ORIGINAL PAPER

Assessing the Utility of a Virtual Environment for Enhancing Facial Affect Recognition in Adolescents with Autism

Esubalew Bekele · Julie Crittendon ·
Zhi Zheng · Amy Swanson · Amy Weitlauf ·
Zachary Warren · Nilanjan Sarkar

Abstract Teenagers with autism spectrum disorder (ASD) and age-matched controls participated in a dynamic facial affect recognition task within a virtual reality (VR) environment. Participants identified the emotion of a facial expression displayed at varied levels of intensity by a computer generated avatar. The system assessed performance (i.e., accuracy, confidence ratings, response latency, and stimulus discrimination) as well as how participants used their gaze to process facial information using an eye tracker. Participants in both groups were similarly accurate at basic facial affect recognition at varied levels of intensity. Despite similar performance characteristics, ASD participants endorsed lower confidence in their responses and substantial variation in gaze patterns in absence of perceptual discrimination deficits. These results add support to the hypothesis that deficits in emotion and face recognition for individuals with ASD are related to fundamental differences in information processing. We discuss implications of this finding in a VR environment with regards to potential future applications and paradigms targeting not just enhanced performance, but enhanced social information processing within intelligent systems capable of adaptation to individual processing differences.

Keywords Autism spectrum disorders · Virtual reality · Facial expressions · Adaptive systems

E. Bekele (✉) · Z. Zheng · N. Sarkar
Department of Electrical Engineering and Computer Science,
Vanderbilt University, 518 Olin Hall 2400 Highland Avenue,
Nashville, TN 37212, USA
e-mail: esubalew.bekele@vanderbilt.edu

J. Crittendon · A. Weitlauf · Z. Warren
Department of Pediatrics and Psychiatry, Vanderbilt University,
Nashville, TN, USA

J. Crittendon · A. Swanson · A. Weitlauf · Z. Warren
Vanderbilt Kennedy Center, Treatment and Research Institute
for Autism Spectrum Disorders, Vanderbilt University,
Nashville, TN, USA

Z. Warren
Department of Special Education, Vanderbilt University,
Nashville, TN, USA

N. Sarkar
Department of Mechanical Engineering, Vanderbilt University,
Nashville, TN, USA

Introduction

With an estimated prevalence of 1 in 88 in United States (CDC 2012), effective treatment of autism spectrum disorders (ASD) is a pressing clinical and public health issue (IACC 2012). Given recent rapid developments in technology, it has been argued that specific computer and virtual reality (VR) based applications could be harnessed to provide effective and innovative clinical treatments for individuals with ASD (Bolte et al. 2010; Goodwin 2008; Parsons et al. 2004). VR technology possesses several strengths in terms of potential application to ASD intervention, including: malleability, controllability, replication ability, modifiable sensory stimulation, and the potential capacity to implement individualized intervention approaches and reinforcement strategies. VR can also depict various scenarios that may not be feasible in a "real-world" therapeutic setting, given naturalistic social constraints and resource challenges (Kandalaft et al. 2012; Parsons and Mitchell 2002). As such, VR appears well-suited for creating interactive skill training paradigms in core areas of impairment for individuals with ASD.

A growing number of studies are investigating applications of advanced VR to social and communication related intervention (Blocher and Picard 2002; Parsons et al. 2004; Mitchell et al. 2007; Welch et al. 2010a, b; Kandalaft et al. 2012; Lahiri et al. 2012, 2013; Ploog et al. 2013; Bekele et al. 2013). Increasingly, researchers have attempted to develop VR and other technological applications that respond not only to explicit human–computer interactions (e.g., utilization of keyboards, joysticks, etc.), but to dynamic interactions such as those incorporating eye gaze and physiological measurements (Wilms et al. 2010; Lahiri et al. 2012; Bekele et al. 2013). However, most of the existing VR environments applied to assistive intervention for children with ASD are designed to build skills based on aspects of performance alone (i.e., correct or incorrect and some performance metrics), and thus may limit individualization of application as they are not capable of responding to individual gaze cues and processing patterns and solely depend on final correct or incorrect recognition of the emotion. VR systems that not only gauge performance on specified tasks but also automatically detect eye gaze or other physiological markers of engagement may hold promise for additional optimization of learning (Lahiri et al. 2012; Welch et al. 2009, 2010a, b). The present study is a preliminary investigation into the development and application of a VR environment capable of utilizing gaze patterns to understand how adolescents with ASD process salient social and emotional cues in faces. Ultimately, the goal is to alter VR interactions by giving the environment the ability to respond to both patterns of performance and gaze. For individuals with ASD, such enhancements may improve attention to and processing of relevant social cues across dynamic interactions both within and beyond the VR environment.

