Design of a novel virtual reality-based autism intervention system for facial emotional expressions identification

E Bekele[1], Z Zheng[1], U Lahiri[2], A Swanson[3], J Davidson[3], Z Warren[3], N Sarkar[1,2]

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

esubalew.bekele@vanderbilt.edu

kc.vanderbilt.edu/autos

ABSTRACT

A virtual reality (VR)-based system for evaluating facial emotion recognition ability of teenagers with autism spectrum disorders (ASD) is presented. This system is integrated with a non-contact eye tracker that allows investigation of eye gaze and eye physiological indices (e.g., blink rate) of the participants while they seek to identify the emotion displayed by the avatars in the VR environment. Performance and eye data of 12 participants (6 children with ASD and 6 typically developing children) are presented.

1. INTRODUCTION

Social communication impairments are among the core deficits of children with autism spectrum disorder (ASD) (Liu et al, 2008). Such social impairments include inability to recognize and understand facial emotional expressions (Dawson et al, 2005). Some suggest that there is significant impairment of understanding the emotional meaning of facial expressions in children with autism (Celani et al, 1999) while others seem to indicate otherwise (Castelli et al, 2005). Literature suggests that technological advancements enable application of emerging technology (Goodwin et al, 2008) such as virtual reality (VR) (Lahiri et al, 2011) and robotics (Bekele et al, 2011). Incorporating implicit cues from sensors such as eye tracker and measurement of peripheral physiological signals may help facilitating individualization and adaptation in ways that is not possible by current performance-based VR systems.

The objective of this work is twofold: (1) development and analysis of an innovative adaptive VR social system with eye tracking and physiological signals monitoring for affect and behavioral adaptation of future social tasks; and (2) performing a usability study for demonstrating the ability of the system in finding individual and group differences in eye gaze data and their implication in a VR-based social interaction and communication task. The scope of this paper is limited to the design of the new system, analysis of basic performance in emotion recognition, and understanding whether eye physiological and behavioral indices can be used to differentiate ASD and typically developing (TD) population for such tasks.

2. SYSTEM DESIGN

The system was designed for a task in which avatars narrated incidents/stories to a participant and then made relevant facial expressions to display a specific emotion. The system captured the participant's eye gaze as well as several peripheral physiological signals.

2.1 Development of the VR environment

The characters were designed to suit the age group that was targeted for the study, (13-17 years old). We designed the characters from mixamo (www.mixamo.com) and animated them in Maya. A total of 7 such avatars (4 males and 3 females) were created and used in this study. We used unity game engine (www.unity3d.com) as it allows customization in modeling, rigging and animations as it integrates seamlessly with 3D modeling and animation software pipeline such as Maya as long as they support exporting in standard formats such as FBX.

2.2 *Development of facial emotional expressions*

Individual facial emotion expressions were given 20 different weights to match them with 20 degrees of each emotion. The universally accepted Ekman's 7 emotional facial expressions (i.e., enjoyment, surprise, contempt, sadness, fear, disgust, and anger) (Ekman, 1993) and 7 phonetic visemes were created using set driven keys with each of them containing 20 weights. The visemes were for lip syncing to go with the storytelling before displaying the facial emotional expressions. Each trial as described in section III consisted of a story telling that might trigger the emotion expression that followed. Each of the 7 expression weights were divided into four animations that correspond to four intensity levels of the seven facial expressions. Each emotion expression was varied from neutral to 4 degrees of emotional expressions (low, medium, high, and extreme as judged by four observers before the study). A total of 16 storylines were manually lip-synced to an avatar and baked to key frames. All lip-syncs were then automatically copied to all the remaining avatars. Figure 1 shows the 4 degrees of an example emotion expression (surprise) with weights varied in steps of 5 from 5 to 20 in a scale of 0-20. The first one is a neutral face.

Figure 1. *Degrees of a surprise emotion expression (neutral, low, medium, high, and extreme).*

2.3 *Integration of the VR social task system and sensory modules*

The system in this study is composed of the main 3D scene running on unity for task presentation (VR task engine), eye tracker application (ET App), and physiological monitoring application (Physiology App, Figure 2). All three applications interact via a network interface in real-time.

