difficulty.

Thus for this Medium Level of Interaction Difficulty, the participant (named as

‘Andrew’) first selects the choice 2 (from the top) of the menu (Figure V-7) to introduce himself to the avatar (named as ‘Alexia’) (represented by (A) in Figure V-6). The avatar responds by saying “Hi. It was a super busy but really enjoyable day.” Then the participant selects the choice 4 (from the top) of the menu (Figure V-7) to ask the avatar about the topic of the conversation (i.e., regarding their decision in going to the zoo) (represented by (B) in Figure V-6). In response to this question, the avatar says “Well, we all really like animals and seeing them in their own habitats. Plus this zoo is the only place in America where people can see animals from China.” This prompts the participant to select choice 5 (from the top) of the menu (Figure V-7) to ask the avatar about the particulars of the conversation topic (i.e., regarding any animal at the zoo that seemed exciting to them) (represented by (C) in Figure V-6). The avatar responds by saying “Sure—it was definitely the Giant Pandas! There was a mom and a dad, and three babies.” Then the participant selects choice 3 (from the top) of the menu (Figure V-7) to ask the avatar regarding the specifics of the conversation topic (i.e., regarding the Giant Pandas) (represented by (D) in Figure V-6). The avatar responds “Well, there are only 5 Giant Pandas in the whole United States, and I saw them! Also, they’re interesting because although they look like big teddy bears, they are very aggressive.” Finally, the participant selects the choice 1 (from the top) of the menu (Figure V-7) to ask the avatar regarding her overall reactions (represented by (E) in Figure V-6). The avatar ends the conversation by responding as “It really was. I had a great time with my friends and it was exciting to see the pandas. I can’t wait to go again!”

- *High Level of Interaction Difficulty*

For a scenario with a high level of interaction difficulty, a participant is asked to retrieve a 'sensitive' piece of information from the avatar using the bidirectional conversation module which comprises of a menu of 7 choices to select from. The structure of the conversation flow is represented by the block diagram (Figure V-8).

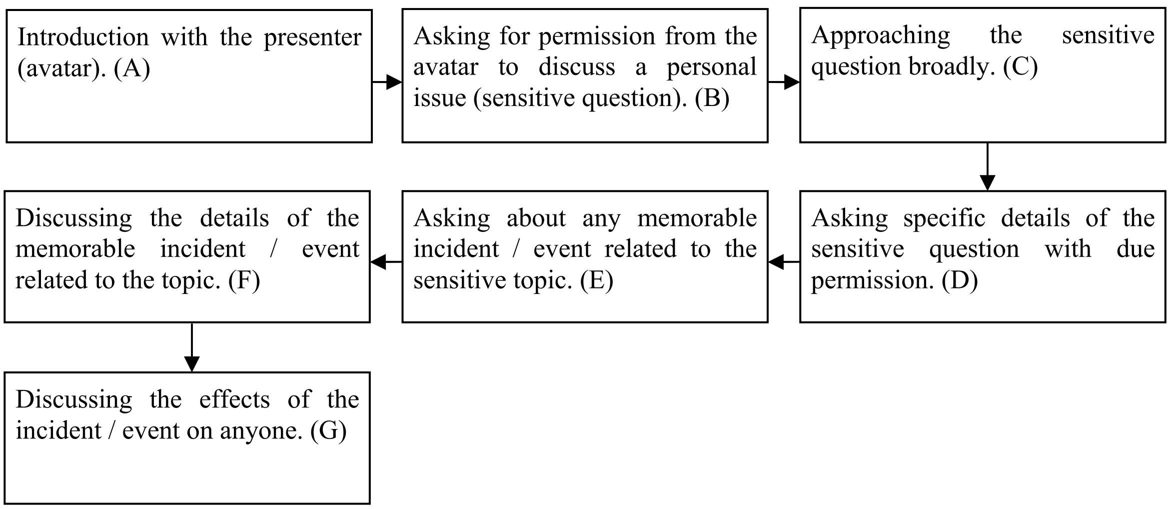


Figure V-8. Block Diagram of the Conversation Threads for High Level of Interaction Difficulty.

For example, after an avatar narrates her experience of playing softball with her best friend Lyndsey and her not liking the softball coach, during the VR-based task presentation, the participant is asked to find out reason for the avatar's not liking the softball coach (i.e., 'sensitive' topic) (Figure V-9). Thus for this High Level of Interaction Difficulty, the participant (named as 'Andrew') first selects the choice 7 (from the top) of the menu (Figure V-9) to introduce himself to the avatar (named as 'Karen') (represented by (A) in Figure V-8). The avatar responds by saying "Sure, I'm glad you liked it. I love my best friend, Lyndsey. Both of us love playing softball." Then the participant selects the choice 5 (from the top) of the menu (Figure V-9) to ask permission from the avatar to discuss a sensitive topic (represented by (B) in Figure V-8). In response to this question, the avatar says "OK. What is it about?" After getting due permission from the avatar, the participant selects choice 3 (from the top) of the menu (Figure V-9) to ask the avatar about the sensitive topic broadly (i.e., regarding the reason behind her not liking the softball coach) (represented by (C) in Figure V-8).

Figure V-9. Snapshot of a bidirectional conversation module for High Level of Interaction Difficulty.

The avatar responds by saying "Sure. You see, our coach was really strict and he didn't spend much time talking to the team. He had a bad habit of shouting at the players on the field. My best friend, Lyndsey is really sensitive and I am too, so we didn't like our coach that much." This prompts the participant to select choice 1 (from the top) of the menu (Figure V-9) to ask the avatar regarding the specific details of the sensitive topic (represented by (D) in Figure V-8). The avatar responds "Usually I'd get really nervous when he yelled and I'd worry I was making a mistake." On hearing that the avatar used to become nervous when the coach yelled at her, the participant selects the choice 2 (from the top) of the menu (Figure V-9) to know about any particular incident related to the sensitive topic (represented by (E) in Figure V-8). The avatar responds by saying "Oh yes. Lots of times. But one time in particular I got really upset." This leads to the next question represented by choice 4 (from the top) of the menu to be selected by the participant to ask the avatar regarding the details of the particular incident when the coach shouted at her (represented by (F) in Figure V-8). In reply to the participant's question, the avatar says "Not at all. I was trying to pitch the ball, but I'm not the best pitcher. The coach shouted at me with a red face, saying, Hey, don't you know how to throw?" Finally, the participant selects the choice 6 (from the top) of the menu (Figure V-9) to ask the avatar regarding the effects that incident had on her (represented by (G) in Figure V-8). The avatar ends the conversation by responding "Lyndsey saw the whole thing! She made me feel better because she agreed he had been mean. Then, she helped me practice pitching and the next day I was a lot better. Our coach didn't yell at me then!"

Design Specifications of the Feedback Given by the Avatars to Facilitate Participants to Continue Bidirectional Conversation

The bidirectional conversation module in our present work also equips the avatar with an ability to execute the role of a facilitator to help the participant to carry on the conversation. For example, with reference to Figure V-5, where the participant is asked to find out from the avatar regarding the experience of the first football game that the avatar had, if the participant ('Andrew') starts the conversation with choice 3 (i.e., an introductory question), then the avatar ('Tom') says "Hi. I am Tom. Yes. I really love football, especially when I get to play!" However, instead of selecting choice 3, if the participant makes an irrelevant choice (e.g., choice 1 or choice 2), then the avatar gives feedback to the participant, saying "I'm sorry, do I know you? Maybe we should introduce ourselves." Thus, the avatar also plays the role of a facilitator during the VR-based conversation. After the introduction is complete, if the participant selects choice 1, instead of choice 2, then the avatar says "It sounds like you want to know about a time I played football. But you haven't asked me about that yet."

Real-time Eye-gaze Monitoring Module

The system captures eye data of a participant interacting with an avatar using Eye-Tracker goggles from Arrington Research (http://www.arringtonresearch.com/). This eye-tracker comes with some basic features (e.g., acquiring raw pupil diameter (PD), raw pupil aspect ratio (PAR), etc.) acquiring capability for offline analysis. In addition, this eye-tracker comes with a Video Capture Module with a refresh rate of 30Hz to acquire a participant's gaze data using the 'Viewpoint' software. We designed the Viewpoint-Vizard handshake module as discussed in Chapter IV to serve as an interface between the two programming platforms. We acquired the raw gaze data using Viewpoint, transformed it to the Vizard compatible format using the handshake interface at a refresh rate of 30 Hz. Subsequently, we applied signal processing techniques, such as windowing, thresholding, etc. to eliminate noise and extract the relevant features.

Gaze Data Processing and Feature Extraction

Raw eye gaze data was acquired with a sampling rate of 30 Hz while the participant wore the eye-tracker goggles and interacted with the avatar during the social communication task. This raw data was subsequently processed to extract the features (Figure V-10), such as, Mean Pupil Diameter (PD_{MEAN}), Mean Blink Rate (BR_{MEAN}), Mean Fixation Duration (FD_{MEAN}), and the Region of Interest (ROI) being looked at by the participant during the interaction.

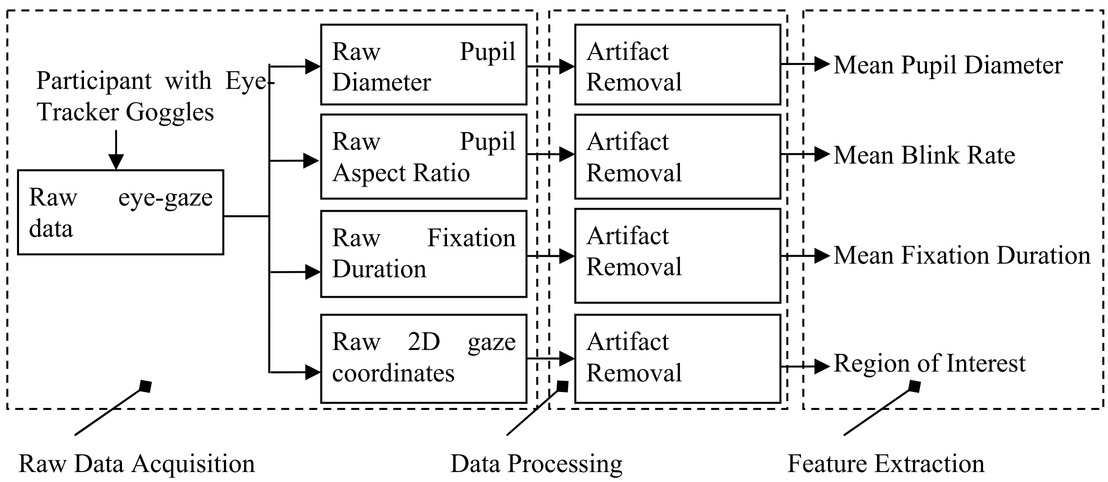


Figure V-10. Block Schematic of Eye-gaze data acquisition and feature extraction.

The raw eye-gaze data acquisition involves the acquisition of gaze data by using the video capture module and the transformation to the Vizard compatible format using the Viewpoint-Vizard handshake module. Thus the parameters of interest are the pupil diameter, the pupil aspect ratio (i.e., the ratio of the major and the minor axes of the pupil image), the fixation duration and the 2D gaze coordinates.

The data processing stage involves the artifact removal.

- For the pupil diameter, artifact removal involves removing the effects due to blinking. Thus value of pupil diameter > 0 is considered.

- We use the pupil aspect ratio (PAR) to compute the blink rate. Although Arrington [Arrington Research, Inc. (2002). Data Collection. In ViewPoint EyeTracker®: PC-60 Software User Guide (pp.47). Scottsdale, Arizona: Arrington Research, Inc.] mentions that blinks can be computed by monitoring the PAR data, Viewpoint software does not provide direct measurement of blink rate. In the present work, we computed the blink rate by considering the number of times the PAR value falls below the lower threshold of 0.5 within a window width of 1 minute. This threshold value for PAR was chosen after several trial test runs detecting blink rate with an accuracy of ±0.05%.
- The recorded data on fixation duration (FD) is first filtered to remove the artifacts due to blinking and noise spikes are eliminated by thresholding. This incorporates filtering the raw data by a moving window having the lower and upper amplitude thresholds of 200 and 450 ms respectively. There are different views on fixation durations with respect to visual stimuli. In one study (Jacob, 1994), fixations have been stated to typically last between 200-600 ms, where blinks of up to 200 ms may occur during a fixation without terminating it and a window of 50 ms lying outside 1^0 of the current fixation has been considered to terminate a fixation. Some researchers have advised to set the lower threshold for fixation as 100 ms (Inhoff, and Radach, 1998). Still others have classified short fixations with FD < 240 ms and long fixations with FD > 320 ms (Graf and Kruger, 1989). In the present study, we compute the fixation duration by using a thresholding window of 200 ms as the lower limit to eliminate the blinking effects and 450 ms as the upper threshold (i.e., up to 1.5

standard deviations from the lower threshold), the reliable data range restricted by noise due to glare effects of cameras of the eye-tracker that we use.

- The 2D gaze coordinates (*x,y*) of the participant's viewing of the presented visual stimulus are recorded. We first remove the points whose coordinates lie beyond the visual stimulus screen. Then our algorithm determines whether one's gaze coordinates correspond to our task-specific segmented regions of the visual stimulus presented to participants. In this present work, we are mainly interested to encourage a participant to interact with an avatar in a socially appropriate way while paying due attention towards the avatar during conversation. This is important as, previous research has indicated atypical visual scanning pattern of children with ASD while viewing the face and the non-face stimuli, in which they tend to look less towards the face (Anderson, Colombo, and Shaddy, 2006) than the non-face objects. Thus, we segment our visual stimulus into two broad regions, face region (Face_ROI) of the avatar and the non-face region (i.e., the entire presented visual stimulus without the face of the avatar).

Design Specifications of the Feedback Given by the System based on the Viewing Pattern of the Participants during Social Conversation

Based on the dynamic fixation pattern of the participant while conversing with the avatar using the bidirectional conversation module, the VR-based gaze-sensitive adaptive response technology provides feedback to the participant at the end of the interaction during each social communication task. Our previous research (Chapter IV) has demonstrated that gaze-based individualized feedback contributes to improving the engagement level and the behavioral viewing pattern of children with ASD. Our present system also has the capability of providing gaze-based individualized feedback (Table V-1).

Table V-1. Rationale behind Gaze-based Feedback.

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

t : Fixation Duration (as a percentage of the total viewing time) of participant's looking towards the Face region of visual stimulus during conversation.

Real-time Peripheral Physiological Signal Acquisition Module

The real-time peripheral physiological signal acquisition module is also one of the sub-systems of our present system. This system is capable of capturing event-marked synchronized peripheral physiological responses of the participants while they participate in the social communication task with the avatars. We do not feedback the inference from the peripheral physiological signals in our present work. Instead we analyze these signals off-line to show the physiological features which are most sensitive to the level of engagement of the participants. Thus this can be a step towards more effective fusion of sensory signals to enable more robust mapping of physiology with one's engagement and thereby help to develop an improved physiology-based behavioral profiling system.

We acquire the physiological signals when a participant interacts with the VR-based social communication task. The VR Task Computer (Figure V-11) is dedicated to the VR-based social communication task. This transmits task-related event-markers to the parallel port of a Physiological Data Acquisition Module which also collects the peripheral physiological signals of the participant during his/her interaction with the VR-based social task. The physiological signals along with the task-related event-markers are acquired and stored by a Physiological Data Logger Computer via an Ethernet Port.

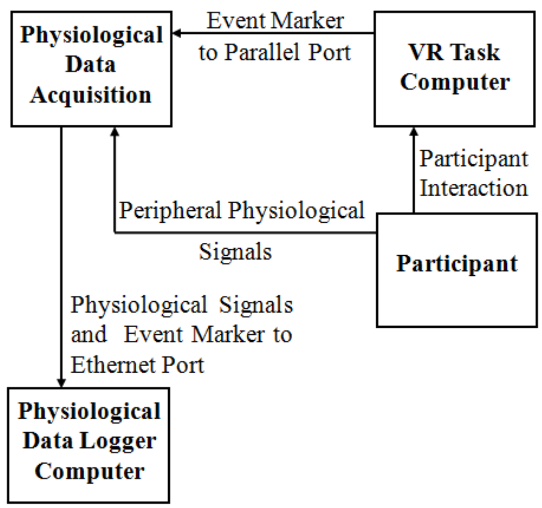


Figure V-11. Block schematic of Real-time acquisition of peripheral physiological signals.

The system acquires peripheral physiological signals using the Biopac MP150 physiological data acquisition system (www.biopac.com). The peripheral physiological signals that we acquire are broadly classified as Cardiovascular activities including electrocardiogram (ECG), impedance cardiogram (ICG), photoplethysmogram (PPG), and phonocardiogram (PCG)/heart sound; Electrodermal activity (EDA) including tonic and phasic responses from skin conductance; Electromyographic activities from corrugator supercilii, zygomaticus major, and upper trapezius muscles; and Peripheral Temperature. The sampling rate is 1000 Hz for all the channels. Appropriate amplification and band-pass filtering are performed. These signals are processed to extract features (detailed description in Chapter III).

Behavioral Engagement Prediction Module

Children with ASD are often characterized by infrequent engagement in social interactions (APA, 1994). Engagement, defined as ‘‘sustained attention to an activity or person’’ (NRC, 2001), is one of the key factors for children with ASD to make substantial gains in communication and social domains (Ruble and Robson, 2006). The engagement of children with ASD is the ground basis for the 'floor-time-therapy' to help them develop relationships and improve their social skills (Wieder and Greenspan, 2005). Thus, if we can engage these children to a social task, then we can teach them social skills. Also, it is well-known that these children demonstrate atypical viewing patterns during social interactions (Rutherford and Towns, 2008) and monitoring eye-gaze can be valuable to design intervention strategies. Several studies have used eye-tracking technology to monitor eye-gaze with static faces (Joseph and Tanaka, 2003; Trepagnier, Sebrechts, and Peterson, 2002) along-with off-line analysis while viewing static scene (Klin et al., 2002). Also a recent study (Wilms et al., 2010) has named VR-based gaze-sensitive system as a ‘tool of the trade’ in social cognitive and affective neuroscience. This study has shown that eye-tracking can be used to drive changes in visual behavior of a virtual character in a gaze-contingent individualized manner. Specifically, it indicates that the gaze behavior of a virtual character can be made responsive to a human observer’s gaze position on the visual stimulus screen while being involved in a joint-attention task. We fully recognize that developing a technology simply asking and reinforcing individuals with ASD to look toward a social target may be a limited enterprise and this is not the ultimate goal of the current study. Instead the present research aims to develop a VR-based gaze-sensitive system that can monitor eye-gaze dynamically during a VR-based social communication task, predict one's level of engagement to a social task based on objective metrics such as, dynamic viewing patterns (e.g., fixation duration), eye physiological indices (e.g., blink rate, pupil diameter) and performance measures (e.g., successful / unsuccessful to retrieve a targeted piece of information through a social conversation) to intelligently adapt itself to improve a child's performance in a social communication task.

Fixation Duration as a Predictor of Engagement

Our previous research (Chapter IV) demonstrates the significance of fixation duration while an individual looks towards the face of the communicator during social communication. Also, Jones et al. (Jones, Carr, and Klin, 2008) have showed that one's fixation duration while looking towards the face region of a speaker indicates social engagement. Fixation duration is a valuable measure, as children with ASD often exhibit lower fixation duration while viewing human faces than the non-human face stimuli (Anderson, Colombo, and Shaddy, 2006) during social interaction. Evidence from literature suggests that, in dyadic communication, eye-gaze information underlying one's expressive behavior (i.e., amount of time a speaker and a listener look at each other) plays a vital role in regulating conversation flow, providing feedback, communicating emotional information, and avoiding distraction by restricting visual input (Argyle and Cook, 1976). For example, a listener looking at the speaker 70% of the time during an interaction has been identified as 'normal while listening' (Colburn, Drucker, and Cohen, 2000; Argyle and Cook, 1976).

In this work, we use certain range of values for the fixation duration as a predictor of one's engagement. To ensure smooth transition from the low engagement to the high engagement state, we also assign a range of values to the engagement label.

Table V-2. Prediction of Engagement from Fixation Duration (FD).

Inference from Fixation Duration	**Engagement Label**
$0\% \leq [((FD_{Face\_ROI})/FD_{TOTAL})*100] \leq 50\%$	1
$50\% < [((FD_{Face\_ROI})/FD_{TOTAL})*100] < 70\%$	2
$[((FD_{Face\_ROI})/FD_{TOTAL})*100] \geq 70\%$	3

FD_{Face_ROI} : Time spent by an individual while looking towards the face region of the avatar.
FD_{TOTAL} : Total time spent by an individual while looking towards the entire presented visual stimulus.

As can be seen from Table V-2, we ascertain a numeric value of 1 to the Engagement Label when a participant's percentage Fixation Duration while looking towards the face of the avatar is between 0 and 50 percent. We give a value of 2, when the participant's percentage Fixation Duration while looking towards the face of the avatar is between 50 and 70 percent. Finally, we give a value of 3, when the participant's percentage Fixation Duration while looking towards the face of the avatar is greater than or equal to 70 percent.

Pupil Diameter as a Predictor of Engagement

Pupil diameter is an important indicator of social engagement with significant pupillary constriction being observed for children with ASD while being engaged in attending to face stimuli (Anderson, Colombo, and Shaddy, 2006). Another study (Gilzernat et al., 2010) has shown the association of reduced pupil diameter with task engagement which they have termed as the phasic mode. Pupil diameter has been described as a reliable and sensitive autonomic measure of attentional engagement and information processing (Anderson, Colombo, and Shaddy, 2006).

As is evident from Table V-3, if the Pupil Diameter (PD) of a participant while interacting with a social situation is greater than that during the previous scenario, we ascertain the lowest value of 1 to the predicted engagement label. Similarly, if one's PD while interacting with the present scenario is less than that during previous situation by 0 to 5 percent, then we use 2 for the engagement label. Further, if the reduction in the PD of a participant from the previous social interaction is greater than 5 percent, then we use the value of 3 for the participant's engagement label.

Table V-3. Prediction of Engagement from Pupil Diameter (PD).

Inference from Pupil Diameter	**Engagement Label**
$PD_{Present} > PD_{Previous}$	1
$PD_{Previous} \geq PD_{Present} \geq 0.95PD_{Previous}$	2
$PD_{Present} < 0.95PD_{Previous}$	3

Blink Rate as a Predictor of Engagement

Blink rate is another important indicator of one's engagement. Literature indicates that there occurs spontaneous inhibition in one's blink rate with increased attentional engagement during visual tasks (Palomba et al., 2000). Some studies have attributed the decrease in one's blink rate with increased engagement to one's attempts to minimize the likelihood of missing important information (Baumstimler, and Parrot, 1971; Kennard, and Glaser, 1964). In a study conducted by Bentivoglio et. al., blink rate for normal subjects was found to decrease from 17 times/min while at rest to 4.5 times/min while being engaged to a reading task (Bentivoglio, et. al., 2004). Increased BR was found in schizophrenic patients in the "relaxed" condition but not in the higher engaged condition (Chen, et. al., 1996). Decreased blink rate was observed by Jensen et al. (Jensen et al., 2009) among children with ASD while being engaged in a task and increased blink rate during task-free condition.