Among the fundamental social impairments of ASD are challenges in appropriately recognizing and responding to nonverbal cues and communication, including challenges recognizing and appropriately responding to facial expressions (Adolphs et al. 2001; Castelli 2005). In particular, individuals with ASD may have impaired face discrimination, slow and atypical face processing strategies, reduced attention to eyes, and unusual strategies for consolidating information viewed on other's faces (Dawson et al. 2005). Although children with ASD have been able to perform basic facial recognition tasks as well as typically developing peers in certain circumstances (Castelli 2005), they have shown significant impairments in efficiently processing and understanding complex, dynamically displayed facial expressions of emotion (Bolte et al. 2006; Capps et al. 1992; Dawson et al. 2005; Weeks and Hobson 1987).

A number of research groups have attempted to utilize computer technology to improve facial affect recognition

(Golan and Baron-Cohen 2006; Golan et al. 2010; LaCava et al. 2010). Detailed and comprehensive reviews of computer-assisted technologies for improving emotion recognition (Ploog et al. 2013) note that specific facial recognition skills can be improved using existing systems. How well these skills transfer to real-world settings, however, remains largely unstudied. One study showed that performance in computerized training correlated with brain activation (Bolte et al. 2006). Otherwise, skill transfer, or generalization of skills beyond the computer to meaningful social situations, has been limited if examined at all.

For multiple reasons, a VR emotion recognition paradigm paired with markers of performance and gaze processing may be a more effective way to teach and generalize skills. Specifically, VR can provide a controlled and replicable environment where specific recognition skills can be tested and taught in a dynamic fashion. Where earlier intervention approaches have tended to rely on static pictures or presentations to teach recognition skills, a VR environment approximates properties of facial expressions as they dynamically develop, appear, and change in reality. Further, the capacity for dynamic presentation means that stimuli can be controlled and altered to teach increasingly subtle displays of facial affect. Skills can be practiced in a variety of virtual scenarios (e.g., altering avatars, context, environment) to promote generalization. In addition, the capacity to embed an eye tracker within a VR environment makes it possible to individualize training sessions beyond simple measures of performance, in this case whether the facial expression is correctly identified. More specifically, eye-tracking may show *how* gaze has been used to process the emotion and as such provide essential feedback to inform intervention (Lahiri et al. 2012). Because research has consistently documented ASD-specific processing differences in this regard (Klin et al. 2002; Pelphrey et al. 2002), developing platforms for gathering such information about affect recognition processing (with the potential for on-line adaptive change of the VR stimuli itself, e.g., highlighting, occluding, and shifting attention to specific features) may yield an efficient training tool for promoting both learning within the environment as well as generalization beyond the virtual environment.

In the current study, we present a brief overview of the VR and eye-tracking system task, results from a preliminary comparison of facial affect recognition performance between our ASD and TD samples (see Bekele et al. 2013), group differences regarding gaze patterns during facial affect detection, and signal-detection analyses examining potential perceptual differences at different levels of emotion expression. Although explicitly not an intervention study, the ultimate aim of this proof-of-concept and user study was to establish the utility of a dynamic VR and eye-tracking system with potential application for

Table 1 Participant characteristics by group

Metric	ASD				TD				Statistics
	Male		Female		Male		Female		t value
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	(df = 18)
IQ	120.1	8.2	113.0	12.0	113.5	11.6	107.5	2.5	ns
Age	14.5	0.9	15.5	1.5	14.5	0.9	15.0	2.0	ns
SRS*	78.0	4.2	87.5	0.5	40.3	3.2	35.5	0.5	19.19
SCQ*	17.1	7.6	17.5	5.5	2.4	1.9	0.0	0.0	6.15

SRS Social Responsiveness Scale, *SCQ* Social Communication Questionnaire

* $p < .05$

intervention platforms. We hypothesized that participants with ASD would show poorer facial affect recognition than the comparison sample, particularly for subtler depictions of affect (e.g., disgust, contempt). We further hypothesized that participants with ASD would have longer response times and less confidence in their recognition decisions. Finally, regarding eye gaze, participants in the ASD group were expected to attend less frequently to relevant facial features during the task than the control group.

