COMPUTATIONAL ANALYSIS OF GAZE BEHAVIOR IN AUTISM DURING INTERACTION WITH VIRTUAL AGENTS

Zeeshan Akhtar[1], *Tanaya Guha*[2]

[1]IIT Kanpur, India [2]University of Warwick, UK

ABSTRACT

Individuals with Autism spectrum disorder (ASD) are known to have significantly impaired social interaction and communication abilities. These impairments are characterized by their difficulties in using and perceiving non-verbal cues, such as facial expressions. The difficulty in processing communicators facial expressions is often attributed to the atypical gaze patterns in individuals with ASD. We present a computational study of gaze behavior in adolescents with ASD during their interaction with virtual agents (avatars) in a virtual reality based social communication platform. We study the implications on the subjects pupil response (pupil diameter changes) and looking pattern (fixation coordinates and duration) when exposed to the avatars demonstrating context-relevant emotional expressions. The data related to fixation and pupil response is collected using a commercial eye-tracker for subjects with and without ASD during their interactions with the avatars. This data is analyzed to investigate how the pupil response dynamics and fixation patterns of the ASD group differ from their typically developing peers. Our results indicate that communicators facial expressions can significantly affect the gaze behavior of the ASD subjects. We also observe reduced complexity in the pupil response dynamics, and lower synchrony between pupil response and fixation pattern in the ASD group.

Index Terms— Autism, emotion, virtual reality, gaze.

1. INTRODUCTION

Autism spectrum disorder (ASD) is a complex and pervasive neurodevelopmental disorder characterized by significant impairment in social interaction and communication abilities. Such impairment is characterized by deficits in perceiving social cues from communicator's face [1], and in using and responding to various non-verbal aspects of communication, such as facial emotional expressions [2, 3, 4]. Impairments in using and processing eye gaze in deriving non-verbal social cues from communicator's face is a well known characteristic of Autism [5, 6, 7]. Research has shown that individuals with ASD face difficulty in perceiving affective facial expressions, and this difficulty is also related to their eye gaze patterns [7]. Another study [8] reported that individuals with ASD, while watching videos of social scenes (social dynamic stimuli), fixate less to the eye region and more to one's body. Several other studies have reported similar observations where subjects with ASD fixate more to human body and other objects as compared to human faces [9, 10].

Recently, virtual reality (VR) based communication systems have emerged as an effective alternative to traditional intervention systems and education programs in Autism [11, 12, 13, 14]. Such systems offer increased accessibility, lower cost and assessment effort, along with a safe, interactive environment for intervention and learning. Virtual simulators, such as *LIFEisGAME* [13] and *virtual cafe* [14], have been created to help children with ASD to improve their emotion recognition skills and performance in social tasks.

This paper presents a computational study of gaze behavior of adolescents with ASD during their interaction with a VR-based social communication system. In particular, we are interested in studying the implications on the subjects' eye physiological index (pupil response measured in terms of pupil diameter changes) and looking pattern (quantified by fixation coordinates and duration) when exposed to virtual agents (*avatars*) demonstrating context-relevant emotional expressions. To acquire relevant eye gaze data, we have used a recently developed VR-based social communication platform designed particularly to help individuals with Autism [15, 12]. This platform creates realistic 3D environments of various social situations, where an avatar narrates a story related to a social situation to a participant (see Fig.1 for an example). Based on the story, the avatar asks a few questions to the participant, and the participant answers using a menu-driven interface. The avatars were designed to display three basic emotional expressions: *happy*, *angry*, *neutral* (see Fig.1). During the course of interaction with the avatars, participant's eye gaze data was collected using a commercially available eye tracking device in terms of two signals: *pupil response* measured in terms of pupil diameter changes, and *fixation coordinate and duration*. The data from each session is labeled with the corresponding to primary emotional expression displayed by the avatar during that session. The eye gaze data thus collected is analyzed using different computational techniques to uncover the subtle differences in gaze behavior between the subjects with ASD and their typically developing (TD) peers in the context of the affective facial expressions displayed by the avatars.

We first study if the ASD and TD groups differ in the ways they fixate on the face vs. other regions in a virtual scene. Next, we investigate the dynamics of their pupil response using an information theoretic measure of data complexity. We consider the hypothesis that the pupil dynamics (a physiological signal) in subjects with ASD will have reduced complexity compared to their TD peers. Our hypothesis is motivated by the existing evidence of physiological signals being associated with atypical, and often reduced measures of complexity under disease or disorder [16, 17]. The notion of complexity is fundamental to time series data, and can be interpreted as the rate at which new information is generated by the underlying time varying system. Thereafter, we analyze pupil response and fixation patterns jointly to understand the synchrony (coordination and dependence) between the two signals for the ASD and TD groups. Our major observations for the ASD group include (i) lesser fixation time on avatar's faces during the display of affective facial expressions, (ii) lower complexity in pupil response dynamics, and (iii) lower synchrony between pupil response and fixation pattern.