Thus, if the Blink Rate (BR) of a participant while interacting with a social situation is greater than that during the previous scenario, we ascertain the lowest value of 1 to the predicted engagement label (Table V-4). If one's BR while interacting with the present scenario is less than that during previous situation by 0 to 5 percent, we use 2 for the engagement label. Finally, we ascertain a value of 3 to the engagement label if the reduction in the BR of a participant from the previous social interaction is greater than 5 percent.

Table V-4. Prediction of Engagement from Blink Rate (BR).

Inference from Blink Rate	Engagement Label
$BR_{Present} > BR_{Previous}$	1
$BR_{Previous} \geq BR_{Present} \geq 0.95BR_{Previous}$	2
$BR_{Present} < 0.95BR_{Previous}$	3

<u>Integration of VR-based social communication module with the Real-time eye-gaze monitoring module to provide Adaptive Response Technology with Dynamic Decision Task Switching based on Overall Predicted Engagement Level</u>

In recent years, VR has been investigated to promote social interactions in individuals with ASD (Parsons, Mitchell, and Leonard, 2004; Tartaro and Cassell, 2007). These systems are able to adapt tasks based only on performance which is an important aspect of potential VR-based intervention systems for children with ASD. However, such adaptation based solely on task performance limits the individualization of application and likely potential generalization of skills. Specifically, performance based virtual social 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 children 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.

Thus, for effective social communication, the system must be intelligent enough to predict both the behavioral engagement level and the performance of a participant during VR-based social communication to promote an adaptive individualized social skill training paradigm.

Rationale behind the Behavioral Engagement

In this work, we predict one's behavioral engagement level from one's behavioral viewing pattern and eye physiological indices. Thus, we monitor one's real-time fixation patterns, blink rate and pupil diameter to predict the behavioral engagement while being involved with the VR-based social communication task. The logic behind the prediction of one's behavioral engagement level is as follows:

- A participant's behavioral engagement is considered as '*Good Enough*' if the cumulative sum of the engagement label, as obtained by real-time monitoring of his/her fixation duration (Table V-2), pupil diameter (Table V-3), and blink rate (Table V-4) is ≥ 6.
- A participant's behavioral engagement is considered as '*Not Good Enough*' if the cumulative sum of the engagement label is < 6.

Rationale behind the Performance Metric

Our present work considers one's performance metric to be measured from his/her ability to extract the target piece of information from the avatar during the VR-based social interaction while using the bidirectional conversation module.

- A participant's performance in the social communication task is considered as '*Adequate*' if he/she scores ≥ 75% of the total score possible in a social conversation belonging to a particular difficulty level.
- Otherwise, the participant's performance in the social conversation task is considered as '*Inadequate*'.

As a step towards social communication skill training, our system requires the participants to carry out the social conversation with the avatar by following the conversation flow threads, as discussed in Section, '*Design Specifications of the Degree of Interaction Difficulty for the Bidirectional Conversation Module'*. Maximum scores that can be acquired while using the bidirectional social conversation module are 30, 50, and 70 for Easy, Medium, and High Level of interaction difficulty respectively. The scores acquired by a participant decreases progressively if he/she makes irrelevant choices at each turn while conversing with the avatar. For example, as shown in Figure V-5, which represents an Easy Level of interaction difficulty, while starting the conversation with the avatar, the participant makes the relevant choice, i.e., selects choice 3 at the first attempt, he/she scores 10 for that selection. But, if he/she mistakenly selects choice 1 or choice 2, and then selects choice 3 at the second attempt, he/she scores 6, while on making a third attempt he/she scores 2. Similar is the case for the Medium and the High Level of interaction difficulty. However, our algorithm allows 2, 3, and 5 misses for the Easy, Medium, and High Level of interaction difficulty, respectively, after which the task progression switches to the next VR-based social task trial.

Our present work tries to fuse the behavioral engagement level with the performance metrics of an individual during social communication task. The system adapts itself intelligently based on the behavioral engagement level and the performance metrics by utilizing a Dynamic Decision Task Switching Module.

Design of Dynamic Decision Task Switching Module

No existing technology (e.g., VR-based systems, robotic systems) specifically addresses how to autonomously detect and flexibly respond to the affective cue, such as, engagement of children with ASD within an intervention paradigm (Bernard-Opitz, Sriram, and Nakhoda-Sapuan, 2001; Dautenhahn and Werry, 2004; Kozima, Nakagawa, and Yasuda, 2007; Michaud and Theberge-Turmel, 2002; Mitchell, Parsons, and Leonard, 2007; Parsons, Mitchell, and Leonard, 2005; Pioggia et al., 2005; Scassellati, 2005; Strickland, 1997; Swettenham, 1996; Tartaro and Cassell, 2007; Trepagnier, et al., 2006). Affective cue, such as engagement is insight into the behavior of children with ASD, and is one of the key factors for these children to make substantial gains in communication and social domains (Ruble and Robson, 2006). The ability to utilize the power of these cues may permit a smooth, natural, and more productive interaction process (Gilleade, Dix, and Allanson, 2005; Kapoor, Mota, and Picard, 2001; Picard, 1997; Prendinger, Mori, and Ishizuka, 2005) especially considering the core social and communicative vulnerabilities that limit individuals with ASD to accurately self-identify affective experiences (Hill, Berthoz, and Frith, 2004). Common in autism intervention, clinicians who work with children with ASD intensively monitor the engagement of the children in order to make appropriate decisions about adaptations to their intervention. The engagement of children with ASD is the ground basis for the "floor-time therapy" to help them develop relationships and improve their social skills (Wieder and Greenspan, 2005). Also, clinicians look out for task performance metric which is positively correlated to the engagement (Blackorby and Cameto, 2005). Given the importance of affective cues (e.g., engagement) in ASD intervention practice (Ernsperger, 2003; Seip, 1996; Wieder and Greenspan, 2005), predicting one's engagement level from implicit measures (e.g., behavioral viewing fixation pattern and eye physiology) to facilitate bidirectional communication may be critical to encourage a child to improve his/her engagement level and performance in social task.

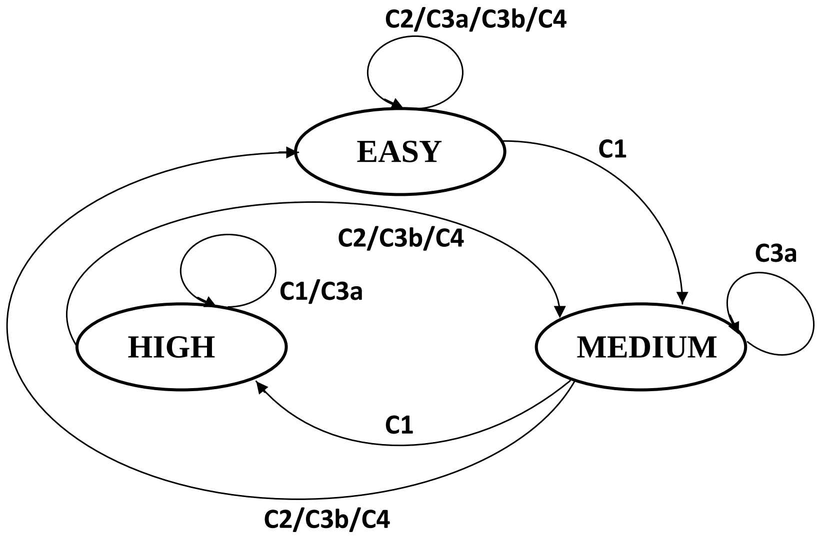
Our present research deals with the development of a Dynamic Decision Task Switching Module that autonomously decides to change the interaction difficulty level with an aim to improve a participant's engagement to the social task. In order to achieve this, we consider one's predicted behavioral engagement, as detected from his/her viewing pattern and eye physiological indices (as discussed above) as '*Good enough*' or '*Not Good Enough*'. In addition, a participant's performance in the virtual social communication task can be '*Adequate*' (e.g., if the participant scores ≥ 75% of the total score possible while extracting intended information from the avatar), otherwise the participant's performance is considered as '*Inadequate*'. Subsequently, we use a rule-governed strategy generator that fuses the information on the predicted behavioral engagement (e.g., '*Good Enough*', or '*Not Good Enough*') and the task performance (e.g., '*Adequate*', or '*Inadequate*') to predict and implement an individualized task modification strategy. The generator has the ability to enhance performance via modifying task difficulty (i.e., increasing/decreasing) and thus provides reengagement strategy in the form of access to preferred level of interaction difficulty for a specific controlled interval of time. The system has the ability to recognize patterns of performance - not just to make a decision based on engagement and performance in one interval. In this way, the system's embodied intelligence is capable of recognizing patterns of success and failure based on its own modifications. The system attempts to promote both engagement and performance, but performance progression is the super-ordinate variable that trumps conflicting decisions in the model and ensure that we do not reward escape/avoidance. We present the individualized task modification strategy based on the composite effect of one's behavioral viewing, eye physiological indices, and performance metric in a social communication task, in Table V-5.

Table V-5. Individualized task modification strategy based on the composite effect of Behavioral Viewing, Eye Physiological Indices, and Performance Metric.

Case No.	**Predicted Behavioral Engagement (from Viewing Pattern and Physiological Indices)**	**Task Performance**	**Predicted Overall Behavioral Engagement**	**Decision taken by Strategy Generator for Task Difficulty Level Modification**
1	Good Enough	Adequate	Engaged	Increase the difficulty level / Maintain at the same difficulty level, if that is the highest.
2	Good Enough	Inadequate	Not Engaged	Decrease the difficulty level / Maintain at the same difficulty level, if that is the lowest.
3	Not Good Enough	Adequate	Semi-Engaged	a) Maintain at the same difficulty level and look for improvement in the next cycle and b) In case of no further improvement, decrease the difficulty level or maintain it, if that is the lowest.
4	Not Good Enough	Inadequate	Not Engaged	Too difficult. So decrease the difficulty level / Maintain it at the same difficulty level, if that is the lowest.

From Table V-5, we find that when a participant's engagement and task performance are sufficient, indicating the overall behavioral engagement as '*Engaged*', then the task progression continues stepwise (i.e., increasing task difficulty after successful completion) to promote continued optimal learning (Case 1). If task performance becomes '*Inadequate*', indicating the overall behavioral engagement as '*Not Engaged*' the task difficulty is lowered (Case 2). If the task performance is '*Adequate*' and the behavioral engagement is '*Not Good Enough*', indicating the overall behavioral engagement as '*Semi-Engaged*' then the performance progression is the super-ordinate variable (Case 3). In this case, the strategy generator maintains the task progression at the same level of difficulty (Case 3a) and look out for an improvement in the next cycle. In case of no improvement, the difficulty level is reduced (Case 3b). Thus, the system has the ability to recognize patterns of one's engagement and performance - not just to make a decision based on engagement and performance in one interval. Further, if both the task performance and the behavioral engagement are '*Inadequate*' and '*Not Good Enough*' respectively, indicating the overall behavioral engagement as '*Not Engaged*', and implying that the task might be too difficult for the participant, then the strategy generator reduces the task difficulty (Case 4).

In the present work we implemented this dynamic task switching module by using a Finite State Machine representation. Finite state machine (Booth, 1967) is a behavior model composed of a finite number of states, transitions between those states, and actions, similar to a flow graph in which one can inspect the way logic runs when certain conditions are met. In our present work, we have three levels of interaction difficulty (such as, Easy, Medium, and High), and the strategy generator provides the logic for the transition from one difficulty level to another. Thus the dynamic task switching used in our present work is represented by the finite state machine representation (Figure V-12).



C1: Case1; C2 : Case2; C3a : Case3a; C3b : Case3b; C4 : Case4

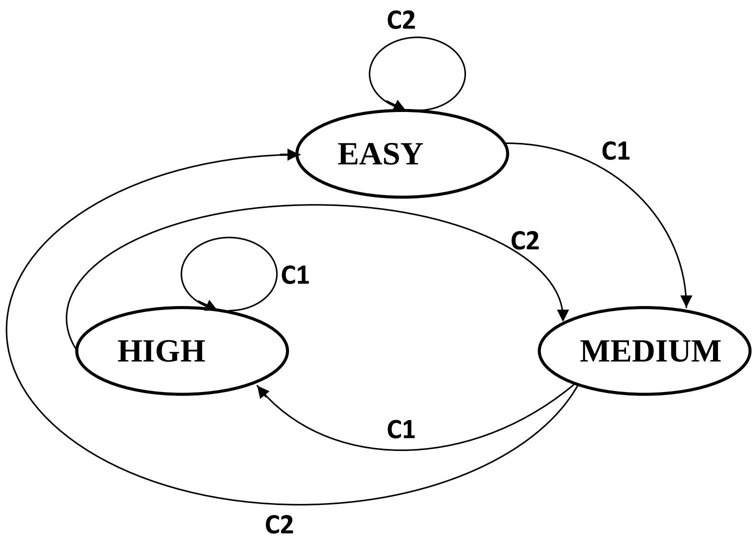
Figure V-12. State Machine Representation of Dynamic Decision Task Switching based on composite effect of one's Behavioral Viewing, Eye Physiological Indices, and Performance Metric.

In the present work, we carried out a comparative analysis between 'a system that predicts social engagement based on the rule-governed composite effect of one's behavioral viewing, eye physiological indices, performance' and 'a system that predicts social engagement based on the performance metric alone'. In order to achieve this, our present system also features switching of task difficulty level based on the performance metric only. We present the task modification strategy based on one's performance metric only in Table V-6.

Table V-6. Task modification strategy based on the Performance Metric.

Case No.	**Task Performance**	**Predicted Overall Engagement**	**Decision taken by Strategy Generator for Task Difficulty Level Modification**
1	Adequate	Engaged	Increase the difficulty level / Maintain at the same difficulty level, if that is the highest.
2	Inadequate	Not Engaged	Decrease the difficulty level / Maintain at the same difficulty level, if that is the lowest.

From Table V-6, we find that when a participant's task performance is '*Adequate*', we consider the overall predicted engagement as '*Engaged*' and the task progression continues stepwise (Case 1). But, if on the other hand, the participant's task performance is '*Inadequate*', which indicates that the participant is '*Not Engaged*', then the system lowers the task difficulty. This task switching is represented by the following State Machine Diagram (Figure V-13).



C1: Case1; C2: Case2

Figure V-13. State Machine Representation of Task Switching based on one's Performance Metric.

Discussion

This chapter presents the detailed design specifications of the developed VR-based Gaze-sensitive Adaptive Response Technology system. This system intelligently fuses the information derived from one's behavioral viewing patterns, variation in one's eye physiological indices and one's performance metric during a VR-based social communication task to predict one's overall engagement to the social task. Based on the predicted overall engagement, the system adaptively responds with an aim to improve one's engagement and performance in the social communication task. By this we hope to foster improved social communication skills among the participants in an individualized manner, and adaptively encourage the participants to improve his/her level of engagement and performance during social interaction. Such a system could provide valuable information to caregivers and clinicians about the specific aspects of social communication. In addition, this will provide an integrated computer and eye physiological profiling system which may serve as a tool for designing intervention strategies.

References

Anderson, C. J., Colombo, J., and Shaddy, D. J. "Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder," J. of Cli. and Exp. Neuropsy., vol. 28, pp. 1238–1256, 2006.

APA, "Diagnostic and statistical manual of mental disorders," (4th ed.), Washington, DC: American Psychiatric Association, 1994.

Argyle, M., and Cook, M. "Gaze and Mutual Gaze," Cambridge, MA: Cambridge Univ. Press, 1976.

Baumstimler, Y., and Parrot, J. "Stimulus generation and spontaneous blinking in man involved in a voluntary activity," J. Exp. Psychol., vol. 88, pp. 95-102, 1971.

Bentivoglio, A.R., Bressman, S.B., Cassetta, E. Carretta, D. Tonali, P., and Albanese, A. "Analysis of blink rate patterns in normal subjects," Movement Disorders, vol. 12, no. 6, pp. 1028-1034, 2004.

Bernard-Opitz, V., Sriram, N., and Nakhoda-Sapuan, S. "Enhancing social problem solving in children with autism and normal children through computer-assisted instruction," Journal of Autism and Developmental Disorders, vol. 31, no. 4, pp. 377-384, 2001.

Blackorby, J. and Cameto, R. "Changes in the School Engagement and Academic Performance of Students with Disabilities," in Engagement and Academics, pp. 1-24, 2005.

Booth, T. "Sequential Machines and Automata Theory," John Wiley and Sons, New York, 1967.

Chen, E.Y.H., Lam L.C.W., Chen, R.Y.L., and Nguyen, D.G.H. "Blink rate, Neurocognitive impairments, and symptoms in Schizophrenia," Biological Psychiatry, vol. 40, no. 7, pp. 597-603, 1996.

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

Colburn, A., Drucker, S., and Cohen, M. "The role of eye-gaze in avatar-mediated conversational interfaces," In SIGGRAPH Sketches and Applications, New Orleans, Louisiana, USA, 2000.

Dautenhahn, K. and Werry, I. "Towards interactive robots in autism therapy: Background, motivation and challenges," Pragmatics and Cognition, vol. 12, pp. 1-35, 2004.

Ernsperger, L. "Keys to Success for Teaching Students with Autism," Future Horizons, 2003.

Garau, M., Slater, M., Bee, S., and Sasse, M. A. "The impact of eye gaze on communication using humanoid avatars," in Proceedings of the SIGCHI conference on human factors in computing systems, pp. 309-316, 2001.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and Cohen, J. D. "Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function," Cognitive, Affective, & Behavioral Neuroscience, vol. 10, pp. 252-269, 2010.

Gilleade, K., Dix, A., and Allanson, J. "Affective videogames and modes of affective gaming: Assist me, challenge me, emote me," in Proceedings of the Digital Games Research Association Conference, 2005.

Graf, W., and Krueger, H. "Ergonomic evaluation of user-interfaces by means of eye-movement data," In M. J. Smith, and G. Salvendy (eds.) Work with Computers: Organizational, Management, Stress and Health Aspects, Elsevier Science Publishers, B.V., Amsterdam, pp. 659-665, 1989.

Hill, E., Berthoz, S., and Frith, U. "Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives," Journal of Autism and Developmental Disabilities, vol. 34, no. 2, 2004.

Inhoff, A. W., and Radach, R. "Definition and computation of oculomotor measures in the study of cognitive processes," In G. Underwood (Ed.), Eye guidance in reading, driving and scene perception. New York: Elsevier, pp. 29-53, 1998.

Itti, L., Dhavale, N., and Pighin, F. “Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention,” In: Proc. Intl. Sym. on Opt. Science and Tech., vol. 5200, pp. 64-78, 2003.

Jacob, R. J. K. “Eye Tracking in Advanced Interface Design,” in Advanced Interface Design and Virtual Environments, ed. W. Barfield and T. Furness, Oxford University Press, Oxford, 1994.

Jenkins, R., Beaver, J. D., and Calder, A. J. “I thought you were looking at me: Direction-specific aftereffects in gaze perception,” Psychological Science, vol. 17, pp. 506-513, 2006.

Jensen, B., Keehn, B., Brenner, L., Marshall, S.P., Lincoln, A.J., and Müller, R.A. “Increased Eye-Blink Rate in Autism Spectrum Disorder May Reflect Dopaminergic Abnormalities,” Intl. Society for Autism Research, Poster Presentation, 2009.

Jones, W., Carr, K., and Klin, A. “Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder,” Arch. Gen. Psychiatry, vol. 65, no. 8, pp. 946-954, 2008.

Joseph, R.M., and Tanaka, J. “Holistic and part-based face recognition in children with autism,” Journal of Child Psychology and Psychiatry, vol. 44, no. 4, pp. 529–542, 2003.

Kapoor, A., Mota, S., and Picard, R. W. “Towards a learning companion that recognizes affect,” in Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition AAAI Fall Symposium, 2001.

Kennard, D.W., and Glaser, G.H. “An analysis of eyelid movements,” J. Nervous Mental Disorder, vol. 139, pp. 31-48, 1964.

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. “Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism,” Archives of General Psychiatry, vol. 59, no. 9, pp. 809-816, 2002.

Kozima, H., Nakagawa, C., and Yasuda, Y. “Children-robot interaction: a pilot study in autism therapy,” Progress in Brain Research, vol. 164, pp. 385-400, 2007.

Michaud, F., and Theberge-Turmel, C. “Mobile robotic toys and autism,” in Socially Intelligent Agents:

Creating Relationships With Computers and Robots, K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds, Eds. Norwell, MA: Kluwer, pp. 125-132, 2002.

Mitchell, P., Parsons, S., and Leonard, A. "Using virtual environments for teaching social understanding to adolescents with autistic spectrum disorders," Journal of Autism and Developmental Disorders, vol. 37, pp. 589-600, 2007.

NRC. (2001). Educating children with autism. National Academy Press, Washington, DC.

Palomba, D., Sarlo, M., Angrilli, A., Mini, A., Stegagno, L. "Cardiac responses associated with affective processing of unpleasant film stimuli," International Journal of Psychophysiology, vol. 36, pp. 45-57, 2000.

Parsons, S., and Mitchell, P. "The potential of virtual reality in social skills training for people with autistic spectrum disorders," Journal of Intellectual Disability Research, vol. 46, pp. 430-443, 2002.

Parsons, S., Mitchell, P., and Leonard, A. "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.

Parsons, S., Mitchell, P., and Leonard, A. "Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments?," Autism, vol. 9, pp. 95-117, 2005.

Pereira, A.F., Yu, C., Smith, L.B., and Shen, H. "A First-Person Perspective on a Parent-Child Social Interaction During Object Play," Proc. 31st Annual Meeting of Cog. Sc. Society, Amsterdam, 2009.

Picard, R. W. "Affective Computing," Cambridge, MA: MIT Press. 1997.

Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., and De Rossi, D. "An android for enhancing social skills and emotion recognition in people with autism," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 13, no. 4, pp. 507-515, 2005.

Prendinger, H., Mori, J., and Ishizuka, M. "Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game," International Journal of Human-Computer Studies, vol. 62, no. 2, pp. 231-245, 2005.

Ruble, L. A., and Robson, D. M. "Individual and environmental determinants of engagement in autism," J. of Aut. and Dev. Dis., vol. 37, no. 8, pp. 1457–1468, 2006.

Rutherford, M. D., and Towns, M. T. “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.

Scassellati, B. “Quantitative metrics of social response for autism diagnosis,” in Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, pp. 585-590, Nashville, Tennessee, 2005.