Method

Participants

Two groups of teenagers between 13 and 17 years of age participated: 10 teenagers with ASD (M = 14.7, SD = 1.1) and 10 typically developing controls (TD: M = 14.6, SD = 1.2) matched by age. Participants in the ASD group were recruited through an existing university clinical research registry. All ASD participants received a clinical diagnosis of ASD from a licensed clinical psychologist and had scores at or above clinical cutoff on the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 2000). Estimates of cognitive functioning for those in the ASD group were available from the registry [e.g., tested abilities from either the Differential Ability Scales (DAS; Elliott 2007), the Stanford Binet (SB; Rold 2003), or Wechsler Intelligence Scale for Children (WISC; Wechsler 2003)]. These core diagnostic assessment procedures had all occurred within 2 years of participation in the current protocol. For children with several measurements within this registry, the most recent available measures were utilized. Participants in the TD group were recruited through an electronic recruitment registry accessible to community families and were administered a Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler 1999) as part of participation in the current study. There were no significant differences between groups regarding overall cognitive abilities (see Table 1).

To index initial autism symptoms and screen for autism risk in the control group, parents of all participants completed the Social Responsiveness Scale (SRS; Constantino and Gruber 2005) and the Social Communication Questionnaire—Lifetime Version (SCQ; Rutter et al. 2003). Scores above 70 on the SRS and above 15 on the SCQ are in the clinically significant range. No child in the comparison group scored in the risk range of either the SRS or SCQ. Table 1 presents sample demographics and descriptive data for both groups.

Overview of Procedure

All informed consent and human subject procedures were approved by our local university institutional review board. Participants attended a 1-h session and were accompanied by a parent. Participants were seated in the experiment room in front of a desktop monitor that displayed the tasks and the desktop remote eye tracker. They were provided with only a very general instruction that they would play a computer game where they would be asked questions about interactions with avatars. After eye tracker calibration, participants were presented with VR tasks where they were asked to identify different kinds of emotional facial expressions as displayed by the avatars. Following the animated facial expression, a menu appeared on-screen with instructions to choose the expression depicted. Response selection automatically prompted the system to begin the next trial.

System Design and Development

Our VR-based facial affect presentation system utilized the popular game engine *Unity* (www.unity3d.com) by Unity Technologies. We used the Unity game engine for task deployment because it allows users to customize modeling,

Fig. 1 Representative avatars

rigging and animations, and also integrates well with 3D modeling and animation software pipelines.

The system was integrated with a remote desktop eye tracker, Tobii X120, with a 120 Hz frame rate that allowed for a free head movement of 30 cm × 22 cm × 30 cm in the width, height and depth dimensions, respectively at a sitting distance of 70 cm. An eye-tracking application was integrated with the VR engine using a network interface that communicated through a distributed network (Bekele et al. 2013). This system network monitored and logged gaze fixation data in relation to time-stamps corresponding to task presentations within the VR environment. To examine differences between social and nonsocial gaze, we predefined seven regions of interest (ROIs) for use in the current protocol. The seven ROIs were: forehead, left and right eyes, nose, mouth, other face areas (i.e., excluding forehead, eyes, nose and mouth regions), and non-face regions. We took the center of the described areas and defined rectangular regions with certain width and height to maximally cover the ROIs without overlaps except for the overall face ROI, which was a composed of a rectangular forehead and an elliptical face. These ROIs were comparable to the areas mentioned in prior work examining scan patterns in ASD samples (see Pelphrey et al. 2002).

Development of Avatar Facial Expressions

Avatars (four males, three females; see Fig. 1) were designed to suit the targeted age group for the study (13–17 years old) using characters from Mixamo (www.mixamo.com) that were animated in Maya (www.autodesk.com).

Emotion expression stimuli were developed based on Ekman's universal facial expressions that include joy, surprise, contempt, sadness, fear, disgust, and anger. Extensive investigation of these emotions has shown high agreement across cultures in assignment of these emotions to faces (Ekman 1993) and these emotions have been utilized in prior works examining emotion recognition and scan patterns with ASD samples (Pelphrey et al. 2002). Each of the seven emotional expressions were divided into four animations that corresponded to four emotion intensity levels (low, medium, high, and extreme; see Fig. 2), which

allowed us to moderate difficulty levels and thus increase response variance.

After creating the animated facial expressions in the VR environment, we tested their characteristics and recognition with a sample of nine typically developing adult university students to validate that the expressions depicted the desired emotions. In the college sample, performance exceeded the minimal threshold for chance for each emotion. Ordered from highest to lowest accuracy, the percentage of correct responses to emotional expressions across all levels of manipulation was: sadness (97 %), joy (92 %), anger (86 %), surprise (81 %), disgust (72 %), contempt (58 %), and fear (53 %). These results mirror the extant literature in facial expression recognition as expressions of disgust, contempt, and fear tend to evoke lower performance than do more basic expressions such as sadness and joy (Castelli 2005).