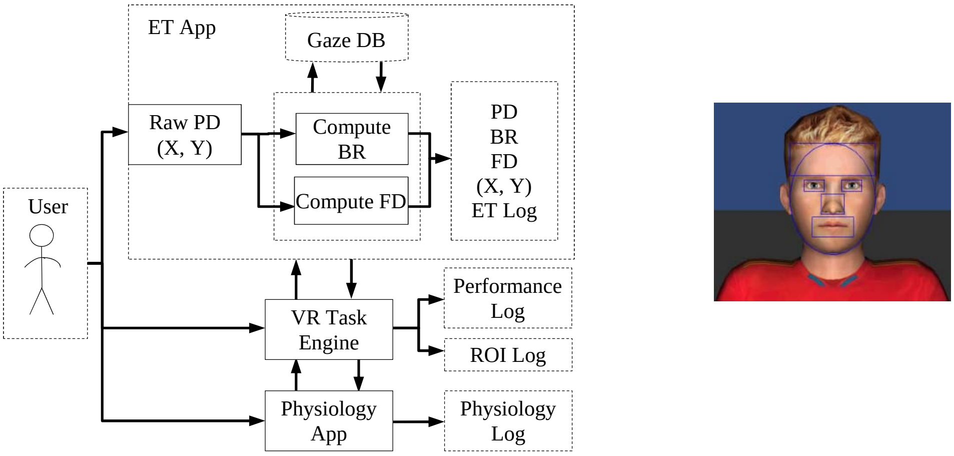

Figure 2. *VR-based system uses network interface between applications (left). The 6 face regions of interest (ROIs) (right) Note: PD: pupil diameter, BR: blink rate, FD: fixation duration, ET: eye tracker.*

The eye tracker used is the remote desktop eye tracker, Tobii X120, which enables tracking eyes non-invasively. The peripheral psychophysiological data was collected using a wireless non-invasive bio signal monitoring device called BioNomadix, by Biopac Inc. (www.biopac.com). BioNomadix is less restrictive in movement and more comfortable for participants than the wired-version. We developed the eye tracker application (ET App, Figure 2) that computes eye physiological (pupil diameter (PD) and blink rate (BR)) and behavioral (fixation durations (FD)) indices from the raw gaze data acquired from the eye tracker for each data point. These indices are known to correlate with one's engagement (Lahiri et al, 2011). The physiological monitoring application (Physiology App, Figure 2) collected raw data and logged it for offline analysis. We monitored 9 channels of peripheral physiological signals for future analysis.

2.4 *Data analysis*

The eye tracker data was analyzed offline to determine variations in gaze patterns between the children with ASD and the control group. A total of 5 ROIs (forehead, left and right eyes, nose, and mouth) were defined to meaningfully compare the data. Any region on the face outside of the 5 ROIs was considered other face region. The rest of the screen outside of the face was marked as non-face resulting in a total of 7 ROIs

©2012 ICDVRAT; ISBN 978-0-7049-1545-9

(Figure 2). The face ROI is the composite of the ellipsoid face and rectangular forehead. Each gaze point was clustered to the defined ROIs and a performance metric for each ROI was computed as the number of points lying in each ROI as a percentage of the total number of gaze points. The pupil diameter was filtered for noise rejection and interpolated linearly for missing data points. After filtering the raw fixation duration from noise spikes, it was classified in to normal fixation and saccade based on the duration. Spurious blinks were removed from the blink rate data using average range of human blink duration.

3. METHODS AND PROCEDURE

3.1 *Usability study and Participants*

A total of 6 participants with ASD (1 female and 5 males) and 6 TD (1 female and 5 males) have completed the study. Average age of ASD participants was 14.83 and that of TD participants was 14.5.

3.2 *Task and procedure*

The presented task was to identify different kinds of facial emotional expressions as displayed by the avatars. Each of the 7 avatars can display all the seven emotions in addition to 16 lip-synced story telling capabilities. There were 28 trials in a typical session corresponding to 4 (the four emotion strengths) times each 7 emotions. Each trial started with a brief (1 minute long) story that might trigger the emotion that follows. The vocal tone and the story lines were carefully selected not to influence the participant's decision making. The participant was expected to make decisions based on only the last 5 seconds of emotion expression after a minute or more of the story. After each trial was over, a menu selection of all the 7 emotions appeared for the participant to choose the emotion he/she thought the avatar just displayed.

4. RESULTS

The following scatter and fixation-saccade diagram shows example raw and behavioral eye data of a participant with autism and a typically developing participant.

Figure 3. *Scatter and fixation-saccade diagrams for of a particular trial of participant with ASD (middle) and that of a TD participant (left and right).*

4.1 *Face and Non-Face regions comparisons*

Clustering of the gaze in to the 5 different regions, other facial areas and non-face areas have been analyzed. Results suggest that the TD group spent slightly more time inside the defined ROIs while the ASD group spent a slightly more time on other facial regions. The ASD group also spent slightly longer time outside of the face region. Significant differences were observed on the mouth and forehead regions (Figure 4). The TD group spent 15.09% ($p < 0.05$) more time on the mouth than their ASD matches. Whereas the ASD group spent 16.34% ($p < 0.05$) more time on the forehead region than their TD counterparts. Given that the majority of time (about 1 minute) was spent on the story telling while only the 5 seconds were dedicated to emotional expression by the avatar, looking more to the mouth area was naturally expected. The results are consistent with ASD population focus on context irrelevant areas.