Seip, J. “Teaching the Autistic and Developmentally Delayed: A Guide for Staff Training and Development,” Delta, BC: Author, 1996.

Strickland, D. “Virtual reality for the treatment of autism,” in Virtual reality in neuropsycho-physiology, G. Riva, Ed., pp. 81-86, Amsterdam: IOS Press, 1997.

Swettenham, J. “Can children with autism be taught to understand false belief using computers?,” Journal of Child Psychology and Psychiatry, vol. 37, no. 2, pp. 157-165, 1996.

Tartaro, A., and Cassell, J. “Using Virtual Peer Technology as an Intervention for Children with Autism,” In: J. Lazar (Ed.), Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations. UK: John Wiley and Sons, Chichester, 2007.

Trepagnier, C., Sebrechts, M.M., and Peterson, R. “Atypical face gaze in autism,” Cyberpsychology and Behavior, vol. 5, no. 3, pp. 213–217, 2002.

Trepagnier, C. Y., Sebrechts, M. M., Finkelmeyer, A., Stewart, W., Woodford, J., and Coleman, M. “Simulating social interaction to address deficits of autistic spectrum disorder in children,” Cyberpsych Behav, vol. 9, no. 2, pp. 213-217, 2006.

Wieder, S., and Greenspan, S. “Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective?,” The Journal of Developmental and Learning Disorders, vol. 9, 2005.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., and Vogeley, K. “It’s in your eyes: using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience,” Social Cognitive and Affective Neuroscience, vol. 5, no.1, pp. 98-107, 2010.

CHAPTER VI

IMPACT OF VIRTUAL REALITY BASED SOCIAL INTERACTIVE GAZE-SENSITIVE SYSTEM WITH ADAPTIVE RESPONSE TECHNOLOGY ON PERFORMANCE AND BEHAVIORAL VIEWING FOR CHILDREN WITH ASD

Introduction

The primary objective of this chapter is to present our findings on the effects of interacting with a Virtual Reality (VR) based gaze-sensitive social communication system equipped with adaptive response technology. A growing number of studies have been investigating the application of VR-based applications to address some of the core deficit areas related to the realm of social communication for children with ASD (Parsons, Mitchell, and Leonard, 2004; Strickland et al., 1996). However, the current VR environments as applied as assistive technologies to tasks involving children with ASD are capable of modifying tasks based only on objective performance characteristics (i.e., correct or incorrect) of responses. Though being able to adapt tasks based on performance is an important aspect of potential VR-based intervention systems for children with ASD, such adaptation based solely on task performance limits the individualization of application and likely potential generalization of skills. Specifically, performance based virtual social interactions do not often involve measurements of or necessitate appropriate subtle, yet critically important, aspects of effective social communication (e.g., eye-gaze, and other forms of social convention). In fact, while many children 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 we focus on the development of a novel VR technology capable of incorporating real-time measurement and flexible adaptation to dynamic gaze patterns of children with ASD. It is a common finding that individuals with ASD often exhibit atypical gaze patterns during social interactions (e.g., greater fixation towards non-social objects than faces) (Cohen, and Volkmar, 1997; Pelphrey et al., 2002). As such, a flexible technology designed to detect, respond to, and potentially enhance appropriate and socially modulated gaze during social interactions could be seen as a tool for potential ASD intervention. Emerging work suggests that integration of a VR-based system with eye-tracking technology appears to be the next logical step towards establishing a gaze-sensitive virtual social interaction. While discussing the importance of such a system, a recent study (Wilms et al., 2010) has named it as a 'tool of the trade' in social cognitive and affective neuroscience. This study has shown that eye-tracking can be used to drive changes in visual behavior of a virtual character in a gaze-contingent individualized manner. Specifically, it indicates that the gaze behavior of a virtual character can be made responsive to a human observer's gaze position on the visual stimulus screen while being involved in a joint-attention task. We fully recognize that developing a technology simply asking and reinforcing individuals with ASD to look towards a social target may be a limited enterprise and this is not the ultimate goal of the current study. Instead the current work represents a first-step in demonstrating the feasibility of potential more complex, sophisticated, robust intervention system designed to detect patterns of gaze, as well as other subtle and necessary components of social communication, in order to develop subtle methods for incorporating these differences in terms of making intelligent and automatic decisions that could be built into complex systems in a virtual environment.

In this chapter, we study the effects of interaction of a group of ASD participants with our designed VR-based gaze-sensitive social communication system equipped with adaptive response technology. Each participant participated in two VR-based social communication tasks on two different sessions. In one session (henceforth referred to as Session1), the participant interacted with the system that adaptively responded based on one's performance metric alone. In the other session (henceforth referred to as Session2), the participant interacted with the system that adaptively responded by predicting one's engagement to the social task, based on the composite effect of one's behavioral viewing, eye physiological indices, and performance metric while participating in the social task. We investigate the effects of interacting with such a system that can intelligently adapt itself based on one's predicted engagement level while participating in the social communication task so far as one's performance and behavioral viewing during the social communication task are concerned.

Experimental Investigation

Participants

A group of 8 adolescents (Male: n=7, Female: n=1) with high-functioning ASD and ages ranging from 13-18 years (Mean = 15.76 years, SD = 1.89 years) participated in this study. Their characteristics are shown in Table VI-1. The majority of male participants is reflective of the autism community, which has been found to have a male to female ratio of 4:1 (Ehlers, and Gillberg, 1993). All ASD participants had a confirmed diagnosis from

Table VI-1. Participant Characteristics. No significant group difference was found for age, and standard score on the PPVT, scores SRS, SCQ, ADOS-G, and ADI-R.

Participant (Gender)	**Age (years)**	**PPVT[a] Standard score**	**SRS[b] Total *T*-score**	**SCQ[c] Total score**	**ADOS-G[d] Total score (cutoff = 7)**	**ADI-R[e] Total score (cutoff = 22)**
Group1						
ASD1 (Male)	17.583	134	80	12	13	49
ASD2 (Male)	16.917	110	73	13	7	33
ASD3 (Male)	14.250	130	89	16	15	34
ASD4 (Male)	13.833	170	92	14	13	53
Group Mean (SD)	15.645 (1.88)	136 (24.98)	83.50 (8.66)	13.75 (1.71)	12 (3.46)	42.25 (10.24)
Group2						
ASD5 (Male)	16.500	92	87	20	-	-
ASD6 (Male)	18.250	97	63	17	9	49
ASD7 (Female)	13.000	133	90	10	7	25
ASD8 (Male)	15.750	126	69	23	11	56
Group Mean (SD)	15.875 (2.18)	112 (20.51)	77.25 (13.28)	17.5 (5.57)	9 (2)	43.33 (16.26)
t-value	0.1839	1.4851	0.7886	1.2878	1.3241	0.1092
p-value	ns	ns	ns	ns	ns	ns
Exact *p*-value	0.8586	0.1881	0.4604	0.2452	0.2428	0.9173

[a]Peabody Picture Vocabulary Test-3rd edition (Dunn and Dunn, 1997)
[b]Social Responsiveness Scale (Constantino, 2002)
[c]Social Communication Questionnaire (Rutter et al., 2003a)
[d]Autism Diagnostic Observation Scale-Generic: Module 3 or 4 depending upon subject's developmental level (Lord et al., 2000)
[e]Autism Diagnostic Interview-Revised (Rutter et al., 2003b)
ns : No Significant group difference.

evaluations by a licensed clinical psychologist using DSM-IV criteria according to their medical records. All but one participant met cutoffs for ASD according to ADOS and ADI-R assessments. ASD5 did not have ADOS or ADI-R records, however his scores on the SRS and the SCQ questionnaires met ASD cutoffs. The participants were categorized in two groups (e.g., Group1 and Group2). Group1 participants were first exposed to VR-based social communication tasks with task-switching based on one's performance metric alone on the first day, followed by VR-based tasks with task-switching based on the composite effects of one's behavioral viewing, eye physiology, and performance metrics on the second day (i.e., Session1-followed by-Session2, as discussed in '*Introduction*'). Group2 participants were exposed to VR-based social tasks in the reverse order, i.e., Session2-followed by-Session1).

All 8 participants underwent the Peabody Picture Vocabulary Test (PPVT) to assess cognitive function (Dunn, and Dunn, 1997). The PPVT is a measure of single-word receptive vocabulary that is often used as a proxy for IQ testing because of its high correlations with standardized tests such as the Wechsler Intelligence Scale for Children (Bee, and Boyd, 2004). It provides standard scores with a mean of 100 and a standard deviation of 15, and the DSM-IV classifies full scale IQ's above 70 as nonretarded (APA, 2000). Participants in this study obtained a standard score of 80 or above on the PPVT measure.

The Social Responsiveness Scale (SRS) is a 65-item, 15-min parent-report questionnaire designed to quantitatively measure the severity of autism-related symptoms. This measure provides an index of ASD-related social competence with questions related to social awareness, social information processing, capacity for reciprocal social communication, social anxiety/avoidance, and autistic preoccupations and traits. The SRS has been shown to correlate on the order of 0.7 with the ADI-R (Constantino et al., 2003). Behaviors and characteristics are rated on a 4-point scale that ranges from "Not True" to "Almost Always True." The SRS generates a total T-score reflecting severity of social deficits in the autism spectrum, as well as five Treatment Subscales: Receptive, Cognitive, Expressive, and Motivational aspects of social behavior, and Autistic Preoccupations. The T-score categorizes measurements in the Normal Range (≤ 59T), Mild to Moderate ASD Range (60T-75T), or Severe Range (≥ 76T)

(Constantino, 2002). Three participants ranked within the Mild to Moderate Range (ASD2, ASD6, and ASD8) with the remaining five falling into the Severe Range (ASD1, ASD3, ASD4, ASD5, and ASD7).

The Social Communication Questionnaire (SCQ) is a brief instrument for the valid screening or verification of ASD symptoms in children that has been developed from the critical items of the Autism Diagnostic Interview (ADI) and compiled into a parent report questionnaire (Rutter et al., 2003a). As in the ADI, these questions tap the three critical autism diagnostic domains of qualitative impairments in reciprocal social interaction, communication, and repetitive and stereotyped patterns of behavior. Among 200 children and adolescents, domain scale scores of the SCQ were significantly correlated with corresponding scores derived from the full ADI ($r = 0.55$ to 0.71, $p < 0.005$) (Berument et al., 1999). Analysis indicated that the SCQ was comparable to the ADI in discriminating ASD from non-ASD, autism vs. mental retardation, and autism vs. other aspects of ASD. A cutoff score of 13 is recommended to maximize valid ascertainment of cases of ASD (specificity) while minimizing errors of omission (sensitivity). The SCQ was designed for use with children over the age of four years with a mental age of at least two years. All participants (except ASD1 and ASD7) met the ASD cutoff for SCQ measure with ASD1 and ASD7 falling off marginally. However, ASD1 and ASD7 met the ASD cutoffs on the ADOS and the ADI-R measures.

The Autism Diagnostic Observation Schedule-Generic (ADOS-G) is a 45-min. semi-structured standardized observational assessment of play, social interaction, and communicative skills that was designed as a diagnostic tool for identifying the presence of autism (Lord et al., 2000). It is organized into four modules, which are distinguished by their appropriateness for use with individuals functioning at different developmental levels, ranging from nonverbal children to highly fluent adults. Each module provides a set of behavioral ratings in five domains: Language and Communication, Reciprocal Social Interaction, Play or Imagination/Creativity, Stereotyped Behaviors and Restricted Interests, and Other Abnormal Behaviors. The scoring algorithm provides cutoffs that can be used to discriminate between a diagnosis of autism, autism spectrum, or non-spectrum. Across all modules, inter-observer agreement for the algorithm score was 0.92, and the test-retest correlation was 0.82. Agreement about diagnostic classification (autism vs. autism spectrum vs. non-spectrum) ranged from 81%-93% (Lord et al., 2000). After coding ratings on the five domains, a total score on the two main components of Communication and Reciprocal Social Interaction equal to or above 7 would indicate autism spectrum, and a score of 10 or more would indicate autistic disorder. All the participants in our study met the cutoff criterion.

The Autism Diagnostic Interview-Revised (ADI-R) is a semi-structured, investigator-based interview for parents/caregivers that was developed for the purpose of diagnostic classification of individuals who may have autism or other pervasive developmental disorders (Rutter et al., 2003b). This interview covers areas of background and history, early development, acquisition and loss of skills, language and communication, social development and play, favorite activities/toys, interests and behaviors, and general behaviors. The ADI-R provides explicit scoring criteria that yield cutoff scores in the domains of social reciprocity, language and communication, and restricted and repetitive activities. The scores from a subset of critical items of the ADI-R are summed to yield scores for each domain; cutoffs are used to determine whether the individual's diagnostic classification is consistent with an autism spectrum disorder. This measure possesses strong psychometric properties in terms of inter-observer agreement, internal consistency, and test-retest reliability. The ADI-R has been found to discriminate autism from non-autism in individuals with mental ages of at least 18 months (Lord et al., 1997). A total score on the four domains: Reciprocal Social Interaction, Communication, Restricted and Repetitive Patterns of Behavior, and Evidence of Abnormal Development before 36 months of age, of the ADI-R equal to or above 22 would indicate autistic disorder (Rutter et al., 2003b). All the participants (except ASD5) in our study met the ADI-R cutoff. For ASD5, the ADI-R score was not available, although ASD5 was above the clinical threshold on the other measures, such as, SRS and SCQ.

A comparative analysis was carried out between the two groups of participants on their age, PPVT scores, and ASD measures such as, SRS, SCQ, ADOS-G, and ADI-R scores. An independent sample t-test between the two groups of participants, as shown in Table VI-1 indicates that no statistically significant group difference exists between the two groups on all the measures. This implies that the two groups are matched on all the above measures.

Procedure

We designed a usability study of the designed system to investigate the implications of the designed VR-based social interactive system with adaptive response technology. The commitment required of interested participants is a total of 2 sessions (lasting for approximately 2.5 hours). The first session runs approximately 1.5 hours, due to two brief adaptation phases (for the participant) with the gathering of the consent/assent. The second session lasts about 1 hour. For each completed session, a participant receives a $15 gift card.

The experiment setup (Figure VI-1) for the usability study, includes a 17" task computer monitor (C1) dedicated to VR-based tasks. For the VR-based tasks, we use Vizard (Worldviz, Santa Barbara, CA), a commercially available Python-based VR design package (discussed in Chapter V). A participant's eye-movement is tracked by using Eye-Tracker goggles (from http://www.arringtonresearch.com/; discussed in Chapter V). Also, a child's physiological data (such, as cardiovascular, electrodermal, electromyographic, and skin temperature) are acquired via wearable biofeedback sensors and Biopac system (MP150 from www.biopac.com; discussed in Chapter V). The data collection system is wearable. The sensors are small, lightweight, non-invasive, and FDA approved. They have been successfully used to collect physiological data of children with ASD in our previous work (Conn, et al., 2008a; Conn, et al., 2008b; Liu, et al., 2008a; Liu, et al., 2008b). All the signal conditioning, and feature extraction routines are written in MATLAB (www.mathworks.com). Computer C1 is connected to the Biopac system via a parallel port to transmit task related event-markers. The physiological signals along with the event markers (e.g., start/end of a social interaction task, performance events) are acquired by the Biopac system and sent over an Ethernet link to the Biopac computer C2 where the physiological signals are stored in a time synchronized manner. Also, eye-data along with task-related event markers and participant's responses while interacting with the VR-based system are logged onto C1. The signal from C1 presenting the VR-based social task are routed to a separate monitor (M1) so that both the participant's parent/caregiver and a clinical observer/therapist can view how the task progresses. Also, both the observers can watch the participant from a video camera view, whose signal is routed to a television, hidden from the participant's view. We video record each session to cross-reference observations made during the experiment.

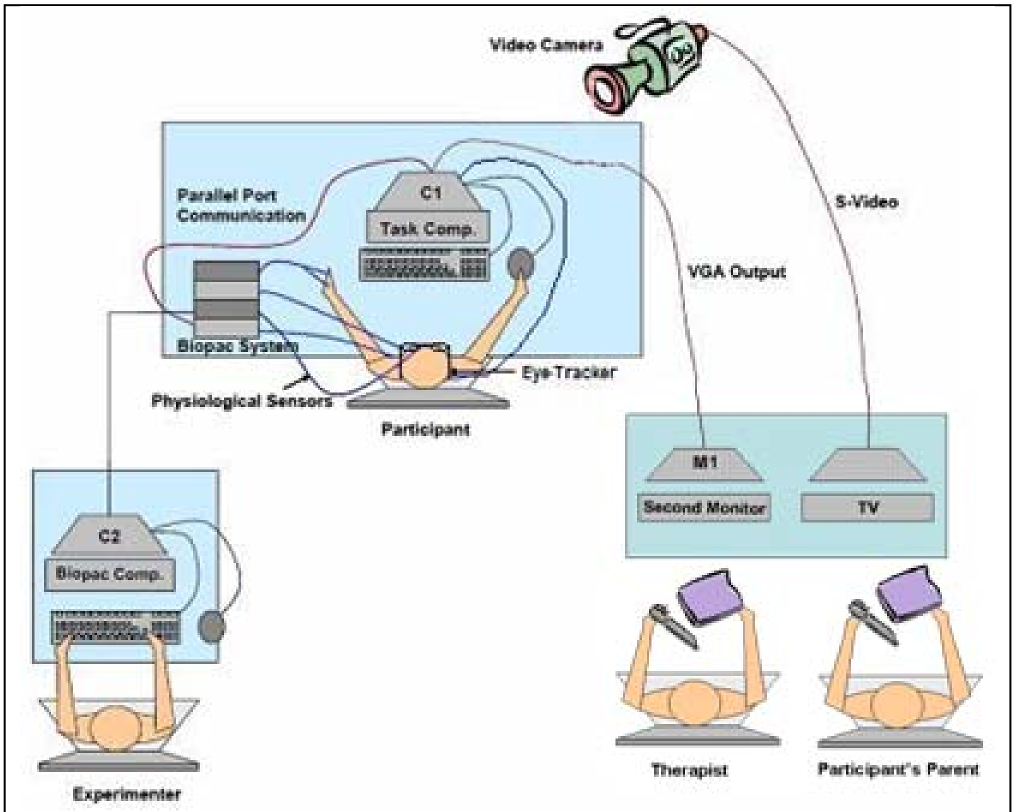


Figure VI-1. Experimental setup.

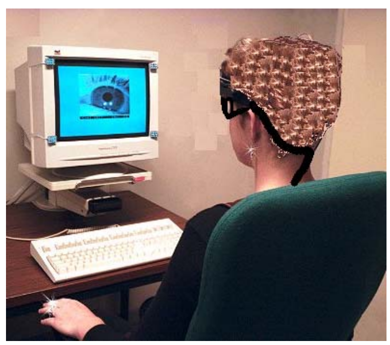
Each participant participated in two VR-based social interaction sessions on two different days. The first session began with the adaptation of the participant. This adaptation stage consisted of two phases. In the first phase, the experimenter briefed the participant about the experiment, the physiological sensors to be used during the experiment, and that they could choose anytime to withdraw from the experiments for any reason, especially if they were not comfortable interacting with the system. This phase ran for approximately 10 min. This was followed by gathering of consent and assent forms for about the next 5 min. Then in the second phase of the adaptation of the participant, the experimenter asked the participant to sit comfortably on a height-

<u>Hello! You're about to interact with your virtual classmates. Here's how things will go:</u>

SETTING UP

1) You'll sit at the computer and we will help you put on some sensors. The sensors are sticky and go on your face, your hands, shoulder, and over your heart. They won't hurt, but they'll feel a little like a band-aid. They tell us information about how your body is responding.

2) You will wear a pair of glasses that have small cameras. These cameras will make a video of your eyes to see where you are looking.



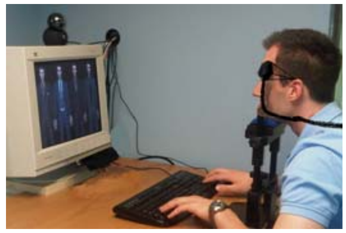
3) To make sure the cameras follow your eyes, we have to "calibrate" your glasses. That means you will look at the computer screen and watch green boxes of collapsing squares. You will keep your head still by using a chin rest. After we calibrate, it is important that you stay still so that the camera gets a good video. If you move, we will need to take a break to recalibrate your glasses.

Figure VI-2. Visual Schedule (part (a))

adjustable chair. The chair was adjusted so that his/her eyes were collinear with the center of the task computer, C1 (Figure VI-1). Then the experimenter walked the

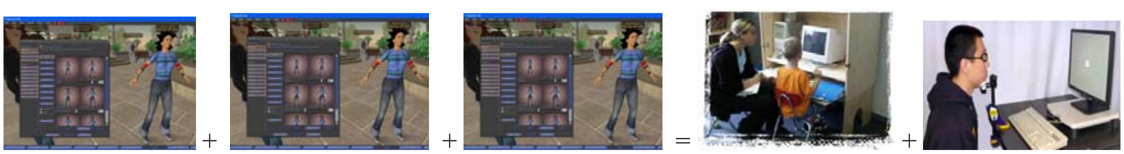
PLAYING THE SESSION

1) First, you will watch your classmates give presentations. Some will be interesting, and some might be boring. Remember to keep your chin on the chin rest.



2) After your classmates give their presentations, you will start a conversation with them by choosing questions in a particular order to ask. Your classmates can tell if you are looking at them or not. Remember to pay attention to them.

3) You will first complete 3 presentations and 3 conversations. Then, you will take a break and we will talk to you about them. When we take a break, you can move around. If you have any questions or comments, this is when you will tell us. After we finish talking, we will recalibrate your glasses and then continue with the session again.



1 + 2 + 3 = Break to Talk + Recalibrate

4) You will then continue the session. After every two presentations and conversations you complete, you will take a break we will talk with you about them. After we finish talking, we will recalibrate your glasses and then continue with the session again.

1 + 2 = Break to Talk + Recalibrate

Figure VI-2. Visual Schedule (part (b))

participant through a visual schedule (Figure VI-2, part (a) and part (b)). This served to contribute to the second phase of adaptation on the one hand along with ensuring consistency in this introductory presentation of the experimental session to the participants on the other hand. This phase ran for approximately 15 min.