Development of Ambiguous Statements

Ambiguous statements, by which we mean a statement that can be potentially attributable to more than one emotional state, were selected from a pool of statements written by clinical psychologists consulting with our VR development team and used as context for each emotion recognition task trial. Such ambiguous statements were necessary so that the participants' choices of emotional state were not influenced by the verbal statements provided by the avatar. The scripts were created with content ranging from incidents at school to interactions with family and friends appropriate to this age group. This initial pool of statements was presented to a focus group of undergraduate students and research assistants who were asked to denote whether these statements could be seen as corresponding to the seven emotional states designated in protocol. Only statements where emotional content was ambiguous (e.g., potentially attributable to the options of two or more emotional states) were included in the final pool of statements (e.g., "Netflix is finally carrying my favorite TV series. The fee is $20 a month."). Final statements were chosen randomly from this available pool and then recorded in a neutral tone of voice by confederates that were similar in age to the study sample.
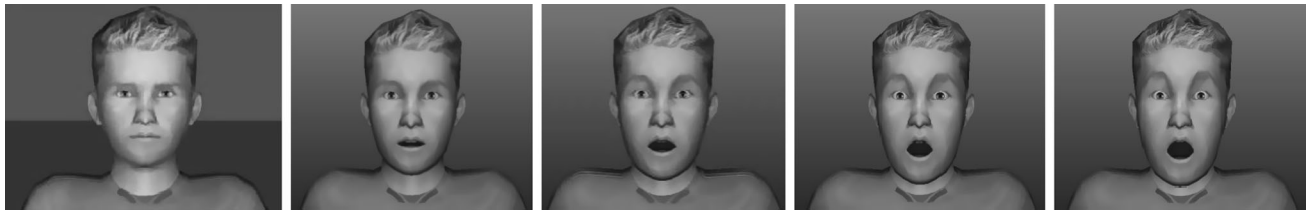
Fig. 2 Example expressions at each level of intensity, from mildest (*left*) to most intense (*right*)

Task Trial Procedure

The system presented a total of 28 trials corresponding to the seven emotional expressions at the four specified levels of intensity. The task order for all 28 trials was randomized across all participants to control for potential order and learning effects. The seven emotions with each emotion having four levels of arousal resulted in 28 emotional expressions. These 28 expressions were pooled at random and assigned to each trial for each participant. Each trial was 15–20 s long. For the first 10–15 s, the avatar maintained a neutral facial expression and narrated a lip-synced statement of ambiguous content. The avatar then produced a facial expression of specified and altered intensity (i.e., moving from neutral to low, medium, high, or extreme emotion) that was displayed for 5 s.

Following the facial expression animation, a menu appeared on-screen prompting participants to identify the emotion displayed from a list of four possible emotions. This menu primarily included choices that were rated as potentially applicable to the story content by the student focus group. Unbeknownst to participants, we recorded response latency in all trials that was defined as the duration of time between the appearance of the menu and the point at which the response was submitted. After selecting an emotion from the menu, participants were asked to complete a 5-point Likert-type rating of their confidence in their response (i.e., 1 = Extremely Confident to 5 = Extremely Unconfident).

Results

Facial Affect Recognition

We examined participant performance using the following metrics: accuracy, response latency, and ratings of response confidence. We utilized a mixed ANOVA approach including intensity as a factor in examining each of these variables. To test our hypothesis that participants with ASD would be less accurate at identifying emotions, we conducted a two-way ANOVA examining performance by group including intensity of emotion as a factor. There

Table 2 Average recognition accuracy by emotion

Emotion	ASD (%)		TD (%)		Statistics	
	Mean	SD	Mean	SD	t value (df = 18)	p value
Anger	75.00	20.41	70.00	28.38	0.45	.66
Contempt	17.50	20.58	15.00	17.48	0.29	.77
Disgust	70.00	30.73	62.50	21.25	0.64	.53
Fear	42.50	20.58	22.50	27.51	1.84	.08
Joy	57.50	23.72	55.00	19.72	0.26	.80
Sadness	90.00	17.48	90.00	17.48	0.00	1.00
Surprise	52.50	21.89	47.50	27.51	0.45	.66