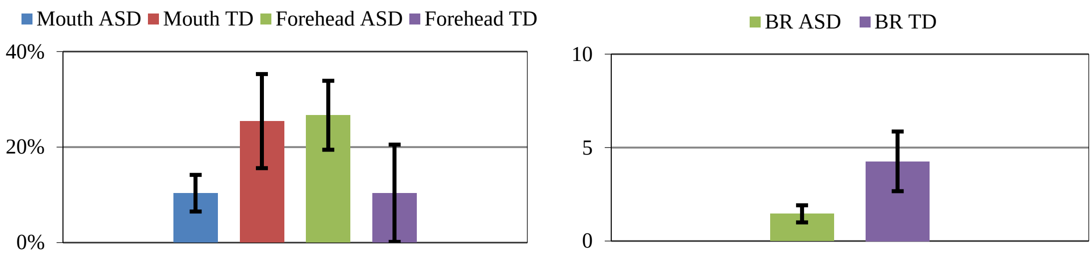

Figure 4. *Plot of percentage time spent on mouth and forehead area and BR by both ASD and TD groups.*

4.2 Behavioral and Physiological Indices

Among the eye physiological indices, blink rate (BR) is the one with significant difference between the two groups. The ASD group has far less (2.81 less blinks per trial, $p < 0.05$) blinks on average per trial than the TD group. The observation of the TD group is consistent with literature which suggests during tasks such as reading, blink rate decreases to 4.5 per minute (Bentivoglio, 1997). The unusually lower result of the ASD group might be attributed to a "sticky" attention, typically exhibited by children with autism.

4.3 Performance and confidence

There were two measures of performance. Correct identification of the presented emotions and how much confident the participant was in their choice. The ASD group performed with higher correctness percentage (13.1% more) than the TD group while the teenagers in the TD group were (11.37% more) more confident in their choices than their ASD. The performance result is inconsistent with expectations, however, factors such as relatively short duration (about 5 sec) of the emotional expression as compared to the story telling portion (close to a minute) could contribute to the outcome. TD teenagers trying to guess the succeeding emotion from the story could also be a possible confounding factor as the story didn't necessarily lead to the emotional expression displayed in a trial. However, these differences were not statistically significant.

5. CONCLUSIONS

We have designed a novel VR-based ASD intervention system that can create highly controlled and versatile emotional expressions. This system integrates an eye tracker to allow gaze analysis to investigate how gaze data and emotion recognition ability are correlated. It also integrates a peripheral physiological data acquisition system. Results of the preliminary analysis of the pilot study to evaluate the capability of the system are encouraging. However, there are several limitations of the current study that warrant consideration. Certainly a much larger study of the current system would be needed to understand how our current findings impact areas of core deficit for individuals with ASD. Additionally, a fundamental challenge to this system is that realistic social interactions, a target for this intervention tool, needs to be developed for demonstrating robust, meaningful change within and more importantly outside the VR platform.

Acknowledgements: This work was supported in part by the National Science Foundation under Grant 0967170 and by the National Institutes of Health under Grant 1R01MH091102-01A1.

6. REFERENCES

Bekele, E., U. Lahiri, et al. (2011). "Development of a novel robot-mediated adaptive response system for joint attention task for children with autism", *IEEE International Symposium on Robot Man Interaction (Ro-Man)*, Atlanta, GA, IEEE: 276-281.

Bentivoglio, A. R., S. B. Bressman, et al. (1997). "Analysis of blink rate patterns in normal subjects", *Movement Disorders* 12(6): 1028-1034.

Castelli, F. (2005). "Understanding emotions from standardized facial expressions in autism and normal development", *Autism* 9(4): 428-449.

Celani, G., M. W. Battacchi, et al. (1999). "The understanding of the emotional meaning of facial expressions in people with autism", *Journal of autism and developmental disorders* 29(1): 57-66.

Dawson, G., S. J. Webb, et al. (2005). "Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies", *Developmental neuropsychology* 27(3): 403-424.

Ekman, P. (1993). "Facial expression and emotion", *American Psychologist* 48(4): 384.

Goodwin, M. S. (2008). "Enhancing and Accelerating the Pace of Autism Research and Treatment", *Focus on Autism and Other Developmental Disabilities* 23(2): 125-128.

Lahiri, U., Warren, Z., Sarkar, N. (2011). "Design of a Gaze-Sensitive Virtual Social Interactive System for Children With Autism", *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19(4): 443-452.

Liu, C., K. Conn, et al. (2008). "Online affect detection and robot behavior adaptation for intervention of children with autism", *Robotics, IEEE Transactions* on 24(4): 883-896.