After getting the verbal confirmation from the participant that he/she was ready to start the experiment, the experimenter placed the peripheral physiological sensors on the participant's body. Then the participant was asked to wear the eye-tracker goggles. 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 (C1). This was followed by the VR-based social communication task. In Session1, the participant viewed an initial instruction screen followed by an avatar giving presentation by narrating a personal story while moving dynamically in the VR world displaying context-relevant social situations (discussed in Chapter V) to the participant. At the end of the VR-based presentation by the avatar, the participant was asked to find out a piece of information from the avatar. The participant then interacted with the avatar socially by using the bidirectional social conversation module (discussed in Chapter V) by selecting one choice at a time from the menu, using a mouse. The avatar responded to the question/statement selected by the participant and after a few back-and-forth turns, the avatar ended the conversation. Then the system moved to the next VR-based social communication task. The first three VR-based social communication tasks helped in selecting the baseline for each participant. Specifically, these three social communication tasks consisted of social tasks of the three interaction difficulty levels, with one in each level. Out of these three social tasks, only one was selected as the baseline depending on the highest performance score achieved by the participant while interacting with the avatar using the bidirectional conversation module. In addition, in order to identify the baseline level of the participant while interacting with the avatar using the bidirectional conversation module (discussed in Chapter V), during these first three social tasks, the avatar did not give any feedback (discussed in Section '*Design Specifications of the Feedback Given by the Avatars to Facilitate Participants to Continue Bidirectional Conversation'* in Chapter V*)* to the participant when the participant made an irrelevant choice. However, the bidirectional conversation modules of the social communication tasks following the baseline were accompanied with appropriate feedback provided by the avatar to the participant so as to facilitate the participant to walk though the conversation process when the avatar felt necessary. At the end of each social communication task, both the clinical observer/therapist and the participant's parent/caregiver rated the participant as to what they thought the level (using a 1-9 scale, with 1-not at all, and 9-very much) of the target affective states of engagement, enjoyment, and anxiety was for the participant during the finished social communication task.

Session2 was similar to Session1, except that in Session2, the system delivered an audio-visual feedback to the participants based on their viewing patterns (discussed in Table V-1). In addition, in Session1, the VR-based social task modification strategy was based only on one's task performance metric while participating in the social communication task (Table V-6). However, in Session2, the VR-based social task modification strategy was based on the composite effect of one's behavioral viewing, eye physiological indices and the task performance metric (Table V-5).

Among the 8 adolescents who participated in the study, 4 participants (ASD1-ASD4) participated first in Session1-followed by-Session2 (henceforth referred to as Group1). In Session1, the VR-based gaze-sensitive social communication system adapted the social tasks presented to the participants based only on the performance metric. In Session2, the VR-based social system intelligently adapted itself based on the engagement level of a participant predicted from the composite effect of his/her behavioral viewing, eye-physiological indices, and the performance metric. The other group (henceforth referred to as Group2) of 4 participants (ASD5-ASD8) was exposed first to the Session2-followed by-Session1. This was carried out to determine whether there existed any ordering effect (Heiman, 2002) of presentation of Session1 and Session2. Also a washout period of approximately 2-4 weeks was maintained between each participant's participation in Session1 and Session2. This washout period was used after a literature review where studies used washout period of 2 weeks (Bolman, and Richmond, 1999; Castner, Williams, and Goldman_Rakic, 2000) and 4 weeks (Zhang et al., 2004; Brownell, 2002).

Results

The objective of this section is to examine the acceptability of the system by the target population and to present the results of an investigation to study the effects of interacting with a VR based gaze-sensitive social communication system equipped with adaptive response technology. We discuss the effects of interaction with such a system so far as one's affective states (e.g., engagement, enjoyment, and anxiety), performance and behavioral viewing during the social communication task are concerned.

System Acceptability

In the current study, we wanted to investigate whether our VR-based gaze-sensitive system with adaptive response technology was acceptable to our participants with ASD. In spite of being given the option of withdrawing from the experiment at any time during their interaction with the system, all the participants completed the sessions. An exit survey carried out at the end of the experiment revealed that all the participants liked interacting with the system particularly while using the bidirectional conversation module, had no problems in wearing the eye-tracker goggles and accepting the peripheral physiological sensors, and understanding the stories narrated by their virtual classmates. When asked about any take-home lesson that they had from the conversation between them and their virtual classmates, most of them (6 out of 8) said that they learned that they should introduce themselves first while speaking to a new friend for the first time and that they should look towards the faces of their friends during conversation. Thus, it is reasonable to infer from this study that our system has a potential to be accepted by the target population.

Feasibility of the System to Create Varying Levels of Engagement, Enjoyment, and Anxiety corresponding to the different Difficulty Levels of VR-based Social Interaction

While the participants participated in the VR-based social communication task, both the clinical observer/therapist and the participant's parent/caregiver rated as to what they thought the level of engagement of the participant was during the finished trial (using a 1-9 scale, with 1:not engaged and 9: very engaged). The same therapist was involved in all of the experiment sessions, which aided in establishing a consistent reporter. As literature review indicates that a clinical observer / therapist's report on the affective states of participants is a reliable measure (Eisenberg et al., 1995) as an experiment design methodology, reports from the therapist are used whenever referring to the participant's affective states. Thus, we investigated the variation in the engagement level of the participants as reported by the therapist corresponding to the three difficulty levels (easy, medium, and high) of social interaction. The Fig. VI-3 indicates that the varying difficulty levels of VR-based social interaction were capable of generating varying levels of participants' engagement, as reported by the therapist. A dependent sample T-test on the participants' engagement level as rated by the therapist corresponding to the 'Easy' and 'High' level of interaction difficulty reveals the engagement levels to be marginally statistically different ($p = 0.0719$) for Session 1 and not significantly different for Session 2.

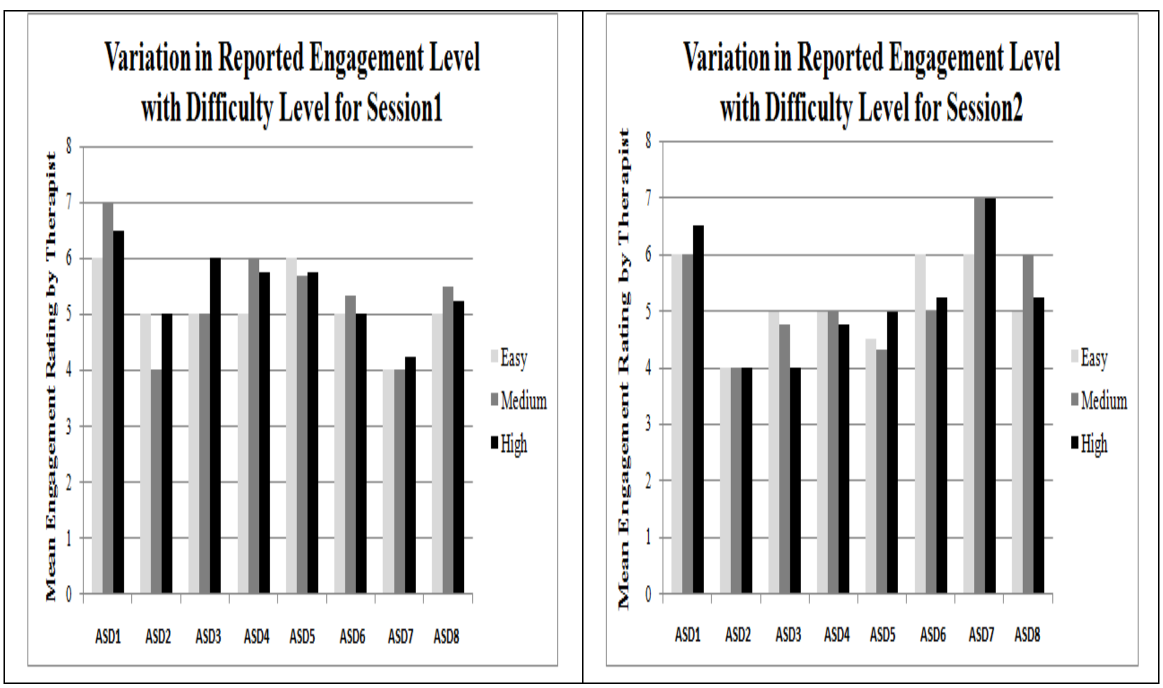


Figure VI-3. Variation in the Reported Engagement level of participants with different Difficulty Levels of social interaction.

We investigated the variation in the participants' level of enjoyment as rated by the therapist (using a 1-9 scale with 1: not enjoyed and 9: very enjoyed). The Fig. VI-4 indicates that variation in the difficulty levels of VR-based social interaction was capable of generating varying levels of participants' enjoyment, as reported by the therapist. A dependent sample T-test on the participants' enjoyment level as rated by the therapist corresponding to the 'Easy' and 'High' level of interaction difficulty reveals the enjoyment levels to be marginally statistically different (p = 0.0543) for Session 1 and not significantly different for Session 2.

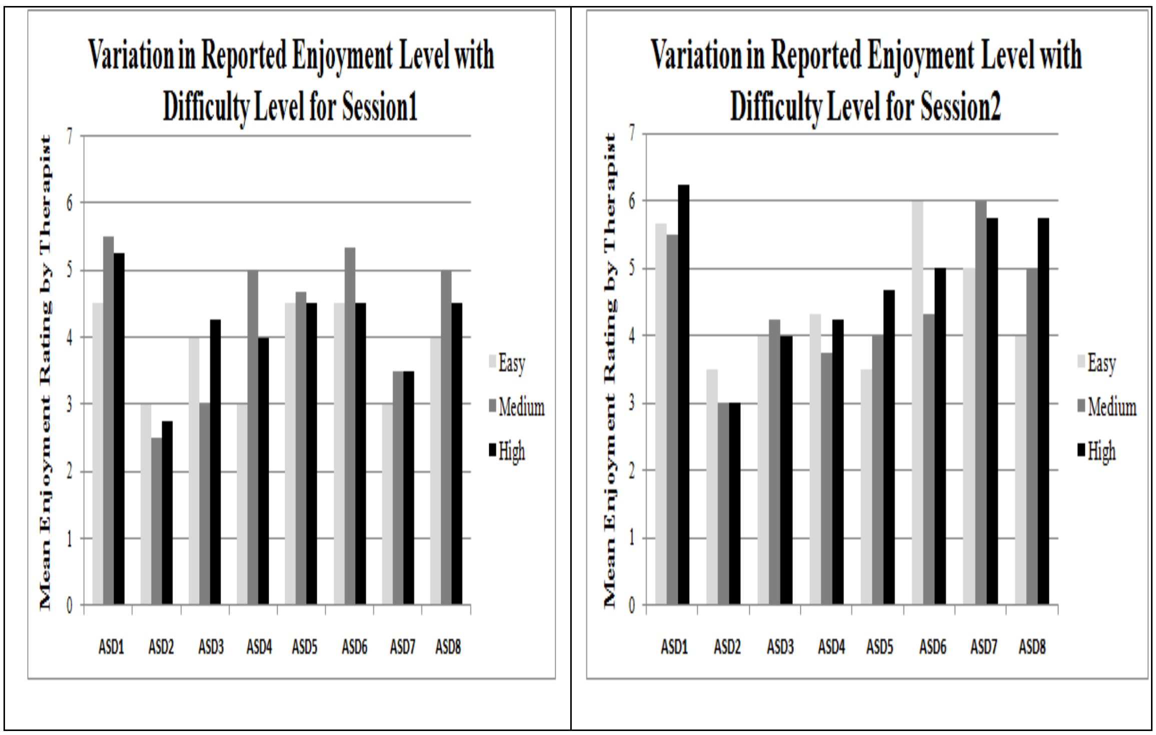


Figure VI-4. Variation in the Reported Enjoyment level of participants with different Difficulty Levels of social interaction.

Similarly, we investigated the variation in the participants' level of anxiety as rated by the therapist (using a 1-9 scale with 1: not anxious and 9: very anxious). Fig. VI-5 shows that a variation in the difficulty levels of VR-based social interaction was capable of generating varying levels of participants' anxiety, as reported by the therapist. A dependent sample T-test on the participants' anxiety level as rated by the therapist corresponding to the 'Easy' and 'High' level of interaction difficulty reveals the anxiety levels to be statistically different (p = 0.0387) for Session 1 and not significantly different for Session 2.

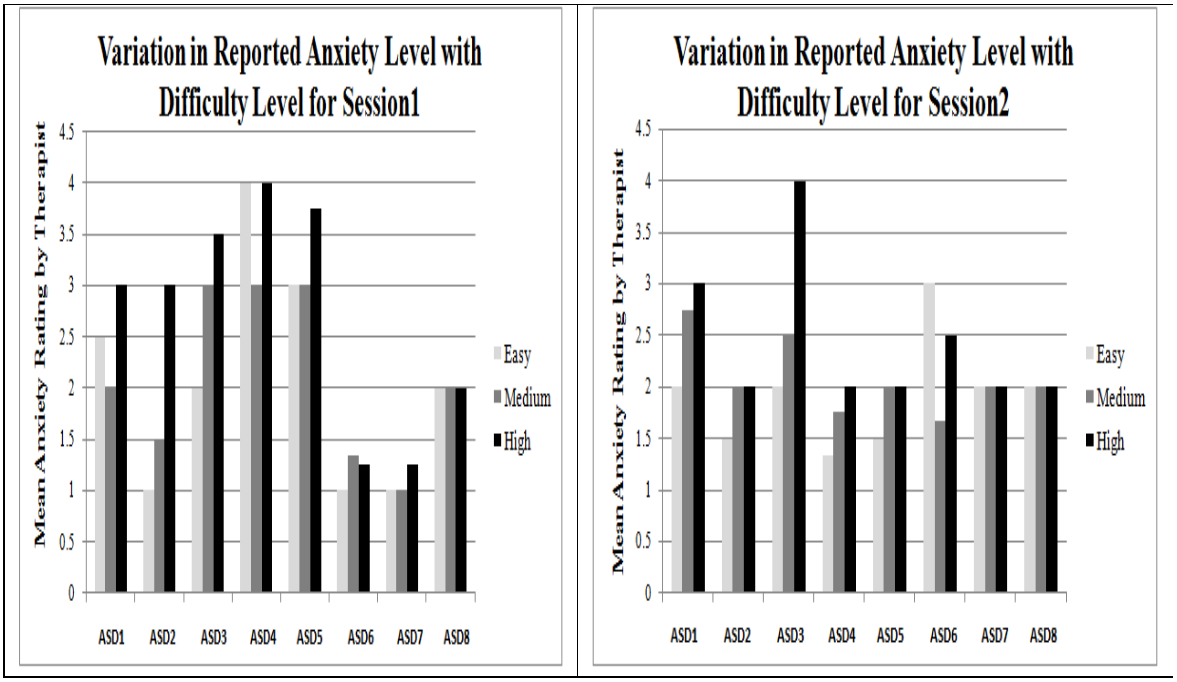


Figure VI-5. Variation in the Reported Anxiety level of participants with different Difficulty Levels of social interaction.

To summarize, we can say that our VR-based gaze-sensitive social interactive system was capable of eliciting varying levels of affective states of engagement, enjoyment, and anxiety among the participants, as is evident from the therapist's ratings. In addition, the dependent sample statistical T-test between the reported measures on the level of the affective states during the trials corresponding to the lowest difficulty level (i.e., 'Easy') and the highest difficulty level (i.e., 'High') indicates that they are statistically different (marginally statistically different for engagement and enjoyment and statistically significantly different for anxiety) for Session1. However, during Session2, while our system adaptively responded based on the participant's engagement level predicted from the composite effect of behavioral viewing, eye physiology, and performance metric, thereby allowing the participant to progress through the social tasks while encouraging socially-appropriate interaction and maintaining the basic levels of comfort, these affective states are found to be non-statistically significant while compared across the 'Easy' and the 'High' difficulty levels. This may imply that our system which used adaptive response technology during Session2 was capable of adequately adapting itself to the participants' predicted engagement level which resulted in bridging the gap in the affective states corresponding to the different difficulty levels.

A Brief Description of Participant Interaction during Session1 and Session2

Here we present a brief description of the VR-based social interaction for each participant during Session1 and Session2.

ASD1:

Session1: This participant interacted in six VR-based social task trials of which two were in Easy level of difficulty (average score 28 out of 30), one in Medium level of difficulty (average score 46 out of 50), and three in High level of difficulty (average score 68.67 out of 70). ASD1 started with Easy level of difficulty as the baseline, followed by one trial of Easy level, then switched to one trial of Medium difficulty level, and then to three trials of High difficulty level. Also ASD1 fixated on the face of the avatar for an average of approximately 49% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in nine VR-based social task trials of which two were in Easy level of difficulty (average score 30 out of 30), four in Medium level of difficulty (average score 48 out of 50), and three in High level of difficulty (average score 68.67 out of 70). ASD1 started with Medium level of difficulty as the baseline, followed by one trial of Medium level, then switched to one trial of Low difficulty level, then to two trials of Medium difficulty level. With decrease in predicted engagement level, the system then offered one trial of Low difficulty level to him. Subsequently, when the predicted engagement level of ASD1 was high, the system having no more games of Medium difficulty level, offered the participant with a trial of High difficulty level. This was followed by two more trials of High difficulty level. Also ASD1 fixated on the face of the avatar for an average of approximately 56% of the time during the VR-based social conversation tasks.

Inference: Thus we find that ASD1 interacted in more VR-based social task trials during Session2 than that during Session1. Also, ASD1 achieved greater performance scores during the trials of Easy level of difficulty and the Medium level of difficulty during Session2 than those during Session1. In addition, ASD1 fixated on the face of the communicator (i.e., the avatar) for greater percentage of the time during the VR-based social communication task during Session2 than that during Session1.

ASD2:

Session1: This participant interacted in five VR-based social task trials with two in Medium level of difficulty (average score 50 out of 50), and three in High level of difficulty (average score 63.33 out of 70). ASD2 started with Medium level of difficulty as the baseline, followed by one trial of Medium level, then switched to three trials of High difficulty level. Also ASD2 fixated on the face of the avatar for an average of approximately 10% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in four VR-based social task trials with all four in Easy level of difficulty (average score 30 out of 30). ASD2 started with Easy level of difficulty as the baseline, followed by one trial of Easy level. Then, on predicting low engagement level of ASD2 and the Easy level of difficulty being the lowest of the three difficulty levels, the system continued to offer ASD2 with trials of Easy difficulty level. ASD2 fixated on the face of the avatar for an average of approximately 30% of the time during the VR-based social conversation tasks.

Inference: We find that ASD2 interacted in less VR-based social task trials during Session2 than that during Session1. However, ASD2 fixated on the face of the communicator (i.e., the avatar) for greater percentage of the time during the VR-based social communication task during Session2 than that during Session1. Thus, though ASD2 could go to High level of difficulty during Session1, this was achieved in socially inappropriate way as is evident from the very less percent of fixation on the face of the avatar during this Session1. In fact, ASD2 was one of the two participants who declined to comment anything on the take-home lessons from the sessions during the exit survey (as discussed in '*System Acceptability*').

ASD3:

Session1: This participant interacted in four VR-based social task trials with all four in High level of difficulty (average score 64 out of 70). ASD3 started with High level of difficulty as the baseline, followed by three trials of High level. Also ASD3 fixated on the face of the avatar for an average of approximately 7% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in seven VR-based social task trials with three in

Easy level of difficulty (average score 30 out of 30), and four in Medium level of difficulty (average score 47 out of 50). ASD3 started with Medium level of difficulty as the baseline, followed by one trial of Medium level, then switched to two trials of Easy difficulty level as the system predicted a low engagement level. However, on getting an improved predicted engagement level, the system offered ASD3 with two numbers of trials of Medium level. Subsequently, with decreased engagement level, the system switched to the Easy difficulty level. ASD3 fixated on the face of the avatar for an average of approximately 29% of the time during the VR-based social conversation tasks.

Inference: ASD3 interacted in more VR-based social task trials during Session2 than that during Session1. In addition, ASD3 fixated on the face of the communicator (i.e., the avatar) for greater percentage of the time during the VR-based social conversation task during Session2 than that during Session1. Thus, though ASD3 could go to High level of difficulty during Session1, this was achieved in socially inappropriate way as is evident from the very less percent of fixation on the face of the avatar during this Session1.

ASD4:

Session1: This participant interacted in four VR-based social task trials with all four in High level of difficulty (average score 61 out of 70). ASD4 started with High level of difficulty as the baseline, followed by three trials of High level. Also ASD4 fixated on the face of the avatar for an average of only approximately 1% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in nine VR-based social task trials with two in Easy level of difficulty (average score 28 out of 30), four in Medium level of difficulty (average score 47 out of 50), and three in High level of difficulty (average score 64.67

out of 70). ASD4 started with Medium level of difficulty as the baseline, followed by one trial of Medium level, then switched to two trials of Easy difficulty level. Then the system predicted improved engagement level of ASD4, thereby offering him with two numbers of trials of Medium level of difficulty. With continued improved predicted engagement level, the system offered three trials of High difficulty level. Also ASD4 fixated on the face of the avatar for an average of approximately 37% of the time during the VR-based social conversation tasks.

Inference: Thus we find that ASD4 interacted in more VR-based social task trials during Session2 than that during Session1. Also, ASD4 achieved greater performance scores during the trials of High level of difficulty during Session2 than that during Session1. In addition, ASD4 fixated on the face of the communicator (i.e., the avatar) for greater percentage of the time during the VR-based social conversation task during Session2 than that during Session1.

ASD5:

Session1: This participant interacted in seven VR-based social task trials with two in Easy level of difficulty (average score 28 out of 30), two in Medium level of difficulty (average score 48 out of 50), and three in High level of difficulty (average score 56 out of 70). ASD5 started with Easy level of difficulty as the baseline, followed by one trial of Easy level, then switched to one trial of Medium difficulty level, then to two trials of High difficulty level, then to one trial of Easy difficulty level, and ultimately with one trial of High difficulty level. Also ASD5 fixated on the face of the avatar for an average of approximately 3% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in eight VR-based social task trials with four in Easy level of difficulty (average score 29 out of 30), two in Medium level of difficulty (average score 48 out of 50), and two in High level of difficulty (average score 64 out of 70). ASD5 started with Easy level of difficulty as the baseline, followed by one trial of Easy level, and then with prediction of reduced engagement level of ASD5, the system continued at the Easy level of difficulty (with the Easy level being the lowest of the three levels of difficulty). On predicting an improved engagement level of ASD5, the system offered him with trial of Medium level of difficulty. With predicted engagement level being high, the system offered him with two trials of High level of difficulty. Again with fall in predicted engagement level of ASD5, the system reduced the difficulty level to Medium followed by a trial of Easy level of difficulty. Also ASD5 fixated on the face of the avatar for an average of approximately 18% of the time during the VR-based social conversation tasks.

Inference: Thus we find that ASD5 interacted in more VR-based social task trials during Session2 than that during Session1. Also, ASD5 achieved greater performance scores during the trials of Easy level of difficulty and the High level of difficulty during Session2 than those during Session1. In addition, ASD5 fixated on the face of the communicator (i.e., the avatar) for greater percentage of the time during the VR-based social conversation task during Session2 than that during Session1.