* $p < .05$

Table 3 Signal detection analysis of performance for discriminability

Levels	d'		C	
	ASD	TD	ASD	TD
Low	1.2302	0.9776	0.7227	0.7793
Medium	1.9249	1.6452	0.5963	0.6426
High	1.6995	1.5917	0.6332	0.6521
Extreme	1.9249	1.5917	0.5963	0.6521

was a significant difference in performance between the intensity levels of emotions (F (1, 3) = 4.8, $p < .005$). Further analysis using Tukey's HSD multiple comparison procedure revealed the low level of intensity was significantly different from the other three levels. There was no significant difference in performance between ASD and TD (F (1, 3) = 2.62, $p > .05$) groups. There was no significant interaction in performance between groups and intensity levels of emotions (F (1, 3) = .1, $p > .05$). Post-hoc analysis examining performance differences by emotion revealed no group differences across any of the included seven emotion groups (see Table 2).

In order to follow-up on the significant difference results for intensity and performance we conducted a signal-detection analysis in an attempt to distinguish between discriminability (d') and subjective bias (C). Results of this analysis suggested that both discriminability and bias were

Table 4 Group differences on performance metrics (out of 28 presentations)

Metric	ASD		TD		Statistics	
	Mean	SD	Mean	SD	t value (df = 18)	p value
Overall accuracy	16.30 (58.22 %)	2.67 (9.53 %)	14.50 (51.79 %)	3.78 (13.49 %)	1.23	.23
Response time* (s)	11.30	4.47	7.50	0.97	2.63	.02
Confidence ratings*	75.52 %	10.69 %	90.09 %	5.75 %	−3.80	.001

* $p < .05$

comparable within and between groups at all but the lowest level of emotion expression (See Table 3).

We also hypothesized that participants in the ASD group would take longer to respond to the dynamically displayed expressions than their typical peers. Response latency was operationalized as the duration of time between the presentation of the on-screen emotion recognition menu and the submission of a response. There was a significant main effect for group (F (1, 3) = 12.4, $p < .001$) between our ASD (mean response time in seconds, M = 11.3, SD = 4.47) and TD sample (M = 7.5, SD = 0.97). There was no main effect significant difference in response latency rating between the intensity levels of emotions (F (1, 3) = 0.59, $p > .05$) nor a significant group by intensity interaction (F (1, 3) = .07, $p > . 05$).

Finally, we hypothesized that confidence ratings would be lower in the ASD group than in the TD group. Despite similar performance accuracy between the two groups, there was a main effect for group (F (1, 3) = 25.97, $p < .001$) regarding confidence ratings with the ASD group significantly lower) (M = 75.52 %, SD = 10.69 %) than the TD group (M = 90.09 %, SD = 5.75 %) (See Table 4). Again there was no effect for intensity (F (1, 3) = 0.79, $p > .05$) nor a significant group by intensity interaction.

Although our facial expression animations were developed from images widely used in research, we ran follow-up analyses to ensure responses were not related to factors unique to the newly developed stimuli. We examined the data for any consistent response patterns among answers that were incorrect (i.e., misclassified facial expression). Specifically, we analyzed response patterns for all incorrect answers across all participants and for all trials. Results revealed that across both groups, participants' most common patterns of error related to misclassification of contempt as disgust (52.5 % of the time), disgust as anger (25 %), and joy as surprise (21.3 %). When examining overall correct identification, participants accurately identified joy, anger, fear, and sadness more than 50 % of the time across all conditions, showing less accuracy with surprise (50 %), contempt (16.25 %), and fear (32.5 %).

Facial Affect and Gaze Patterns

Our primary analytic approach for interpreting gaze patterns was to analyze between group differences in time spent looking toward predefined locations, or ROIs. We examined differences in looking between face and non-face regions as well as time spent on critical areas of the face tied to the dynamic display of emotions (e.g., mouth, forehead, eyes). These analyses were conducted for two distinct time periods: (1) the scripted conversation portion of the task, wherein the avatar maintained a neutral expression while talking, and (2) the period of time during which the avatar shifted facial expression from neutral to the targeted expression.

We then further examined the percentage of time participants spent looking at different parts of the avatar's face during the neutral and emotional expression conditions (see Table 5). Participants in both groups spent similar amounts of time looking at the avatar's face as well as eyes, nose, and "other face areas." Significant differences emerged in the amount of time spent looking at the avatar's mouth and forehead, however, with ASD participants looking more at the forehead (neutral: M = 22.98 %, t = 2.56, $p < .05$; emotion: M = 19.46 %, t = 2.41, $p < .05$) and TD participants looking more at the mouth (neutral: M = 28.46 %, t = −3.76, $p < .01$; emotion: M = 26.19 %, t = −3.24, $p < .01$) across both conditions. A visual map of these different looking times is presented in Fig. 3.