ASD6:

Session1: This participant interacted in seven VR-based social task trials with two in Easy level of difficulty (average score 30 out of 30), two in Medium level of difficulty (average score 50 out of 50), and three in High level of difficulty (average score 58.67 out of 70). ASD6 started with Easy level of difficulty as the baseline, followed by one trial of Easy level, then switched to one trial of Medium difficulty level, then to three trials of High difficulty level, and finally to one trial of Medium difficulty level. Also ASD6 fixated on the face of the avatar for an average of approximately 28% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in seven VR-based social task trials with two in Easy level of difficulty (average score 30 out of 30), two in Medium level of difficulty (average score 48 out of 50), and three in High level of difficulty (average score 66 out of 70). ASD6 started with Easy level of difficulty as the baseline, followed by one trial of Easy level, and then with prediction of improved engagement level of ASD6, the system offered him with a trial of Medium level of difficulty. With decrease in predicted engagement level of ASD6, the system maintained at the Medium level of difficulty. With predicted engagement level going high, the system offered him with three trials of High level of difficulty. Also ASD6 fixated on the face of the avatar for an average of approximately 29% of the time during the VR-based social conversation tasks.

Inference: Though ASD6 interacted in same VR-based social task trials during Session2 as that during Session1, he achieved greater performance score during the trials of High level of difficulty during Session2 than those during Session1. However, ASD6 showed a very less improvement in the percent of time spent by him in fixating on the face of the communicator (i.e., the avatar) during the VR-based social conversation task during Session2 than that during Session1.

ASD7:

Session1: This participant interacted in five VR-based social task trials with two in Medium level of difficulty (average score 50 out of 50), and three in High level of difficulty (average score 67.33 out of 70). ASD7 started with Medium level of difficulty as the baseline, followed by one trial of Medium level, and then switched to three trials of High difficulty level. Also ASD7 fixated on the face of the avatar for an average of approximately 42% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in four VR-based social task trials with all being of the High difficulty level (average score 69 out of 70). ASD7 started with High level of difficulty as the baseline, followed by three trials of High level. Also ASD7 fixated on the face of the avatar for an average of approximately 69% of the time during the VR-based social conversation tasks.

Inference: Though ASD7 interacted in less VR-based social task trials during Session2 than that during Session1, she achieved greater performance score during the trials of High level of difficulty during Session2 than those during Session1. In addition, ASD7 showed improvement in the percent of time spent by her in fixating on the face of the communicator (i.e., the avatar) during the VR-based social communication task during Session2 than that during Session1.

ASD8:

Session1: This participant interacted in five VR-based social task trials with two in Medium level of difficulty (average score 50 out of 50), and three in High level of difficulty (average score 68.67 out of 70). ASD8 started with Medium level of difficulty as the baseline, followed by one trial of Medium level, and then switched to three trials of High difficulty level. Also ASD8 fixated on the face of the avatar for an average of approximately 7% of the time during the VR-based social conversation tasks.

Session2: This participant interacted in four VR-based social task trials with all being of the High difficulty level (average score 68 out of 70). ASD8 started with High level of difficulty as the baseline, followed by three trials of High level. Also ASD8 fixated on the face of the avatar for an average of approximately 36% of the time during the VR-based social conversation tasks.

Inference: ASD8 interacted in less VR-based social task trials during Session2 than that during Session1, and also he was the only participant to achieve a slightly lower performance score during the trials of High level of difficulty during Session2 than those during Session1. However, ASD8 showed improvement in the percent of time spent by him in fixating on the face of the communicator (i.e., the avatar) during the VR-based social communication task during Session2 than that during Session1. In fact, ASD8 was one of the two participants who declined to comment anything on the take-home lessons from the sessions during the exit survey (as discussed in '*System Acceptability*').

Quantitative Analysis of Performance of Participants during Trials (VR-based social communication tasks) for the Session1 and Session2

The engagement of children with ASD is the ground basis for the 'floor-time-therapy' to help them develop relationships and improve their social skills (Wieder, and Greenspan, 2005). Clinicians who work with children in autism intervention intensely monitor affective cues, e.g., engagement in order to make appropriate decisions about adaptations to their intervention and reinforcement strategies. Thus our hypothesis was that if we can allow a computer to recognize the engagement level of a child in terms of his/her behavioral viewing pattern, eye physiological indices, and performance during VR-based social communication tasks and apply this information as a means of taking appropriate decisions about the adaptation of the child to the social task, then it may contribute to improved social task performance. Our usability study comprised of two sessions, namely Session1 and Session2. In Session1, the task switching was based only on one's performance metric alone and in Session2, the task switching was based on the composite effect of one's behavioral viewing, eye physiology, and performance metric during the VR-based social task.

In order to carry out a quantitative analysis of the performance of the participants while they interacted with VR-based social communication system during Session1 and Session2, we computed the weighted performance score similar to other studies (Javier, 2007; Hirsch et al., 2004). For this we first computed the normalized weighted performance score (Table VI-2). Specifically, the weight of the social communication task is considered as '1' for the 'Easy' difficulty level, '2' for the 'Medium' difficulty level, and '3' for the 'High' difficulty level of the VR-based bidirectional social communication module. In order to carry out a comparative analysis among the performance of the participants, each of whom participated in different VR-based social communication tasks (of varying numbers of trials and of difficulty levels), we need to compute normalized values of the performance scores achieved by the participants during the Session1 and Session2. The formulae that we have used to compute the normalized scores are as follows:

Let us consider that the VR-based social task trials of 'Easy', 'Medium', and 'High' difficulty levels have weights designated by 'x', 'y', and 'z' respectively. Also, let a participant acquires an average performance score of '$XAvg$' (out of maximum possible score of '$XMax$' (i.e., 30) for trials of 'Easy' difficulty level), '$YAvg$' (out of maximum possible score of '*YMax*' (i.e., 50) for trials of 'Medium' difficulty level), and '*ZAvg*' (out of maximum possible score of '*ZMax*' (i.e., 70) for trials of 'High' difficulty level).

Case1- A participant interacted with VR-based social task trials of 'Easy', 'Medium', and 'High' difficulty levels. The Weighted Performance Score Achieved (Normalized) is:

$$\mathbf{PERF.SCORE(Normalized)} = \frac{\left(\frac{X}{X+Y+Z}XAvg\right)+\left(\frac{Y}{X+Y+Z}YAvg\right)+\left(\frac{Z}{X+Y+Z}ZAvg\right)}{\left(\frac{X}{X+Y+Z}XMax\right)+\left(\frac{Y}{X+Y+Z}YMax\right)+\left(\frac{Z}{X+Y+Z}ZMax\right)} \text{……………(VI.1)}$$

Case2- A participant interacted with VR-based social task trials of 'Easy', and 'Medium' difficulty levels. The Weighted Performance Score Achieved (Normalized) is:

$$\mathbf{PERF.SCORE(Normalized)} = \frac{\left(\frac{X}{X+Y}XAvg\right)+\left(\frac{Y}{X+Y}YAvg\right)}{\left(\frac{X}{X+Y}XMax\right)+\left(\frac{Y}{X+Y}YMax\right)} \text{……………………………..(VI.2)}$$

Case3- A participant interacted with VR-based social task trials of 'Medium', and 'High' difficulty levels. The Weighted Performance Score Achieved (Normalized) is:

$$\mathbf{PERF.SCORE(Normalized)} = \frac{\left(\frac{Y}{Y+Z}YAvg\right)+\left(\frac{Z}{Y+Z}ZAvg\right)}{\left(\frac{Y}{Y+Z}YMax\right)+\left(\frac{Z}{Y+Z}ZMax\right)} \text{……………………………..(VI.3)}$$

Case4- A participant interacted with VR-based social task trials of 'Easy' difficulty level only. The Weighted Performance Score Achieved (Normalized) is:

$$\mathbf{PERF.SCORE(Normalized)} = \frac{XAvg}{XMax} \text{……………………………………………(VI.4)}$$

Case5- A participant interacted with VR-based social task trials of 'Medium' difficulty level only. The Weighted Performance Score Achieved (Normalized) is:

$$\mathbf{PERF.SCORE(Normalized)} = \frac{YAvg}{YMax} \text{...} \text{(VI.5)}$$

Case6- A participant interacted with VR-based social task trials of ‘High’ difficulty level only. The Weighted Performance Score Achieved (Normalized) is:

$$\mathbf{PERF.SCORE(Normalized)} = \frac{ZAvg}{ZMax} \text{...} \text{(VI.6)}$$

Table VI-2. Summary of Performance Progression for participants of Group1 during Session1 and Session2.

	Session	Performance Score Achieved (a)	Difficulty Level Wt. (b)	Max.Possible Score (c)	Weighted Perf. Score (Norm) (d)
ASD1	1	26	1	30	0.96
		30	1	30	
		46	2	50	
		66	3	70	
		70	3	70	
		70	3	70	
	2	50	2	50	0.98
		50	2	50	
		30	1	30	
		46	2	50	
		46	2	50	
		30	1	30	
		66	3	70	
		70	3	70	
		70	3	70	
ASD2	1	50	2	50	0.94
		50	2	50	
		58	3	70	
		66	3	70	
		66	3	70	
	2	30	1	30	1.00
		30	1	30	
		30	1	30	
		30	1	30	
ASD3	1	70	3	70	0.91
		70	3	70	
		62	3	70	
		54	3	70	
	2	50	2	50	0.95
		50	2	50	
		30	1	30	
		30	1	30	
		46	2	50	
		42	2	50	
		30	1	30	
ASD4	1	58	3	70	0.87
		62	3	70	
		58	3	70	
		66	3	70	
	2	50	2	50	0.93
		42	2	50	
		30	1	30	
		26	1	30	
		46	2	50	
		50	2	50	
		66	3	70	
		62	3	70	
		66	3	70	

Thus, we find from Table VI-2, that all participants in Group1, i.e., ASD1-ASD4 showed an improvement in the normalized weighted performance score that they achieved during Session2 than that during Session1.

Similarly, we investigated the performance progression for participants of Group2 for Session1 and Session2. From Table VI-3, we find that all the participants of Group2 (except ASD8) showed an improvement in the normalized weighted performance score that they achieved during Session2 than that during Session1. ASD8 showed a small decrement (1.59% from its normalized score during Session1) in the performance score. ASD8 was one of the participants who declines to comment on any take-home lesson from the sessions during exit survey (mentioned in Section '*System Acceptability*').

Table VI-3. Summary of Performance Progression for participants of Group2 during Session1 and Session2.

	Session	Performance Score Achieved (a)	Difficulty Level Wt. (b)	Max.Possible Score (c)	Weighted Perf. Score (Norm) (d)
ASD5	1	30	1	30	0.86
		26	1	30	
		50	2	50	
		58	3	70	
		44	3	70	
		46	2	50	
		66	3	70	
	2	30	1	30	0.93
		30	1	30	
		30	1	30	
		46	2	50	
		66	3	70	
		62	3	70	
		50	2	50	
		26	1	30	
ASD6	1	30	1	30	0.90
		30	1	30	
		50	2	50	
		58	3	70	
		66	3	70	
		52	3	70	
		50	2	50	
	2	30	1	30	0.95
		30	1	30	
		50	2	50	
		46	2	50	
		62	3	70	
		70	3	70	
		66	3	70	
ASD7	1	50	2	50	0.97
		50	2	50	
		62	3	70	
		70	3	70	
		70	3	70	
	2	70	3	70	0.99
		70	3	70	
		66	3	70	
		70	3	70	
ASD8	1	50	2	50	0.99
		50	2	50	
		66	3	70	
		70	3	70	
		70	3	70	
	2	66	3	70	0.97
		70	3	70	
		70	3	70	
		66	3	70	

Then we carried out statistical analysis to determine whether the improvement in the normalized performance score of the participants while interacting with the VR-based social situations during Session1 and Session2 was statistically significant. For this, we (a) first computed a dependent sample T-test for all the participants (ASD1-ASD8) on the normalized performance scores during Session1 and Session2, (b) carried out a dependent sample T-test for each group of participants separately, i.e., for Group1 (ASD1-ASD4) and for Group2 (ASD5-ASD8), and finally (c) performed an independent sample T-test between the normalized performance scores achieved by Group1 and Group2 during Session1 and Session2 to determine whether the presentation of Session1 and Session2 had any statistically significant ordering effect.

A dependent sample T-test for the participants' normalized performance score between Session1 and Session 2 (as mentioned in point (a) above) indicates that they are statistically significantly ($p = 0.0102$) different (Table VI-4). In addition, a dependent sample T-test carried out on the normalized weighted performance score separately for the Group1 and Group2 between Session1 and Session 2 (as mentioned in point (b) above) indicates that they are statistically significantly ($p = 0.0235$) different for Group1 (Table VI-4), but not statistically significant for Group2 (Table VI-4).

Table VI-4. Comparative Analysis of Performance Progression for Group1 and Group2 between Session1 and Session2.

		Normalized Weighted Performance Score				
		Session1	**Session2**	**Session1**	**Session2**	
Group1	ASD1	0.96	0.98	0.92	0.96	Mean
	ASD2	0.94	1.00	0.04	0.03	SD
	ASD3	0.91	0.95	0.0235		p-value
	ASD4	0.87	0.93	4.2720		t-value
Group2						
	ASD5	0.86	0.93	0.93	0.96	Mean
	ASD6	0.90	0.95	0.06	0.02	SD
	ASD7	0.97	0.99	0.2255		p-value
	ASD8	0.99	0.97	1.5216		t-value
Mean		0.93	0.96			
SD		0.05	0.03			
p-value		0.0102				
t-value		3.4814				

As mentioned in the Section '*Procedures*', participants in Group1 (ASD1-ASD4) participated in the VR-based social communication tasks first in Session1-followed by-Session2. However, the participants in Group2 (ASD5-ASD8) participated first in

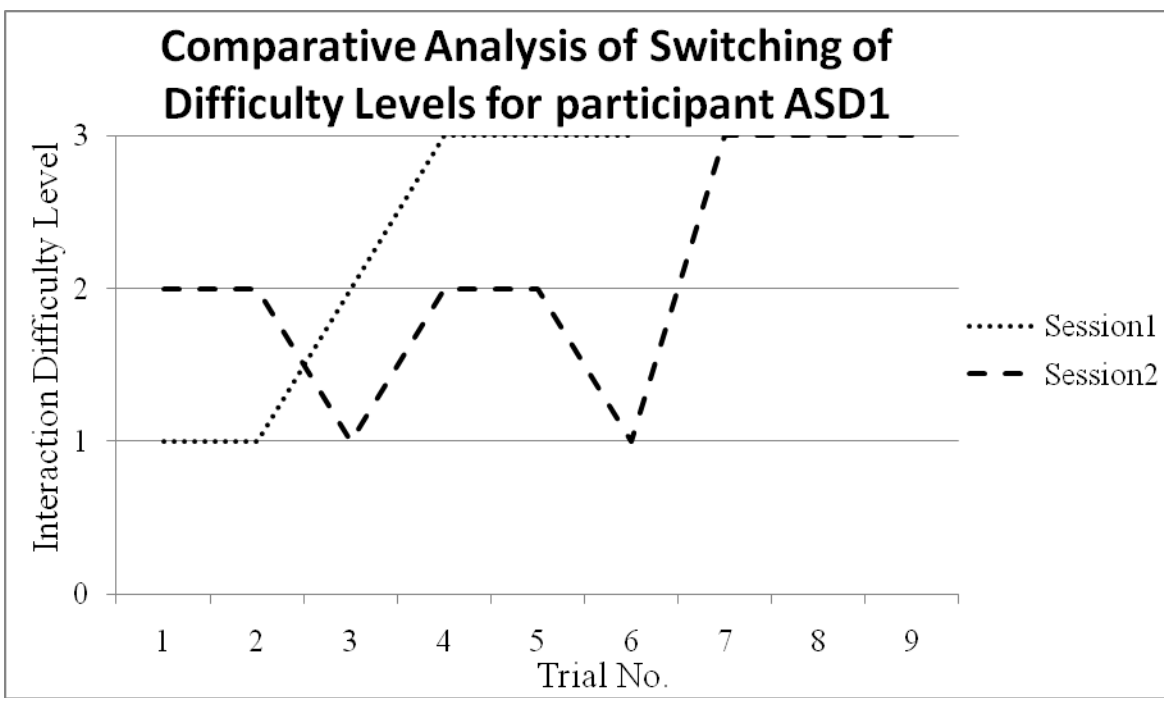
Table VI-5. Comparative Analysis of Performance Progression across Group1 and Group2 for Session1 and for Session2.

		Normalized Weighted Performance Score			
		Session1	**Session1**		
Group1	ASD1	0.96	0.86	ASD5	***Group2***
	ASD2	0.94	0.90	ASD6	
	ASD3	0.91	0.97	ASD7	
	ASD4	0.87	0.99	ASD8	
Mean		0.92	0.93		
SD		0.04	0.06		
p-value		0.7885			
t-value		0.2805			
		Session2	**Session2**		
Group1	ASD1	0.98	0.93	ASD5	***Group2***
	ASD2	1.00	0.95	ASD6	
	ASD3	0.95	0.99	ASD7	
	ASD4	0.93	0.97	ASD8	
Mean		0.96	0.96		
SD		0.03	0.02		
p-value		0.8280			
t-value		0.2270			

Session2-followed by-Session1. We carried out a statistical analysis (as mentioned in point (c) above) to determine whether there was any ordering effects due to the order of presentation of Session1 and Session2 VR-based social tasks. Thus an independent sample T-test carried out on the normalized performance scores achieved across Group1 and Group2 for each of Session1 and Session2 indicates that they are not statistically significantly ($p = 0.7885$ for Session1 and $p = 0.8280$ for Session2) different, as can be seen from Table VI-5. Thus, we can say that there were no significant ordering effects due to the order of presentation of VR-based social tasks of Session1 and Session2.

Progression of VR-based Social Communication Tasks for Session1 and Session2

Here we discuss in details the patterns of task progression for two participants in each of the two groups while they participate in the VR-based social communication tasks during Session1 and Session2.



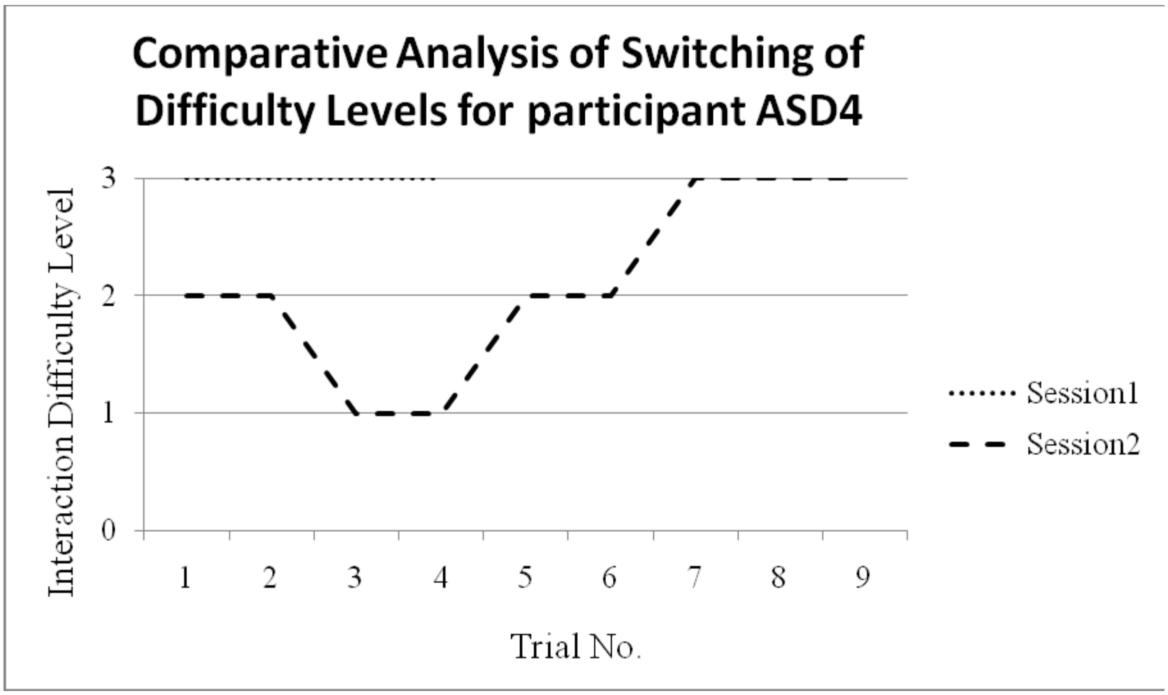
Interaction difficulty level=1: 'Easy' difficulty level; Interaction difficulty level=2: 'Medium' difficulty level; Interaction difficulty level=3: 'High' difficulty level.

Figure VI-6. Comparative Analysis of progression of VR-based social communication tasks during Session1 (task switching based on performance metric) and Session2 (task switching based on the composite effect of performance metric, behavioral viewing, and eye-physiology) for ASD1.

Participant ASD1 (Group1) progressed through six VR-based social communication tasks during Session1 and nine during Session2 (Table VI-2). The nature of progression through the social tasks during the Session1 and Session2 is represented in Fig. VI-6.

Thus during Session1 (when task switching was based on the performance metric alone), ASD1 started with VR-based social communication task of 'Easy' difficulty level as the baseline (i.e., Trial1), continued in 'Easy' difficulty level in Trial2, then moved to the 'Medium' difficulty level in Trial3, and finally to the 'High' difficulty level from Trial4-Trial6. On the other hand, Session2 equipped with the adaptive response technology predicted the engagement level of ASD1 and switched the difficulty levels based on his engagement level. During Session2, ASD1 starts at 'Medium' difficulty level as the baseline (Trial1), remains at the 'Medium' difficulty level in Trial2. Then the strategy generator (discussed in Chapter V) of the adaptive response technology predicted a lower engagement level of ASD1 which switched the task presented to ASD1 to 'Easy' difficulty level. On detecting an improvement in the engagement level of ASD1, the strategy generator moved ASD1 to the 'Medium' difficulty level in Trial4. Again, the predicted engagement level of ASD1 went to low in this trial along with 'Successful' performance (discussed in Chapter V). Thus the strategy generator maintains the difficulty level i.e., 'Medium' in Trial5. On further prediction of low engagement level of ASD1, the strategy generator lowered the difficulty level to 'Easy' in Trial6. In this trial, our adaptive response technology was capable of improving the predicted engagement level of ASD1 and now with all 'Medium' difficulty levels being executed by ASD1 our system offered the VR-based social task of 'High' difficulty level in Trial7. Thereafter, our adaptive response technology predicted an improved engagement level of ASD1 and progressed ASD1 through the ‘High’ difficulty level tasks of Trial7 to Trial9.

Let us consider the case of VR-based social communication task progression for participant ASD4 (Group1). The participant ASD4 moved through four Trials during Session1 and nine Trials during Session2 (Table VI-2). A detailed analysis of the task progression pattern for ASD4 is presented in Fig. VI-7.