Discussion

In this preliminary investigation, we used a novel VR system and a range of newly developed avatar facial expressions to examine performance and process differences between adolescents with and without ASD on tasks of facial emotion recognition. Our hypotheses were partially supported, such that group differences emerged when examining response latency, eye gaze, and confidence in responses, but not participants' ability to correctly identify emotions.
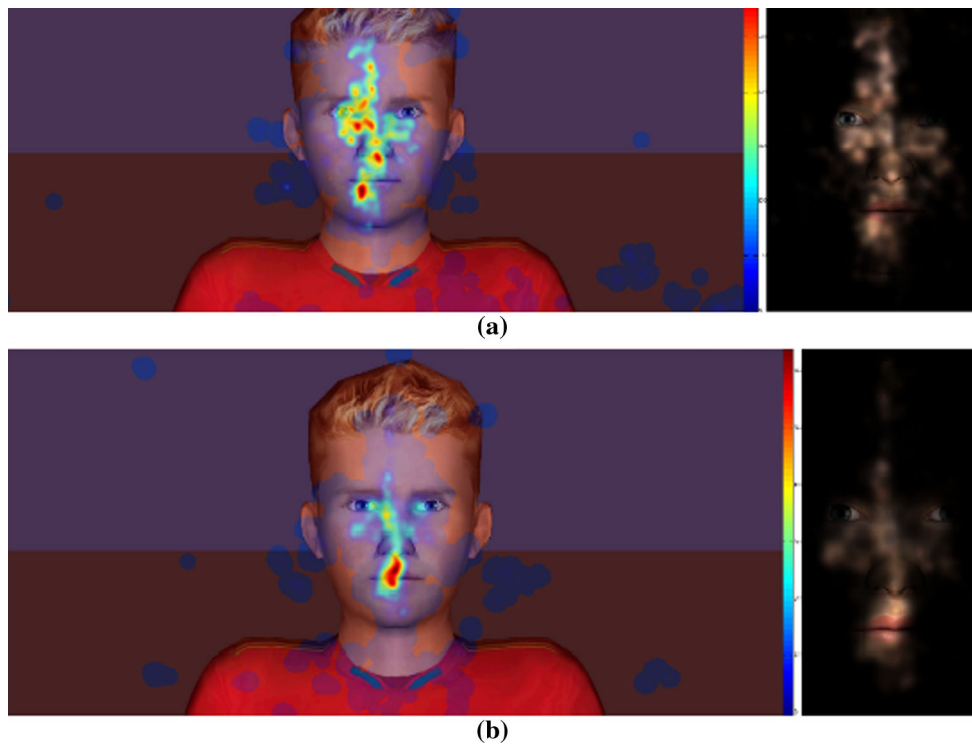
Fig. 3 Intergroup comparison gaze visualizations (*heat maps* and *masked scene maps*) **a** combined gaze in the ASD group for all the trials and all participants **b** combined gaze in the TD group for all the trials and all participants

Table 5 Amount of time spent looking at mouth and forehead ROIs across groups

	Neutral			Emotion expression		
	Mouth*	Forehead*	Face	Mouth*	Forehead*	Face*
ASD (%)	11.91	22.98	86.87	14.09	19.46	85.09
TD (%)	28.46	9.31	80.37	26.19	8.14	91.50
Statistics						
t value (df = 18)	−3.76	2.56	−1.47	−3.24	2.41	−2.23
p value	.0014	.0196	.1582	.0045	.0269	.0388

* $p < .05$

Contrary to our expectations, no significant difference was found between groups in terms of overall performance accuracy. Previous work has noted that children, particularly children with very high cognitive abilities as was the case in the current study, are capable of identifying basic static emotions at a high level of accuracy (Castelli 2005). However, we had originally hypothesized that our dynamic displays of emotion within the VR environment might prove challenging for children with ASD. Further, we had designed the task such that lower levels of emotional expression would hopefully be quite challenging to discriminate to avoid specific floor effects of this assessment. While this task did challenge accuracy in performance (e.g., low level of accuracy across tasks), it did so across group. Ultimately, this result of comparable performance

supports the possibility that if children with ASDs have difficulty with affect recognition, the problem may not be at the level of basic naming or recognition which mirrors decades worth of extant literature on affect recognition using static images (Adolphs et al. 2001; Castelli 2005; Dawson et al. 2005) rather in complex recognition tasks and in the presence of social context stimuli (Table 4).