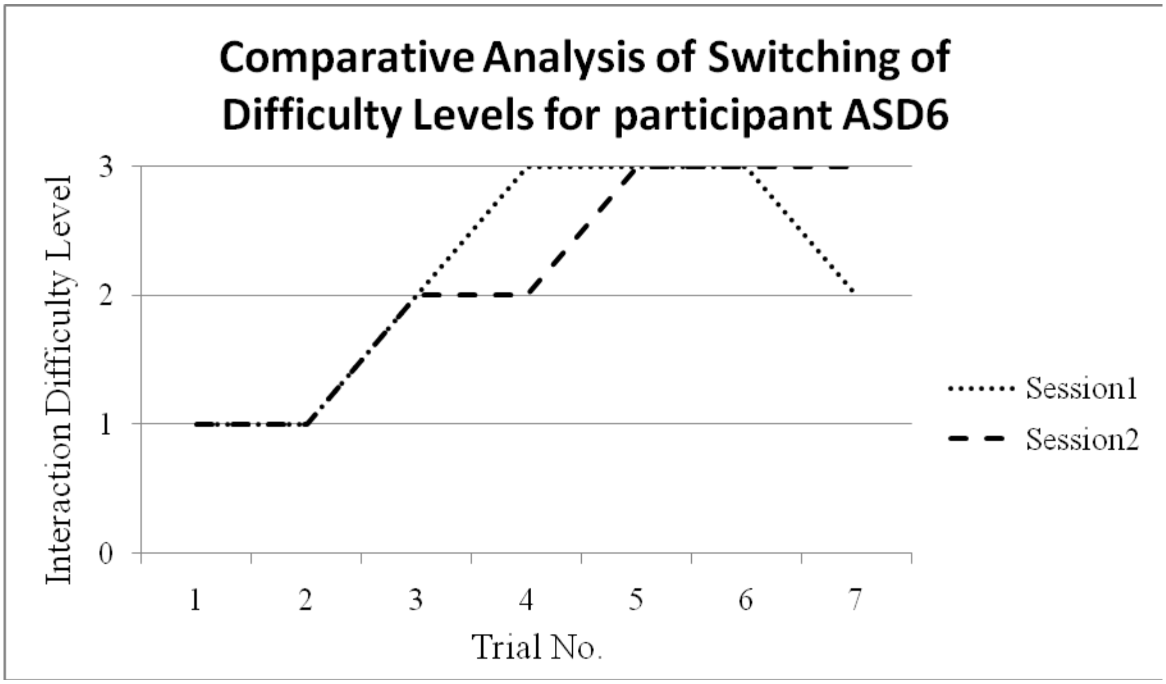


Interaction difficulty level=1: ‘Easy’ difficulty level; Interaction difficulty level=2: ‘Medium’ difficulty level; Interaction difficulty level=3: ‘High’ difficulty level.

Figure VI-7. Comparative Analysis of progression of VR-based social communication tasks during Session1 (task switching based on performance metric) and Session2 (task switching based on the composite effect of performance metric, behavioral viewing, and eye-physiology) for ASD4.

As can be seen from Fig. VI-7, during Session1, ASD4 started at the ‘High’ difficulty level as the baseline (Trial1). Further, ASD4 continued at the ‘High’ difficulty level for the subsequent trials (i.e., Trial2 – Trial4). But, we get a completely different picture for VR-based social task progression for ASD4 during Session2. During Session2, ASD4 started with the VR-based social communication task of ‘Medium’ difficulty level as the baseline (Trial1). Then ASD4 continues in the 'Medium' difficulty level during Trial2. At the end of Trial2, the strategy generator predicted a lower engagement level of ASD4 which caused ASD4 to be shifted to the 'Easy' difficulty level in Trial3. At the end of Trial3, the strategy generator detected a low predicted engagement level along with 'Successful' performance, thereby causing the adaptive response technology to maintain the same difficulty level, i.e., 'Easy' with hopes of regaining the engagement level of ASD4 during Trial4. This strategy worked out well and the strategy generator then detected an improved engagement level of ASD4 at the end of Trial4. Thus, the adaptive response technology offered a task of 'Medium' difficulty level in Trial5. Again, the strategy generator detected a reduced predicted engagement level of ASD4 at the end of Trial5. Similar to Trial3, the strategy generator maintained the task at the same difficulty level, but this time at 'Medium' difficulty level during Trial6. This strategy worked out for ASD4. Subsequently, the strategy generator detected an improved predicted engagement level of ASD4 and thereby continued the VR-based task presentation at the 'High' difficulty level from Trial7 to Trial9.

Next let us consider the case of progression of VR-based social communication tasks for participant ASD6 (Group2). ASD6 progressed though seven Trials during each of Session1 and Session2 (Table VI-3). The task progression pattern for ASD6 is presented in Fig. VI-8. From Fig. VI-8, it can be seen that during Session1, ASD6 started at the 'Easy' difficulty level as the baseline (Trial1), continued at the 'Easy' difficulty level in Trial2, then moved to 'Medium' difficulty level in Trial3. In Trial4, ASD6 moved to 'High' difficulty level and remained at the same difficulty level up to Trial6. Subsequently, ASD6 moved down to 'Medium' difficulty level in Trial7. Although the

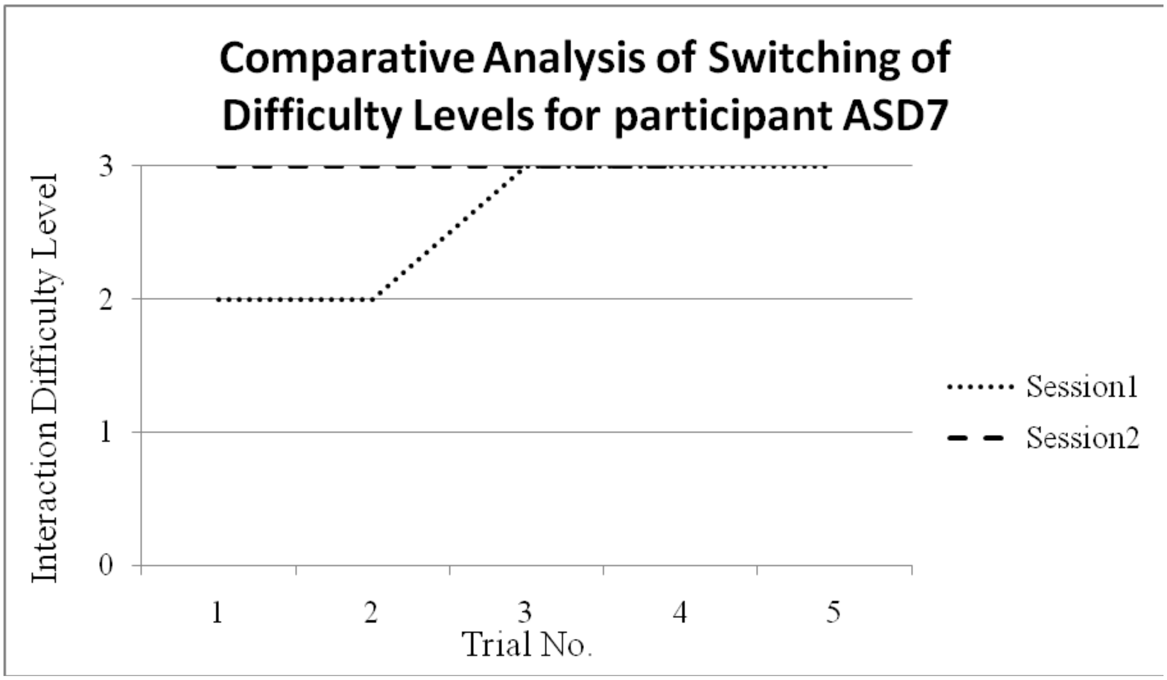


Interaction difficulty level=1: ‘Easy’ difficulty level; Interaction difficulty level=2: ‘Medium’ difficulty level; Interaction difficulty level=3: ‘High’ difficulty level.

Figure VI-8. Comparative Analysis of progression of VR-based social communication tasks during Session1 (task switching based on performance metric) and Session2 (task switching based on the composite effect of performance metric, behavioral viewing, and eye-physiology) for ASD6.

same number of trials is executed by ASD6 during Session2, yet we get a completely different picture for VR-based social task progression during Session2. During Session2, ASD6 started with the VR-based social communication task of ‘Easy’ difficulty level as the baseline (Trial1). Then ASD6 continued in the ‘Easy’ difficulty level during Trial2. At the end of Trial2, the strategy generator predicted an increased engagement level of ASD6 which caused ASD6 to be shifted to the ‘Medium’ difficulty level in Trial3. At the end of Trial3, the strategy generator detected a low predicted engagement level along with ‘Successful’ performance, thereby causing the adaptive response technology to maintain the same difficulty level, i.e., ‘Medium’ with the hope of regaining the engagement level of ASD6 during Trial4. This strategy of the strategy generator worked out well and the strategy generator then detected an improved engagement level of ASD6 at the end of Trial4. Thus, the adaptive response technology offered a task of 'High' difficulty level in Trial5. Thereafter, the strategy generator detected a continued high engagement level of ASD6, thereby causing ASD6 to carry on with the 'High' difficulty level up to Trial7.

Finally, we consider the case of progression of VR-based social communication tasks



Interaction difficulty level=1: 'Easy' difficulty level; Interaction difficulty level=2: 'Medium' difficulty level; Interaction difficulty level=3: 'High' difficulty level.

Figure VI-9. Comparative Analysis of progression of VR-based social communication tasks during Session1 (task switching based on performance metric) and Session2 (task switching based on the composite effect of performance metric, behavioral viewing, and eye-physiology) for ASD7.

for participant ASD7 (Group2). ASD7 progressed through five Trials during Session1 and four trials during Session2 (Table VI-3). The task progression pattern for ASD7 is presented in Fig. VI-9. It can be seen from Fig. VI-9, that during Session1, ASD7 started at the 'Medium' difficulty level as the baseline (Trial1), continued at the 'Medium' difficulty level in Trial2, then moved to 'High' difficulty level in Trial3 and remained at the ‘High’ difficulty level up to Trial5. However, during Session2, we find that ASD7 started with the VR-based social communication task of ‘High’ difficulty level as the baseline (Trial1). Then ASD7 continued in the ‘High’ difficulty level during Trial2. Thereafter, the strategy generator detected a high engagement level of ASD7 which caused ASD7 to stay at the ‘High’ difficulty level up to Trial4.

Individual Analysis of variation in Behavioral Viewing Pattern during Session1 and Session2

For the behavioral viewing pattern, we have considered the fixation duration (FD) while the participants look at the Face_ROI of the avatar during VR-based social communication task as a percentage of the total viewing time. This metric is particularly important as children with ASD are characterized by atypical viewing pattern in which they tend to fixate less towards the face of the communicator during social conversation (Jones, Carr, and Klin, 2008). In dyadic communication, eye-gaze serves at least five distinct communicative functions (Argyle, and Cook, 1976; Kendon, 1967): regulating conversation flow, providing feedback, communicating emotional information, communicating the nature of interpersonal relationships and avoiding distraction by restricting visual input. Eye-gaze helps control the flow of turn taking in conversations. For example, the person who is listening uses eye gaze to indicate whether he/she is paying attention, while the person who is speaking uses it to track whether the listener is still engaged in the conversation (Colburn, Drucker, and Cohen, 2000). Thus in order to encourage the participants to carry out VR-based interaction with the avatars in socially appropriate ways, our system provided gaze-based feedback (discussed in Chapter V) during Session2.

We investigated to determine whether the gaze-based individualized feedback provided by our system during Session2 has contributed to any improvement in behavioral viewing pattern among the participants. In particular, we were interested to determine the impact of the gaze-based individualized feedback on the participants' behavioral viewing pattern during dyadic communication with the avatar while using the bidirectional conversation module (discussed in Chapter V).

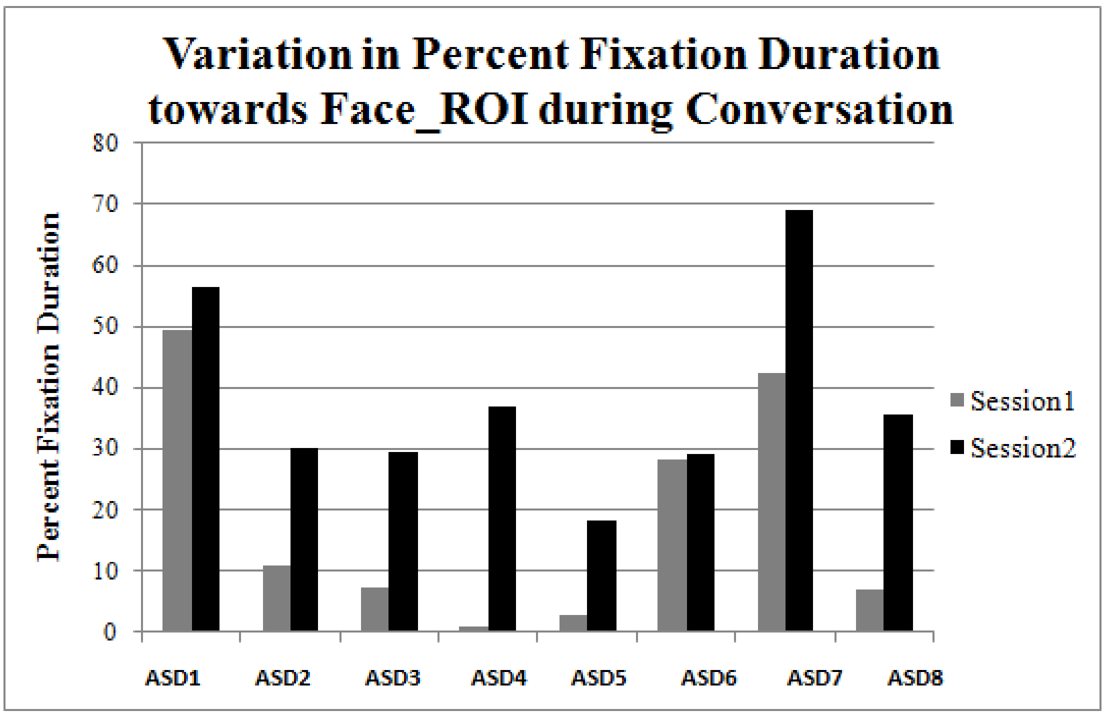


Figure VI-10. Variation in individual Percent Fixation Duration while looking towards the Face_ROI of the avatars during VR-based social conversation.

Thus from Fig. VI-10, we find that for each participant, there had been an improvement in the behavioral viewing pattern in terms of greater attention towards the face region of the avatar during the VR-based social conversation. However, the improvement for ASD6 is quite less (approximately, 1%). A dependent sample T-test between the percent fixation duration while looking towards the Face_ROI of the avatars during Session1 and Session2 indicate that the variation in the behavioral viewing pattern for the group was statistically significantly different ($p = 0.002$).

Group Analysis of variation in Behavioral Viewing Pattern with Baseline, Last Trial, and Rated Engagement Level during Session1 and Session2

Here, we present the group analysis of the behavioral viewing pattern of the participants in terms of their fixation duration (FD) while they look at the Face_ROI of the avatar during VR-based social communication task (which comprised of the participant's role as audience to the avatar's presentation and also the participants' role as social communicator while using the bidirectional conversation module) as a percentage of the total viewing time. Engagement is defined as "sustained attention to an activity or person" (NRC, 2001). In addition, Jones et al. (Jones, Carr, and Klin, 2008) have showed that one's FD while looking towards the face region of a speaker indicates social engagement. Further, FD is a valuable measure, as children with ASD often exhibit lower FD while viewing human faces than the non-human face stimuli (Anderson, Colombo, and Shaddy, 2006) during social interaction. Thus increased FD towards the face of the communicator during social communication has been shown to be indicative of greater engagement.

Our results indicate that the percentage fixation duration of the group of participants increased both from the baseline to the last trial and also with the increase in engagement of the participants, as rated by the therapist, particularly during Session2 (Fig. VI-11). From Fig. VI-11, it can be seen that during Session1, the percent fixation duration of the participant group while looking towards the face of the avatar during VR-based social communication decreased from Baseline to last trial and showed variation with increase in engagement rating. But, during Session2, where our system switched tasks based on the predicted engagement level of the participants, we find from Fig. VI-11, that all the participants fixated on the face_ROI of the avatar more during the last trial than that during the baseline. This indicates that the feedback on the behavioral viewing pattern of the participants given by our system during Session2 has encouraged the participants to improve their behavioral viewing. Also, we find that during Session2, the participants'

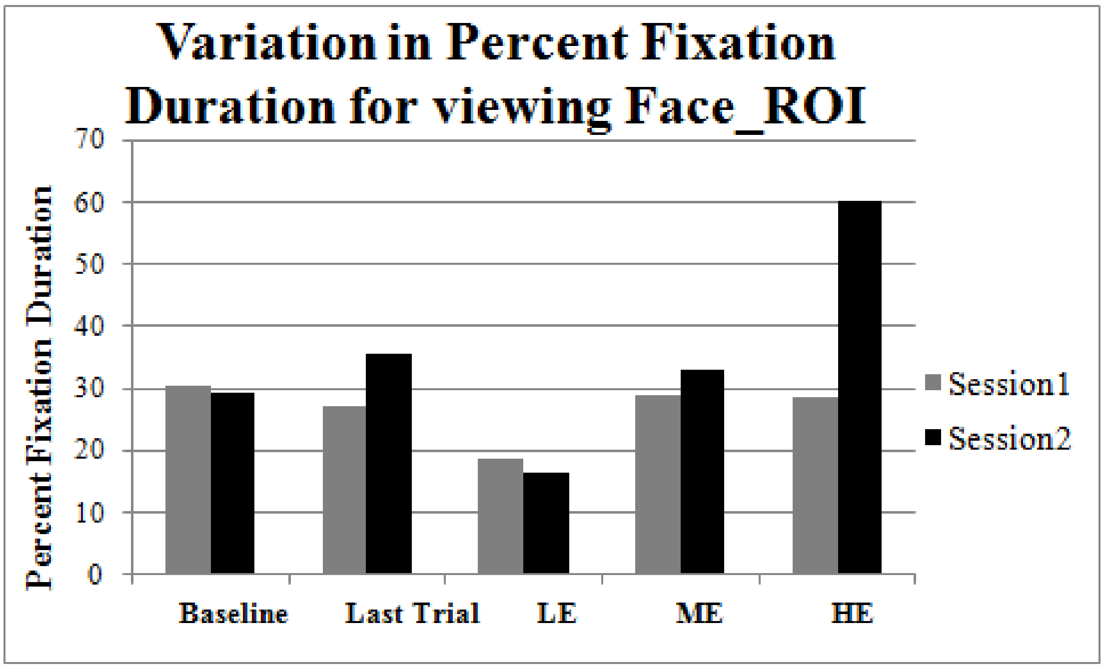


Figure VI-11. Variation in the Group Percent Fixation Duration while looking towards the Face_ROI with Baseline, Last Trial, and Level of Engagement (as rated by Therapist).
Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged). Low Engagement (LE) : corresponds to therapist's engagement rating of 1-3.
Medium Engagement (ME) : corresponds to therapist's engagement rating of 4-6.
High Engagement (HE) : corresponds to therapist's engagement rating of 7-9.

behavioral viewing pattern in terms of increased fixation on the face of the communicator (i.e., the avatar) during social conversation, improved with increased engagement level of the participants (as rated by the therapist).

Discussion

This chapter presents the results to show the effects of interacting with our developed system that is capable of intelligently adapting itself based on the predicted engagement level. The developed system is capable of switching VR-based social tasks based on one's performance metric alone (Session1) and also capable of bringing about progression of virtual social tasks based on the composite effect of one's behavioral viewing, eye physiology, and the performance metric (Session2). The results show that such a system is acceptable to the participants with ASD. Additionally, interaction with such a system featuring varying levels of social interaction difficulty can elicit variations in the affective states (e.g., engagement, enjoyment, and anxiety) level of the participants.

More importantly, the results presented in this chapter show that if we allow a computer to recognize the engagement level of an individual in terms of his/her behavioral viewing pattern, eye physiological indices, and performance during VR-based social communication tasks and apply this information as a means of flexibly taking appropriate decisions about the adaptation of the individual to the social task, then it may contribute to improved social task performance and also behavioral viewing pattern. In fact, in order to achieve effective social communication skills, one must not only acquire adequate social task performance measures, but also be able to carry out conversation in socially appropriate way (e.g., paying proper attention towards the face of the communicator). The investigation results presented in this chapter show the efficacy of the VR-based gaze-sensitive adaptive response technology to encourage individuals with

ASD to improve social communication skills, both in terms of improved performance metric and also in terms of improved behavioral viewing pattern of the participants during social conversation.

Also note that the VR-based social communication tasks offered by our system had their own limitations. For example, the presented social tasks offered limited back-and-forth conversation turns (such as, six, ten, and fourteen back-and-forth conversation turns between the participant and his/her virtual peer, i.e., the avatar, corresponding to the 'Easy', 'Medium' and 'High' level of interaction difficulty). In addition, these tasks offered limited challenge to the participants when compared to other available computer-based games, such as Pong, Anagram, etc. Also, the participants in our study were high-functioning adolescents with ASD who might find some of the social tasks somewhat less challenging than those on the low-functioning spectrum.

However, in spite of the limitations of the current system, the VR-based gaze sensitive social interactive system with adaptive response technology was capable of eliciting variations in affective states, performance scores and behavioral viewing patterns among the participants with ASD. With further improved and more challenging interaction tasks, we may expect greater variation in the affective states, performance scores, and behavioral viewing patterns among this target population.

In short, the VR-based gaze-sensitive adaptive response technology which can intelligently adapt itself based on one's predicted engagement level has the potential to promote improved task performance along with encouraging socially appropriate mechanisms (such as improved attention to the face of the communicator) during social communication. Thus this work demonstrates the efficacy and impact of VR-based gaze-sensitive social communication system with adaptive response technology to serve as an effective tool for developing tailored interventions for individuals with ASD. In a sense, deploying such technological tools could make targeted and personalized intervention a reality for these individuals and could be incorporated into complex intervention paradigms aimed at improving functioning and quality of life for older children, adolescents, and adults with ASD.

References

Anderson, C. J., Colombo, J., and Shaddy, D. J. “Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder,” J. of Cli. and Exp. Neuropsy., vol. 28, pp. 1238–1256, 2006.

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

Argyle, M., and Cook, M. "Gaze and Mutual Gaze," Cambridge, MA: Cambridge Univ. Press, 1976.

Bee, H., and Boyd, D. “The Developing Child,” 10th ed. Boston: Pearson, 2004.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., and Bailey, A. “Autism Screening Questionnaire: Diagnostic Validity,” British Journal of Psychiatry, vol. 175, pp. 444-451, 1999.

Bolman, W. M., and Richmond, J. A. “A double-blind, placebo-controlled, crossover pilot trial of low dose dimethylglycine in patients with Autistic Disorder,” Journal of Autism Developmental Disorders, vol. 29, no. 3, pp. 191-194, 1999.