Hypothesized significant group differences did emerge when additional performance metrics and gaze patterns were examined. Children in the ASD group overall took longer to respond to stimuli and had lower confidence in their answers than participants in the TD group. In addition, participants in the TD group showed generally symmetrical eye gaze to relevant areas on the face during the animations and demonstrated a significant bias toward

relevant components of the facial expression (i.e., mouth ROI). However, those in the ASD group showed a pattern of more distributed attention with less focus to relevant stimulus features (Fig. 3).

Notably the riggings utilized to display animations for the involved avatars had a majority of connection features (i.e., relevant features that dynamically changed to display the emotion) within the very region of interest where TD participants focused a majority of their gaze (i.e., mouth ROI). Each facial rig for each character contains 6 out of the 17 facial bones (35 %) compared to just one bone on the forehead (5.88 %) and 4 bones around each eye (23.5 %). Moreover, by the design of the avatars, the eyes covered relatively less area than the other facial structures and fewer eye gaze data points lied in the eye areas resulting in insignificant difference in gaze towards the eyes than the other areas such as mouth and forehead. This result mirrors previous findings regarding facial processing and eye gaze for people with ASD (Castelli 2005; Celani et al. 1999; Dawson et al. 2004; Hobson 1986; Anderson et al. 2006; Hobson et al. 1988), suggesting atypical attention to irrelevant features. Such a difference in gaze attention may help explain why participants in the ASD group took longer to achieve similar results as the TD group regarding what facial expressions were being displayed (Ploog 2010). The faster response times of the TD group may be due to focusing quickly on relevant features when making judgments about social stimuli. Indeed, previous research has shown that when children with ASD are cued to look at salient facial ROI during a facial emotion recognition task, they focused on non-core facial feature areas (i.e., the areas of face that have less relevant features needed for emotion recognition) than the control group (Pelphrey et al. 2002; Klin et al. 2002). Further, our discriminability analysis suggests that the differences that emerged between groups were not reflective of perceptual differences in discriminating visual stimuli across emotional intensity levels. Rather difference seems to be at their core related to how such information is gathered and synthesized. As such, developing tools such as VR displays of emotion that can dynamically display emotive expressions and potentially guide and alter gaze processing and attention to enhance facial recognition, may prove a valuable intervention approach over time. Specifically, such a system might be capable of enacting changes not simply recognizing emotions, but changes related to how such emotions are recognized. Further, addressing social vulnerabilities on a processing level may result in changes that more powerfully generalize than current approaches for enhancing skills in social interaction, as real-world social interactions often require fast and accurate interpretation of, and response to, others' verbal and nonverbal communication. Finally, it is important to note that this difference

in processing of social information is likely not circumscribed to emotion processing and may be tied to many other challenges and vulnerabilities associated with ASD. As such, work addressing differences in processing not just performance may be important in designing intervention paradigms across other areas of skill vulnerability.

The present study is preliminary in nature. Our goal was to test participant response to our novel animated facial expression stimuli. Cognitive scores in the current sample are higher than general population norms, which may limit generalization of findings. This sample characteristic may be irrelevant, however, as all of our teenaged participants (both with and without ASD) were less accurate in identifying more subtle expressions than the college students included in a previous verification sample. Observed response variability, then, may be better explained by maturation of social cognition rather than general cognitive ability, per se. Our study's small sample size, although characteristic of exploratory studies in general, weakens the statistical power of the results. Readers are thus cautioned against making broad inferences at a population level from these data. Further, the work highlights processing differences but ultimately does not demonstrate that we can alter this processing component within a VR context. Despite this major limit, our results represent a potential meaningful step towards developing a dynamic VR environment for ASD intervention that focuses on intervening on this target. In future work, we plan to refine our stimuli and further develop our VR program to test the ability of the system to advance social and individualize interventions that focus on altering the fundamental methods in which individuals take in social information. We anticipate that providing a safe environment in which to practice such social skills, with ongoing monitoring of performance, engagement, and processing will lead to individuals with ASD having more confidence and successful navigation of parallel tasks in the real-world.

References

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience, 13*(2), 232–240.

Anderson, C. J., Colombo, J., & Shaddy, D. J. (2006). Visual scanning and pupillary responses in young children with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology, 28*(7), 1238–1256.

Bekele, E., Zheng, Z., Swanson, A., Crittendon, J., Warren, Z., & Sarkar, N. (2013). Understanding how adolescents with autism

respond to facial expressions in virtual reality environments. *IEEE Transactions on Visualization and Computer Graphics, 19*(4), 711–720.