Brownell, M. D. “Musically adapted social stories to modify behaviors in students with autism: four case studies,” Journal of Music Therapy, vol. 39, no. 2, pp. 117-144, 2002.

Castner, S. A., Williams, G. V., and Goldman_Rakic, P. S. “Reversal of Antipsychotic-induced Working Memory Deficits by Short-term Dopamine D1 Receptor Stimulation,” in Hyman, S.E. (ed.), The Science of Mental Health: Schizophrenia, Psychology Press, vol. 287, pp. 288-290, 2000.

Cohen, D. J., and Volkmar, F. R. “Handbook of Autism and Pervasive Developmental Disorders,” 2nd Ed., New York, NY: John Wiley and Sons, 1997.

Colburn, A., Drucker, S., and Cohen, M. “The role of eye-gaze in avatar mediated conversational interfaces,” in SIGGRAPH Sketches and Applications, New Orleans, Louisiana, USA, 2000.

Conn, K., Liu, C., Sarkar, N., Stone, W., and Warren, Z. “Towards affect-sensitive assistive intervention technologies for children with autism,” in Affective Computing: Focus on Emotion Expression, Synthesis and Recognition, J. Or, Ed. Vienna, Austria: ARS/I-Tech Education and Publishing, 2008a.

Conn, K., Liu, C., Sarkar, N., Stone, W., and Warren, Z. “Affect-sensitive assistive intervention technologies for children with autism: an individual-specific approach,” in Proceeding of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, August

2008b.

Constantino, J. N. “The Social Responsiveness Scale,” California: Western Psych. Serv., 2002.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., Metzger, L. M., Shoushtari, C. S., Splinter, R., and Reich, W. “Validation of a brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised,” Journal of Autism and Developmental Disorders, vol. 33, no. 4, pp. 427-433, 2003.

Dunn, L. M., and Dunn, L. M. “PPVT-III: Peabody Picture Vocabulary Test,” 3rd ed. Circle Pines, Minnesota: American Guidance Service, 1997.

Ehlers, S., and Gillberg, C. “The Epidemiology of Asperger syndrome: A total population study,” Journal of Child Psychology and Psychiatry, vol. 34, no. 8, pp. 1327-1350, 1993.

Eisenberg, N., Fabes, R., Murphy, B., Maszk, P., Smith, M., and Karbon, M. “The role of emotionality and regulation in children's social functioning: a longitudinal study,” Child Development, vol. 66, pp. 1360-1384, 1995.

Heiman, G. W. “Research Methods in Psychology,” 3rd Edition. Boston & New York. Houghton Mifflin Company, 2002.

Hirsch, S., Frank, T.L., Shapiro, J.L., Hazell, M.L., and Frank, P.I. “Development of a questionnaire weighted scoring system to target diagnostic examinations for asthma in adults: a modelling study,” BMC Fam Pract. vol. 5, pp. 30-38, 2004.

Javier, S. C-R. “Weighted Average Score of Customer Needs as Critical Input for QFD,” QFD International Journal, 2007.

Jones, W., Carr, K., and Klin, A. “Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder,” Arch. Gen. Psychiatry, vol. 65, no. 8, pp. 946-954, 2008.

Kendon, A. "Some functions of gaze direction in social interaction,” Acta Psychologica, vol. 26, pp. 22- 63, 1967.

Liu, C., Conn, K., Sarkar, N., and Stone, W. “Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder,” International Journal of Human-Computer Studies, vol. 66, issue 9, pp. 662-677, September 2008a.

Liu, C., Conn, K., Sarkar, N., and Stone, W. “Online affect detection and robot behavior adaptation for intervention of children with autism,” IEEE Transactions on Robotics, vol. 24, issue 4, pp. 883-896, 2008b.

Lord, C., Pickles, A., McLennan, J., Rutter, M., Bregman, J., Folstein, S., Fombonne, E., Leboyer, M., and Minshew, N. “Diagnosing autism: analyses of data from the Autism diagnostic interview,” Journal of Autism and Developmental Disorders, vol. 27, pp. 501–517, 1997.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. "The Autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism," Journal of Autism and Developmental Disorders, vol. 30, no. 3, pp. 205-223, 2000.

NRC. “Educating children with autism,” National Academy Press, Washington, DC, 2001.

Parsons, S., Mitchell, P., and Leonard, A. "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.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., and Piven, J. "Visual scanning of faces in autism," Journal of Autism and Developmental Disorders, vol. 32, pp. 249–261, 2002.

Rutter, M., Bailey, A., Berument, S., Lord, C., and Pickles, A. "Social Communication Questionnaire," Los Angeles, California: Western Psychological Services, 2003a.

Rutter, M., Le Couteur, A., and Lord, C. "Autism diagnostic interview revised WPS edition manual," Los Angeles, California: Western Psychological Services, 2003b.

Strickland, D., Marcus, L. M., Mesibov, G. B., and Hogan, K. "Brief report: two case studies using virtual reality as a learning tool for autistic children," Journal of Autism and Developmental Disorders, vol. 26, no. 6, 1996.

Wieder, S., and Greenspan, S. "Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective?," The Journal of Developmental and Learning Disorders, vol. 9, 2005.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G.R., and Vogeley, K. "It's in your eyes using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience," Social Cognitive and Affective Neuroscience, vol. 5, no. 1, pp. 98-107, 2010.

Zhang, L., Plotkin, R. C., Wang, G., Sandel, M. E., and Lee, S. "Cholinergic Augmentation With Donepezil Enhances Recovery in Short-Term Memory and Sustained Attention After Traumatic Brain Injury," Archives of Physical Medicine and Rehabilitation, vol. 85, no. 7, pp. 1050-1055, 2004.

CHAPTER VII

UNDERSTANDING PSYCHOPHYSIOLOGICAL RESPONSE WITH VIRTUAL REALITY BASED ADAPTIVE SOCIAL INTERACTIVE GAZE-SENSITIVE SYSTEM FOR CHILDREN WITH ASD

Introduction

The primary objective of this chapter is to present an analysis on the psychophysiological effects of interacting with a Virtual Reality (VR) based gaze-sensitive social communication system equipped with adaptive response technology. Children with ASD often have communicative impairments (both verbal and nonverbal), particularly regarding expression of affective states (APA, 2000; Green et al., 2002; Schultz, 2005). They often experience states of emotional or cognitive stress measured as Autonomic Nervous System activation without external expression (Picard, 2009) challenging their interests in learning and communicating. Clinicians involved in interventions must overcome these communication impairments generally exhibited by children with ASD by adeptly inferring the affective (e.g., engagement, enjoyment, and anxiety) cues of the children to adjust the intervention accordingly. However, the vulnerabilities characterizing the communicative impairments of children with ASD place limits on traditional conversational and observational methodologies. There is a growing consensus that endowing a computer with an ability to understand implicit affective cues should permit more meaningful and natural human-computer interaction

(Picard, 1997; Reeves, and Nass, 1996). Thus, for affective computing, we choose the implicit measure by using the physiological signals. The physiological signals are continuously available and are not necessarily directly impacted by the communicative impairments (Ben Shalom et al., 2006; Groden et al., 2005; Toichi, and Kamio, 2003). As such, physiological signal acquisition may represent a methodology for gathering rich data despite the potential communicative impairments of children with ASD.

In this chapter, we present our offline analysis of the impact of interaction with our developed system on physiological signals. Out of the three affective states (e.g., engagement, enjoyment, and anxiety) we carried out investigation based on the engagement of the participant, since in the present study we are mainly interested with the participant's engagement level during the VR-based social communication task. Thus we studied the effects of varying engagement level of the participants, as rated by the therapist, on the physiological signals, while the participants interacted with our system.

The results could provide valuable information to caregivers and clinicians about the specific affect-eliciting aspects of social communication such that this feedback could drive behavioral interventions that scaffold skills from basic levels of comfort. Investigation of the physiological signals may help in isolating physiological features which are more sensitive to one's engagement and thereby lead to the development of a more robust adaptive controlled system. In future, such a system can fuse the discriminatory physiological signals from the peripheral physiology (e.g., cardiovascular, electrodermal, electromyographic, etc.) and the eye physiology (e.g., blink rate and pupil diameter) for a more robust individualized adaptive system.

Experimental Investigation

Procedure

In the present study each participant participated in two sessions. Each participant was walked through the Adaptation Phase (discussed in Chapter VI). The participant was positioned in front of a task computer (C1, Fig. VI-1). Then the peripheral physiological sensors from Biopac were placed on the participant's body. The peripheral physiological signals recorded in this work are the same as those described in Chapter III with the features listed in Appendix A. These signals were collected using a Biopac MP150 system (biopac.com) and small wearable sensors were placed on a participant's left eyebrow (Corrugator Supercilii EMG), left cheek (Zygomaticus Major), upper back/lower neck muscle on right (Upper Trapezius EMG), chest (ECG and Heart Sound), neck and torso (ICG), ring and pointer finger of left hand (GSR), middle finger of left hand (PPG), and thumb on the participant's left hand (Skin Temperature). Participants used their right hand to click a mouse for interaction with the VR system. The sensors have been successfully used to collect physiological data of typical individuals (Rani, Liu, and Sarkar, 2006) and our previous work with participants with ASD (as discussed in Chapter III). The eye-tracker goggles from Arrington were then calibrated for the participant's eyes (discussed in Chapter VI). Data acquired from the eye-tracker was used to compute the real-time pupil diameter, and blink rate of the participant (discussed in Chapter V) during the VR-based social communication task. This was followed by the participant's participation in the VR-based social communication Task.

Results

The objective of this section is to present the results of an investigation to study the effects of interacting with a VR based gaze-sensitive social communication system equipped with adaptive response technology. We discuss the results of the offline analysis to show how various physiological responses are influenced when the participants interact with such a system.

Impact of Varying Engagement Levels on the Physiological Signals

We studied the implications of varying engagement level of the participants, as rated by the therapist, on their physiological signals, while the participants interacted with our system. The therapist rated the participants' engagement level using a 1-9 scale (1: not engaged, 9: very engaged). We segregated the engagement rating into Low Engagement (LE: for engagement rating 1-3), and High Engagement (HE: for engagement rating 7-9). Subsequently, we investigated the implications of varying engagement on the peripheral physiological signals (broadly categorized as electrocardiographic (ECG), skin temperature (SKT), galvanic skin response (GSR) and electromyographic (EMG)), eye physiological signals (namely, pupil diameter (PD) and blink rate (BR)).

Table VII-1. Group Analysis of Physiological Features for Low Engagement (LE) and High Engagement (HE) for Session1.

		Feature Spec.	**LE**	**HE**	**Significance (p-value)**
Peripheral Physiology	ECG	pep mean (ms)	106.66	158.70	2.3621e-06*
	SKT	temp mean (˚F)	92.73	90.16	0.0058
	EMG	Cfreq mean (Hz)	62.50	141.32	2.5809e-06*
		Cemg std (µV)	0.08	0.05	0.0326
		blink amp (µV)	0.23	0.08	6.0924e-05*
		blink std (µV)	0.15	0.05	0.0318
		Zemg mean (µV)	-1.42e-06	7.87e-09	0.0465
		Zemg std (µV)	0.12	0.01	2.9516e-10*
Eye Physiology	EYE	blink rate mean (blinks/min)	149.46	6.93	0.0153

Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged).
Low Engagement (LE) corresponds to therapist's rating on engagement of 1-3.
High Engagement (HE) corresponds to therapist's rating on engagement of 7-9.
* :$p < 0.001$.

From Table VII-1, it can be seen that a number of features from the peripheral physiological signals and one feature from the eye physiology are found to be statistically significantly different for the LE and the HE states (as rated by the therapist) of the participants for Session1.

When investigated for Session2, also a number of features from peripheral physiological signals, and one feature from eye physiology are found to be statistically significantly different (Table VII-2).

Table VII-2. Group Analysis of Physiological Features for Low Engagement (LE) and High Engagement (HE) for Session2.

		Feature Spec.	LE	HE	Significance (p-value)
Peripheral Physiology	ECG	pep mean (ms)	124.14	97.26	0.0275
		pep std (ms)	39.47	57.85	0.0198
		imp ibi std (ms)	125.84	172.36	0.0267
		HR (beats/min)	82.29	96.39	0.0017
		mean ibi ppg (ms)	713.63	610.17	0.0023
		ppgpeak mean (μV)	0.58	0.26	0.0087
	SKT	temp mean (˚F)	93.67	92.25	0.0101
	EMG	Cemg std (μV)	0.03	0.02	0.0368
		Temg std (μV)	0.02	0.08	0.0145
Eye Physiology	EYE	blink rate mean (blinks/min)	73.22	4.46	0.0151

Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged).
Low Engagement (LE) corresponds to therapist's rating on engagement rating of 1-3.
High Engagement (HE) corresponds to therapist's rating on engagement of 7-9.
* :p < 0.001.

Thus during both Session1 and Session2, our investigation results show that a number of peripheral physiological (e.g., ECG, EMG, etc.) and one eye physiological (e.g., blink rate) features are found to vary statistically significantly with the engagement level (as rated by the therapist) of the participant group.

Understanding the Psychophysiological Response (Selected Eye Physiological and Peripheral Physiological Features) with Varying Engagement Level

We carried out further investigation to analyze and understand the variation in some selected features and thereby determine whether the nature in the variation is similar to non-VR based studies. Among the eye physiological signals, we chose the Blink Rate (blinks/min) since it is found to be statistically significant with varying levels of one's engagement for both Session1 (Table VII-1) and Session2 (Table VII-2). In addition, we also analyzed the somatic and autonomic responses – heart (Cardiovascular (ECG)), and skin (Electrodermal (EDA)) which have been referred to as hallmarks of affective response (Cacioppo et al., 2000). Thus, among these peripheral physiological signals, we chose some signals broadly categorized as Cardiovascular, and Electrodermal. Among the Cardiovascular features, we considered the Heart Rate (bpm), and the mean interbeat interval Pulseplethysmogram (IBI_PPG mean) signal (ms). Among the Electrodermal features, we chose the mean of Tonic Amplitude (μS) and the mean of the Phasic Amplitude (μS).

- Variation in the Eye Physiological Feature

Literature indicates Blink Rate (BR) as an important measure of affective state. In a study conducted by Bentivoglio et al., mean BR for normal subjects was found to decrease from 17 times/min while at rest to 4.5 times/min while reading (i.e., in attentive condition) (Bentivoglio et al., 2004). Increased BR was found in schizophrenic patients in the "relaxed" condition but not in the "attentive" condition (Chen et al., 1996) and in children with ASD (Jensen et al., 2009) during task-free periods. Thus the BR has been shown to decrease with increased engagement to a task for non-VR based studies. In the present study, we obtained similar findings so far as the variation in BR with engagement is concerned.

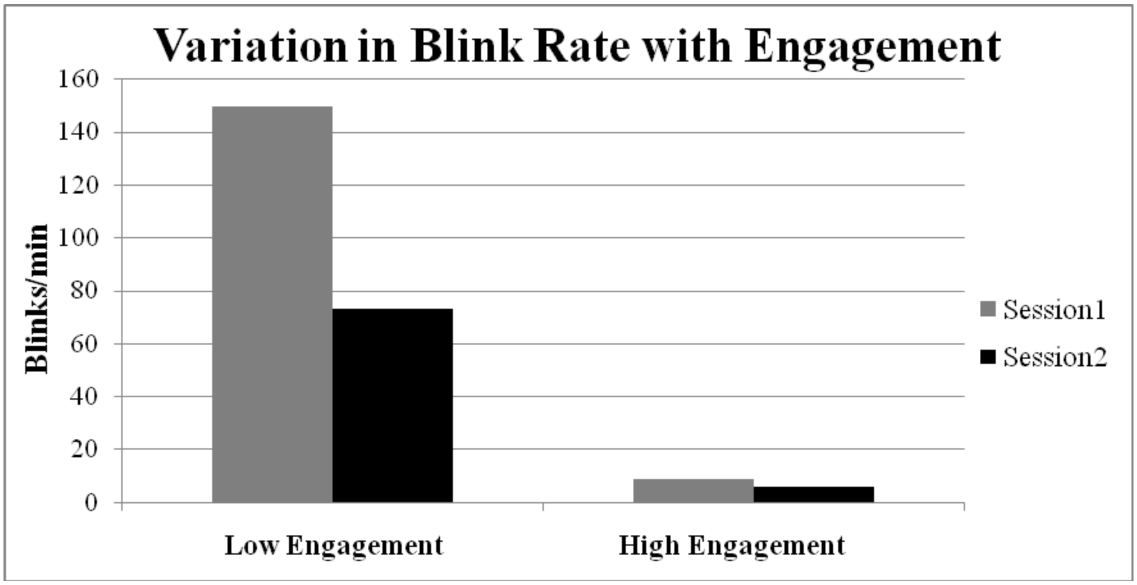


Figure VII-1. Variation in the Group Blink Rate with Engagement.
Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged).
Low Engagement corresponds to therapist's rating on engagement of 1-3.
High Engagement corresponds to therapist's rating on engagement of 7-9.

Thus from Fig. VII-1, we find that the blink rate (BR) decreases from Low Engagement to High Engagement state for both the Sessions 1 and 2, similar to non-VR based studies.

- Variation in Cardiovascular features

With sustained attention, the parasympathetic activity of the Autonomic Nervous System (ANS) is suppressed (Weber, Van der Molen, and Molenaar, 1994; Ravaja, 2002) resulting in sympathetic activation of the ANS. Also, Selvaraj et al., showed that vasoconstriction (sympathetic activation) has a very noticeable effect on the interbeat interval of the Pulseplethysmogram (IBI_PPG) pulse (Selvaraj, Santhosh, and Anand, 2007).

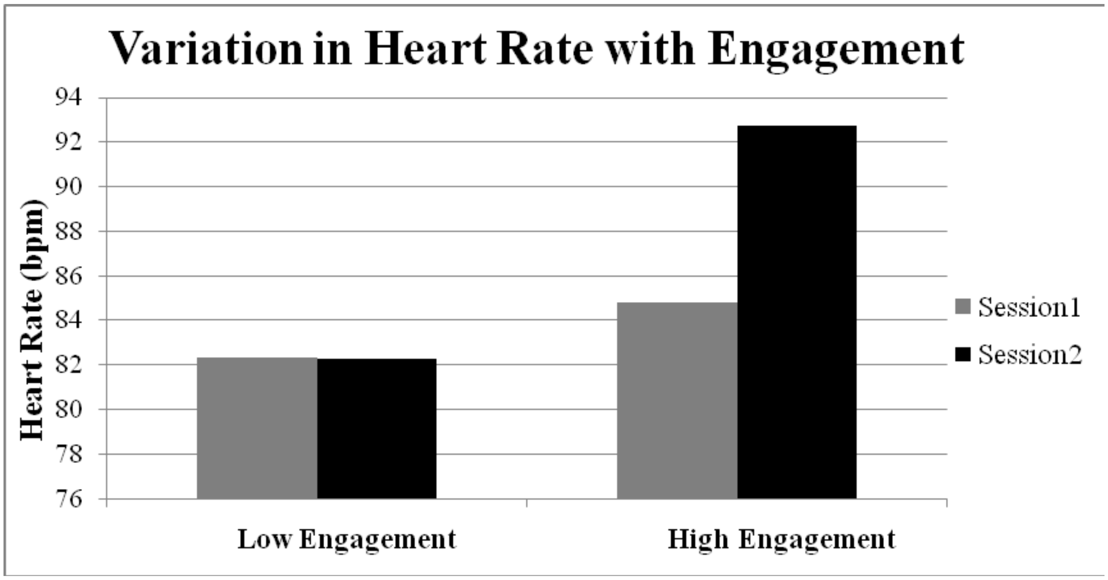


Figure VII-2. Variation in the Group Heart Rate with Engagement.
Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged). Low Engagement corresponds to therapist's rating on engagement of 1-3.
High Engagement corresponds to therapist's rating on engagement of 7-9.

From Fig. VII-2, we find that with increased attention and engagement, the Heart Rate (HR) of the participants increases for both the Session1 and 2, similar to the observation made for non-VR based studies. Also, we find that the increase in HR from the Low Engagement state to the High Engagement state is greater for Session2 than that for Session1. This implies a greater improvement in the engagement level of the participants from Session1 to session2 and the same is reflected from the variation in the HR of the participant group.

So far as the mean interbeat interval Pulseplethysmogram (IBI_PPG) is concerned, sympathetic activity results in a decrease in IBI_PPG mean and vice-versa for parasympathetic activity.

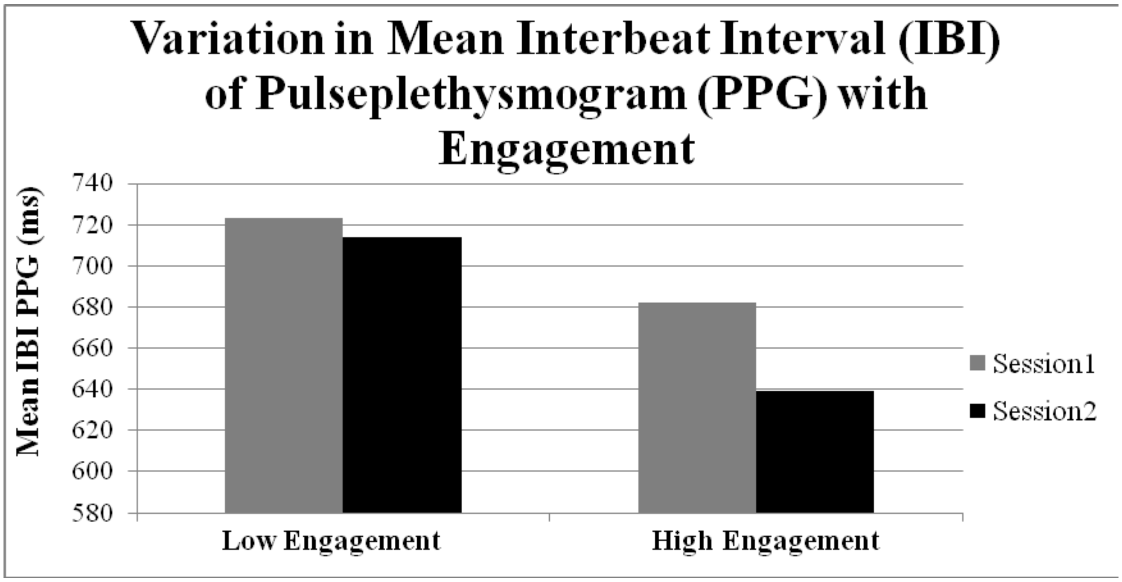


Figure VII-3. Variation in the Group mean Interbeat Interval of Pulseplethysmogram with Engagement.
Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged). Low Engagement corresponds to therapist's rating on engagement of 1-3.