Blocher, K., & Picard, R. W. (2002). Affective social quest: emotion recognition therapy for autistic children. In K. Dautenhahn, A. Bond, L. Canamero, & B. Edmonds (Eds.), *Socially intelligent agents—creating relationships with computers and robots* (pp. 1–8). Dordrecht: Kluwer.

Bolte, S., Hubl, E., Feineis-Matthews, S., Prvulovic, D., Dierks, T., & Poustka, F. (2006). Facial affect recognition training in autism: Can we animate the fusiform gyrus? *Behavioral Neuroscience, 120*(1), 211–216.

Bolte, S., Golan, O., Goodwin, M. S., & Zwaigenbaum, L. (2010). What can innovative technologies do for autism spectrum disorders? *Autism, 14*(3), 155–159.

Capps, L., Yirmiya, N., & Sigman, M. (1992). Understanding of simple and complex emotions in non-retarded children with autism. *Journal of Child Psychology and Psychiatry, 33*(7), 1169–1182.

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

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders, 29*(1), 57–66.

Centers for Disease Control and Prevention (CDC). (2012). Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, United States, 2008. *Morbidity and Mortal Weekly Report, 61*(SS03), 1–19.

Constantino, J. N., & Gruber, C. P. (2005). *The Social Responsiveness Scale*. Los Angeles, CA: Western Psychological Services.

Dawson, G., Webb, S. J., Carver, L., Panagiotides, H., & McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science, 7*(3), 340–359.

Dawson, G., Webb, S. J., & McPartland, J. (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*, 384–392.

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome and high functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology, 18*(2), 589–615.

Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., et al. (2010). Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. *Journal of Autism and Developmental Disorders, 40*(3), 269–279.

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.

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry, 27*(3), 321–342.

Hobson, R. P., Ouston, J., & Lee, A. (1988). Emotion recognition in autism: Coordinating faces and voices. *Psychological Medicine, 18*(4), 911–923.

Interagency Autism Coordinating Committee (IACC). (2012). IACC Strategic Plan for Autism Spectrum Disorder Research—2012 Update. December 2012. Retrieved from the U.S. Department of Health and Human Services Interagency Autism Coordinating Committee website: http://iacc.hhs.gov/strategic-plan/2012/index.shtml.

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2012). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-012-1544-6.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, American Medical Association, 59*(9), 809–816.

LaCava, P., Rankin, A., Mahlois, E., Cook, K., & Simpson, R. (2010). A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD. *Autism, 14*(3), 161–178.

Lahiri, U., Warren, Z., & Sarkar, N. (2012). 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.

Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z., & Sarkar, N. (2013). Design of a virtual reality based adaptive response technology for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 21*(1), 55–64.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–225.

Mitchell, P., Parsons, S., & Leonard, A. (2007). Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(3), 589–600.

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disabilities Research, 46*(5), 430–443.

Parsons, S., Mitchell, P., & Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 34*(4), 449–466.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*(4), 249–261.

Ploog, B. O. (2010). Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 40*(11), 1332–1349.

Ploog, B. O., Scharf, A., Nelson, D., & Brooks, P. J. (2013). Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(2), 301–322.

Rold, G. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Nelson Education. Rolling Meadows, IL: Riverside.

Rutter, M., Bailey, A., Lord, C., & Berument, S. (2003). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children (WISC-IV)* (4th ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Weeks, S. J., & Hobson, R. P. (1987). The salience of facial expression for autistic children. *Journal of Child Psychology and Psychiatry, 28*(1), 137–152.

Welch, K., Lahiri, U., Liu, C., Weller, R., Sarkar, N., & Warren, Z. (2009). An affect-sensitive social interaction paradigm utilizing

virtual reality environments for autism intervention. *Human-Computer Interaction: Ambient, Ubiquitous and Intelligent Interaction, 56*(12), 703–712.

Welch, K., Sarkar, N., & Liu, C. (2010a). Affective modeling for children with autism spectrum disorders: Application of active learning. In L. Berhardt (Ed.), *Advances in medicine and biology* (pp. 297–310). New York: Nova Science Publishers.

Welch, K. C., Lahiri, U., Sarkar, N., & Warren, Z. (2010b). An approach to the design of socially acceptable robots for children with autism spectrum disorders. *International Journal of Social Robotics, 2*(4), 391–403.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G., & Vogeley, K. (2010). It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognition and affective neuroscience. *Social Cognitive and Affective Neuroscience, 5*(1), 98–107.