From Fig. VII-2, we find that with increase in engagement level from Low Engagement to High Engagement, the HR increases for both Session1 and 2 which prompts a decrease in the mean IBI PPG. In fact, we find from Fig. VII-3 that as the participants move from Low Engagement to High Engagement state, the mean IBI PPG decreases (as is expected). However, the decrease in the mean IBI PPG from Low Engagement to High Engagement state is greater for Session2 than that for Session1.

- Variation in Electrodermal Features

Electrodermal activity (EDA), commonly known as skin conductance, is an important psychophysiological index of arousal (Lang et al., 1993). As people experience arousal, their sympathetic nervous system is activated, resulting in increased sweat gland activity and skin conductance (Ravaja et al., 2006). EDA consists of two main components e.g., tonic response and phasic response. Tonic skin conductance refers to ongoing skin conductance level in the absence of any discrete environmental events. Phasic skin conductance refers to event-related momentary increase in skin conductance.

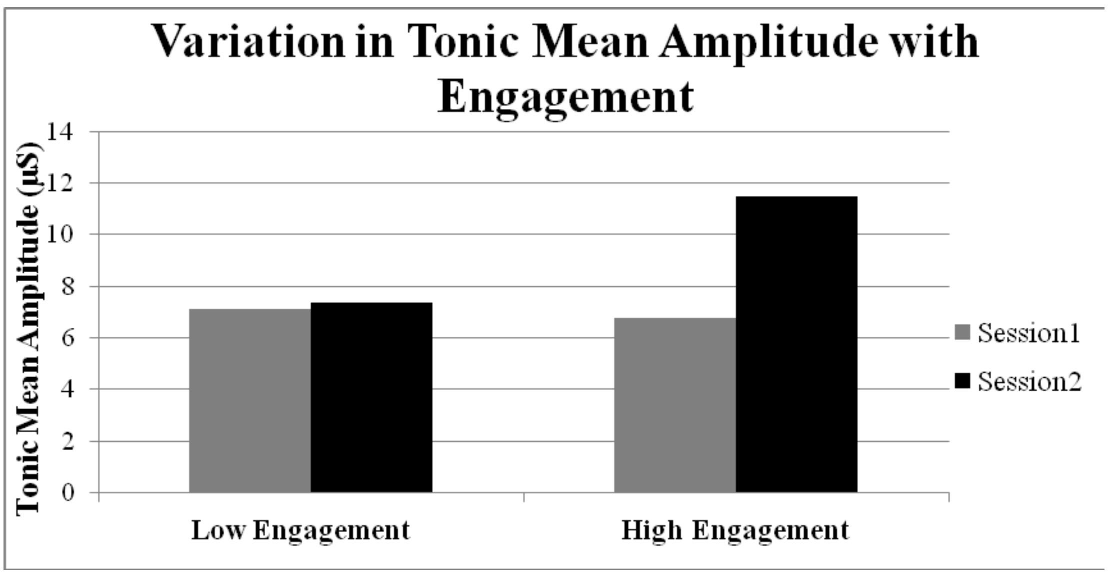


Figure VII-4. Variation in the Group Tonic mean Amplitude with Engagement.
Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged). Low Engagement corresponds to therapist's rating on engagement of 1-3.
High Engagement corresponds to therapist's rating on engagement of 7-9.

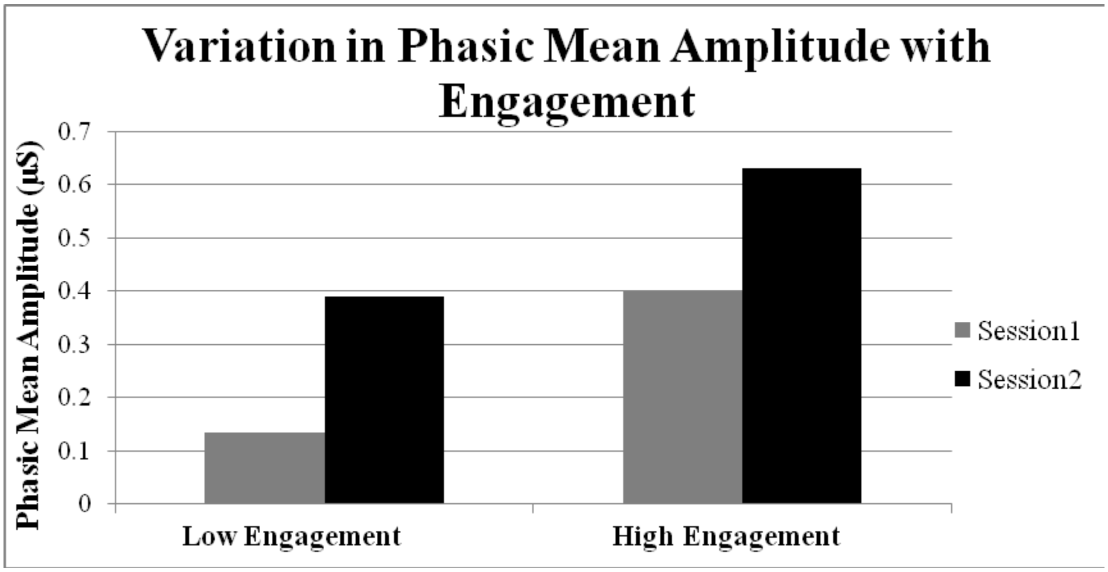


Figure VII-5. Variation in the Group Phasic mean Amplitude with Engagement.
Note : Therapist rated participants engagement level on a 1-9 scale (1-not engaged; 9-most engaged). Low Engagement corresponds to therapist's rating on engagement of 1-3.
High Engagement corresponds to therapist's rating on engagement of 7-9.

We find from Fig. VII-4 and Fig. VII-5, that although the Tonic component shows a decrease from the Low Engagement state to the High Engagement state for Session1, the Tonic component increases from the Low to High Engagement for Session2 similar to the findings from non-VR based studies. However, the mean Phasic amplitude increases from the Low Engagement to the High Engagement state (Fig. VII-5) during both Session1 and 2, similar to the findings in non-VR based literature.

Discussion

This chapter presents the results of offline analysis to study the how various physiological responses are influenced when the participants interact with our system capable of switching VR-based social tasks based on one's performance metric alone (Session1) and that capable of bringing about progression of virtual social tasks based on the composite effect of one's behavioral viewing, eye physiology, and the performance metric (Session2). Investigation into the offline analysis of the eye physiological and peripheral physiological signals collected from the participants during the VR-based social communication tasks reveals the efficacy of the system to cause variations in the physiological signals.

More importantly, the results presented in this chapter show that if we allow a computer to recognize the engagement level of an individual in terms of his/her behavioral viewing pattern, eye physiological indices, and performance during VR-based social communication tasks and apply this information as a means of flexibly taking appropriate decisions about the adaptation of the individual to the social task, then it may contribute to psychophysiological variations similar to non-VR based studies. Thus this work demonstrates the efficacy of VR-based gaze-sensitive social communication system with adaptive response technology to serve as an effective tool for developing tailored interventions for individuals with ASD using a physiology-based approach. In a sense, deploying such technological tools could make targeted and personalized intervention a reality for these individuals and could be incorporated into complex intervention paradigms aimed at improving functioning and quality of life for older children, adolescents, and adults with ASD.

References

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

Ben Shalom, D., Mostofsky, S. H., Hazlett, R. L., Goldberg, M. C., Landa, R. J., Faran, Y., McLeod, D. R., Hoehn-Saric, R. "Normal physiological emotions but differences in expression of conscious feelings in children with highfunctioning autism," Journal of Autism and Developmental Disorders, vol. 36, no. 3, pp. 395-400, 2006.

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., and Albanese, A. “Analysis of blink rate patterns in normal subjects,” Movement Disorders, vol. 12, no. 6, pp. 1028-1034, 2004.

Chen, E. Y. H., Lam, L. C. W., Chen, R. Y. L., and Nguyen, D. G. H. “Blink rate, Neurocognitive impairments, and symptoms in Schizophrenia,” Biological Psychiatry, vol.40, no. 7, pp. 597-603, 1996.

Green, D., Baird, G., Barnett, A.L., Henderson, L., Huber, J., and Henderson, S.E. “The severity and nature of motor impairment in Asperger’s syndrome: a comparison with specific developmental disorder of motor function,” Journal of Child Psychology and Psychiatry and Allied Disciplines, vol. 43, no. 5, pp. 655–668, 2002.

Groden, J., Goodwin, M.S., Baron, M.G., Groden, G., Velicer, W.F., Lipsitt, L.P., Hofmann, S.G., and Plummer, B. “Assessing cardiovascular responses to stressors in individuals with autism spectrum disorders,” Focus on Autism and Other Developmental Disabilities, vol. 20, no. 4, pp. 244–252, 2005.

Jensen, B., Keehn, B., Brenner, L., Marshall, S. P., Lincoln, A. J., and Müller, R. A. “Increased Eye-Blink Rate in Autism Spectrum Disorder May Reflect Dopaminergic Abnormalities,” Intl. Society for Autism Research, Poster Presentation, 2009.

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. “Looking at pictures: Affective, facial, visceral, and behavioral reactions,” Psychophysiology, vol. 30, pp. 261–273, 1993.

Picard, R. W. “Affective Computing,” Cambridge, MA: MIT Press. 1997.

Picard, R.W. “Future Affective Technology for Autism and Emotion Communication,” Philosophical Transactions of the Royal Society B: Biological Sciences, 2009.

Rani, P., Liu, C., and Sarkar, N. “Affective feedback in closed loop human-robot interaction,” in Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot interaction, Salt Lake City, Utah, 2006.

Ravaja, N. “Presence-related influences of a small talking facial image on psychophysiological measures of emotion and attention,” Proceedings of the 5th Annual International Workshop Presence 2002. Porto, Portugal: University Fernando Pessoa, 2002.

Ravaja, N., Saari, T., Salminen, M., Laarni, J., and Kallinen, K. “Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation,” Media Psychology, vol. 8, pp. 343-367, 2006.

Reeves, B., and Nass, C. I. “The media equation: how people treat computers, televisions, and new media as real people and places,” New York: Cambridge University Press, 1996.

Schultz, R.T. “Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area,” International Journal of Developmental Neuroscience, vol. 23, pp. 125–141, 2005.

Selvaraj, N., Santhosh, J., and Anand, S. “Feasibility of Photoplethymographic Signal for Assessment of Autonomic Response using Heart Rate Variability Analysis,” 3rd Kuala Lumpur International Conference on Biomedical Engineering, vol. 15, Part 10, pp. 391-395, 2007.

Toichi, M., and Kamio, Y. “Paradoxical autonomic response to mental tasks in autism,” Journal of Autism and Developmental Disorders, vol. 33, no. 4, pp. 417–426, 2003.

Weber, E. J. M., Van der Molen, M. W., and Molenaar, P. C. M. "Heart rate and sustained attention during childhood: Age changes in anticipatory heart rate, primary bradycardia, and respiratory sinus arrhythmia," Psychophysiology, vol. 31, pp. 164-174, 1994.

CHAPTER VIII

CONTRIBUTIONS AND FUTURE WORK

Contributions

Impairments in social communication skills are thought to be core deficits in children with Autism Spectrum Disorder (ASD). There is growing consensus that appropriately individualized intensive behavioral and educational interventions can improve core social communication vulnerabilities seen in individuals with ASD. However, there are potent barriers related to accessing and implementing appropriately individualized intensive intervention services (e.g., limited access to and availability of appropriately trained professionals, lack of available data suggesting which interventions will work better for specific children, concerns about efficacy and generalization regarding certain interventions, and exorbitant costs. Given these barriers, researchers are employing technology to develop more accessible, quantifiable, intensive and individualized intervention services for core deficit areas related to ASD. In recent years, several assistive technologies, particularly Virtual Reality (VR), have been investigated to promote social interactions in this population. However, 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. Also, 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. Thus researchers have been trying to link VR with gaze measurement in tasks such as joint-attention tasks. However, the currently available systems though may automatically detect and respond based on one's viewing pattern, cannot objectively identify and predict social engagement, understand viewing patterns, and psychophysiological effect of the specific child based on attentive indices.

Given the promise of VR-based gaze-sensitive social interaction to influence one's affective states, behavioral viewing patterns, and performance in the social task, the development of a VR-based gaze-sensitive social interactive system that can integrate the objective metrics and adapt itself to promote improved social communication skills among the children with ASD is critical. Our present research bridges this gap by closing the loop by developing a novel Virtual Interactive system with Gaze-sensitive Adaptive Response Technology that can seamlessly integrate VR-based tasks with eye-tracking techniques to intelligently encourage a participant to engage in social communication tasks while maintaining the niceties of social interactions. Specifically, such a system is capable of objectively identifying and quantifying the dynamic viewing patterns, subtle changes in eye physiological responses in real-time, and performance metric of a participant and adaptively responding in an individualized manner to foster improved social communication skills among the participants in an individualized manner. Thus, the contributions of this dissertation can be broadly categorized into two major areas, namely, (i) System Development and (ii) Development of new paradigms for technology-assisted intervention.

System Development

This involves developing intelligent software platforms that (a) can detect subtle variations in one's peripheral physiological, eye-physiological signal features, and behavioral viewing patterns in real-time and (b) seamlessly integrate these information with the VR-platform to take intelligent decisions regarding the adaptation of the individual to the VR-based social tasks. Thus, we present the

1. Design and development of a physiology-based assessment tool that identifies specific aspects of VR-based social interaction inducing affective response (e.g., engagement, enjoyment, and anxiety) in individuals with ASD. The VR-based social communication system discussed in Chapter III is capable of systematic manipulation of specific aspects of social communication. Specifically, the virtual peers (i.e., the avatars) within this system can display varying amounts of eye contact, and can vary proximity to the participant as they interact socially with the participants. The design is evaluated through a usability study that combines ratings reported from a clinical observer with physiological responses indicative of affective states of the participants, both being collected when the participants interact in the VR-based social tasks with the avatars. In the usability study, a number of peripheral physiological features, broadly categorized as cardiovascular (ECG), electrodermal (EDA), electromyographic (EMG), etc., were examined for a group of ASD and Typically Developing (TD) adolescents during social communication task presented on a VR platform for elicitation of multiple affective states. The investigation results show that the VR system provokes variations in both affective ratings and physiological signals to changes in social experimental stimuli for participants with ASD and TD participants. Thus, this work provides a vital step towards development of future social interventions using technologies such as VR for the ASD population. Since physiological signals have been shown to be differentiated during

social interaction within a virtual environment, the signals could be a useful measure in real-time VR-assisted social skill intervention, an important therapeutic instrument for addressing the core deficits in the ASD population.

2. Design and development of a VR-based gaze-sensitive social interactive system capable of providing individualized feedback based on the real-time viewing pattern of an individual interacting with the VR platform. Chapter IV presents the design details of such a system and also describes the investigation results from a usability study. Results indicate that gaze-based individualized feedback can lead to an improvement in the behavioral viewing patterns and the engagement level of participants with ASD during computer mediated VR-based social communication tasks. In addition, the usability study shows the feasibility of measuring eye physiological indices such as blink rate and pupil diameter in real-time and that they can be correlated to the emotion recognition capability of the participants with ASD. Thus, it is reasonable to believe that such a system could be used in intervention, perhaps as a supplementary tool, to allow an individual with ASD to enhance his/her social communication skills.

3. Design and development of an intelligent VR-based gaze-sensitive system with adaptive response technology. The system, as presented in Chapter V, intelligently fuses the information derived from an individual's behavioral viewing, eye physiological indices, and performance metrics through a rule-governed strategy generator during VR-based social communication tasks. Thus, the embodied intelligence of the VR-based gaze-sensitive system encourages a participant by adaptively responding in an individualized manner to participate in social communication task with improved engagement and subsequently improved performance during the social task. The results

of a usability study shows that if we allow a computer to recognize the engagement level of an individual in terms of his/her behavioral viewing pattern, eye physiological indices, and performance during VR-based social communication tasks and apply this information as a means of flexibly taking appropriate decisions about the adaptation of the individual to the social task, then it may contribute to improved social task performance. In addition, the investigation results also indicate that the VR-based gaze-sensitive adaptive response technology has the potential to promote improved task performance along with encouraging socially appropriate mechanisms (such as improved attention to the face of the communicator) to foster improved social communication skills among the individuals with ASD.

Development of new paradigms for technology-assisted intervention

This involves developing new paradigms for technology-assisted intervention. Specifically, the VR-based gaze-sensitive adaptive response technology for social communication for children with ASD intelligently fuses one's behavioral viewing, eye physiological indices and performance metric to predict one's engagement level to promote social communication skills among the target population. The presented research shows for the first time the capability of an intelligent closed loop system that adaptively responds based on the composite effect of one's behavioral viewing, eye physiological indices and performance metric during a social task to encourage social communication skills as opposed to an open loop system that responds based only on one's performance metric alone. The intelligent adaptive closed loop system provides a comprehensive platform for fostering socially appropriate mechanisms utilizing rule-governed strategy generator implemented using finite state machine automaton. Such a system with adaptive response technology has the potential to serve as an effective tool for developing intensive, individualized, and tailored interventions for individuals with ASD. In a sense, deploying such technological tools could make targeted and personalized intervention a reality for these individuals and could be incorporated into complex intervention paradigms aimed at improving functioning and quality of life for older children, adolescents, and adults with ASD.

The results of the usability study are promising. However, a much larger study must be conducted before such findings can be generalized. The presented usability study shows, in principle, that the VR-based gaze-sensitive system with adaptive response technology has the potential to be used as a supplement to real-life social skills training tasks in an individualized and intensive manner. However, we acknowledge that current findings, particularly toward skill improvement, are preliminary and limited in nature. While demonstrating proof of concept of the technology and trends of 'improved' social communication skills in a VR-based social task, questions about the practicality, efficacy, and ultimate benefit of the use of this and other technological tools for demonstrating clinically significant improvements in terms of ASD impairment remain, which will eventually be addressed by empirical investigation in the future.

Future Work

Our integrated technology fuses the behavioral viewing, eye physiological indices, and performance metrics of an individual with an aim to foster improved social communication skills among the participants in an individualized manner by adaptively encouraging the participants to continue social interaction. Though our system is capable of capturing event-marked synchronized peripheral physiological responses, such as, ECG, EDA, EMG, etc. during VR-based interaction, we did not feedback the inference from these peripheral physiological signals at this stage of research. Presently, we analyzed these responses off-line so that we can systematically isolate the most sensitive physiological features for future online feedback. In the future, an overall integrated system that fuses the behavioral viewing, most sensitive physiological features (eye-physiology and peripheral physiological signals) as derived from the investigation in the presented work, and the performance metrics, can be applied. Thus this can be a step towards more effective fusion of sensory signals to enable more robust mapping of physiology with one's engagement and thereby help to develop an improved physiology-based behavioral profiling system.

Note that the presented work requires physiological sensing that has its own limitations. For example, one needs to wear physiological sensors and put on the eye-tracker goggles, and use of such sensors could be restrictive under certain circumstances. However, none of the participants in our previous studies and in our presented study had any objection in either wearing the physiological sensors or in putting on the eye-tracker goggles. Given the rapid progress in wearable computing with small, non-invasive sensors and wireless communication, physiological sensors can be worn in a wireless manner, which could alleviate possible constraints on experimental design. Also, with increased research on remote desktop-mounted eye-tracker, experimental design may become even more simplified. In future the proposed system can be integrated with wireless sensors and remote eye-tracker thereby allowing a wider range of ASD population to be involved in the study.

Future work may also involve designing socially-directed interaction experiments with embodied robots interacting with children with ASD while systematically varying various aspects of social communication. For example, the real-time VR-based adaptive response technology described in the presented work can be integrated with 3D humanoid robot so as to produce realistic life-like social interaction with children with ASD.

APPENDIX A

Table A-1. Peripheral Physiological Indices

Physiological Response	**Features Derived**	**Label Used**	**Unit of Measurement**
Electrocardiogram	Sympathetic power from ECG	Sym	unit/square second (unit/s^2)
	Parasympathetic power from ECG	Para	unit/s^2
	Very Low Frequency Power from ECG	VLF	unit/s^2
	Ratio of powers	Para/VLF Para/Sym VLF/Sym	No unit
	Mean of IBI	IBI_ECGmean	milliseconds (ms)
	SD of IBI	IBI_ECGstd	Standard Deviation (SD, ms)
Photoplethysmogram	Mean amplitude of the peak values of the PPG signal	PPG_Peakmean	microvolt (μV)
	Maximum amplitude of the peak values of the PPG signal	PPG_Peakmax	μV
	Mean of IBI of PPG	IBI_PPGmean	ms
	SD of IBI of PPG	IBI_PPGstd	ms
	Mean Pulse Transit Time (PTT)	PTTmean	ms
	SD Pulse Transit Time (PTT)	PTTstd	ms
Heart Sound	Mean of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound	D3_HSmean D4_HSmean D5_HSmean	No unit
	SD of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound	D3_HSstd D4_HSstd D5_HSstd	No unit
Bioimpedance	Mean Pre-Ejection Period (PEP)	PEPmean	ms
	SD Pre-Ejection Period (PEP)	PEPstd	ms
	Mean of IBI of ICG	IBI_ICGmean	ms
	SD of IBI of ICG	IBI_ICGstd	ms
Electrodermal Activity	Mean tonic activity level	Tonicmean	microsiemens (μS)
	Slope of tonic activity	Tonicslope	μS/s
	Mean amplitude of skin conductance response (phasic activity)	Phasicmean	μS
	Maximum amplitude of skin conductance response (phasic activity)	Phasicmax	μS
	Rate of phasic activity	Phasicrate	peaks/min
Electromyographic activity	Mean of Corrugator Supercilii activity	Cormean	μV
	SD of Corrugator Supercilii activity	Corstd	μV
	Slope of Corrugator Supercilii activity	Corslope	μV/s
	Mean of IBI of blink activity	IBI_Blinkmean	s
	Mean amplitude of the peak values of blink activity	Blink_Peakmean	μV

Physiological Response	**Features Derived**	**Label Used**	**Unit of Measurement**
Electromyographic activity	Mean amplitude of blink activity	Blinkmean	μV
	SD of blink activity	Blinkstd	μV
	Mean of Zygomaticus Major activity	Zygmean	μV
	SD of Zygomaticus Major activity	Zygstd	μV
	Slope of Zygomaticus Major activity	Zygslope	μV/s
	Mean of Upper Trapezius activity	Trapmean	μV
	SD of Upper Trapezius activity	Trapstd	μV
	Slope of Upper Trapezius activity	Trapslope	μV/s
	Mean and Median frequency of Corrugator, Zygomaticus, and Trapezius	Cfreqmean Zfreqmean Tfreqmean Cfreqmedian Zfreqmedian Tfreqmedian	Hertz
Temperature	Mean temperature	Tempmean	Degree Fahrenheit (°F)
	Slope of temperature	Tempslope	°F/s
	SD of temperature	Tempstd	°F