2. VR-BASED DATA ACQUISITION

The VR platform used for this work presents a realistic 3D environment of social situations, where a virtual agent (avatar) narrates a

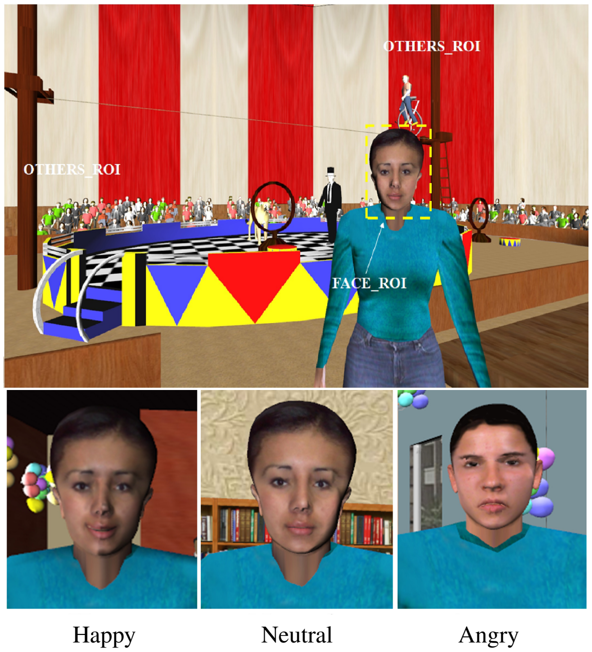

Fig. 1. (a) A sample social situation in the VR-based communication system we employed for data acquisition. The scene is annotated with two regions of interest (ROI): face and others. (b) Avatars displaying happy (left), neutral (center) and angry (right) expressions.

story related to different social situations to a participant [12, 15]. The avatars narrate social stories that have been shown to help in teaching social skills to individuals with ASD in previous studies [18]. The avatars show context-relevant emotion during the storytelling process, and are lip synced with prerecorded audio files for the narration. The avatars are capable of displaying three basic emotions, such as happy, angry and neutral (see Fig. 1). The graphical user interface for the VR intervention system (VR task environment and the 3D avatars) was designed using the VIZARD software from Worldviz.

We recruited eight high functioning ASD and eight TD individuals for our study. The ASD participants were above the threshold of clinical measures for Autism i.e. Social Responsiveness Scale [19] mean(std) = 69.1(6.9), and Social Communication Questionnaire [20]: 14.0(5.1). All TD participants were well below the clinical measures. The ages of ASD participants (14.7(3.3)) and TD participants (15.5(3.3)) were not significantly different ($p = 0.33$). The participants were asked to listen to the social stories narrated by the avatars with context-relevant emotions based on its experience in a social situation. At the end of each story, the avatar asked few questions related to the narrated story. During each interactive storytelling session, the participants' gaze data was collected. A lightweight wearable eye tracker (ViewPointEyeTracker- 2.9.2.5 from Arrington Research) was used to acquire this data in a time synchronized manner with the VR task. Participants' raw gaze data (pupil diameter, fixation coordinates and duration) was acquired in real time and stored in a database. Later, minor preprocessing of the data was performed on the raw data to remove any noise due to blinking and movement artifacts. Based on the recorded fixation coordinates, we labeled the participants' fixation patterns corresponding to the two Regions of Interest (ROIs): *face* and *others* (see Fig. 1).

The participants were free to withdraw from the study had they

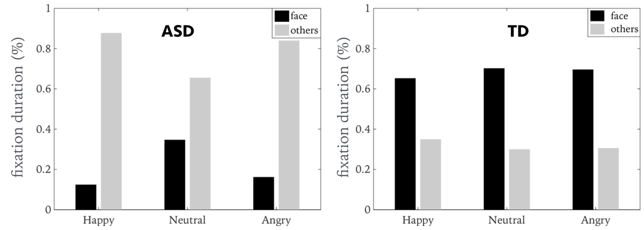

Fig. 2. Comparison of fixation duration (in % of total time) between *face ROI* and *Others ROI*

felt uncomfortable at any point of time. Nevertheless, all of them completed the study without reporting any difficulty. An exit survey conducted at the end of the study showed that all the participants liked interacting with our VR system and expressed interest in future participation.

3. DATA ANALYSIS AND INTERPRETATION

Our computational study aims at understanding the differences in gaze patterns between the ASD and TD groups as they interact with the avatars. We first study if the groups differ in the way the fixate on the face vs. other regions. Next, we investigate the dynamics of their pupil response using an information theoretic measure of data complexity. We then analyze pupil response and fixation duration jointly to understand the synchrony (coordination and dependence) between the two quantities for the ASD and TD groups.

3.1. Fixation on face vs. other region

As mentioned in Section 2, using the fixation coordinates, we could differentiate between the fixation of the participants into *face* and *others* ROIs. Fig. 2 plots the average fixation duration of the ASD and TD subjects for the *face ROI* and the *others ROI* for the ASD and TD groups. Clearly, ASD subjects fixate more on the regions other than the avatar's face while TD subjects fixate more on the face region. This observation is consistent with previous studies on eye gaze where ASD subjects were found to focus more on the body of the communicator than eye region [8, 9, 10]. Also note that when the avatars display emotional expressions, such as happy or angry, the ASD subjects fixate even less on the face region while for the same condition the TD subjects fixate more on the face region.

3.2. Complexity of pupil response

In this section, we investigate whether the ASD and the TD groups have similar or differing patterns of complexity in their pupil response. Although the interpretation of complexity varies with the physiological parameters being studied and the developmental condition being investigated, there is significant evidence for various pathological processes being associated with atypical and often reduced measures of physiological complexity [16, 17, 21]. We hypothesize that subjects with ASD will exhibit lower complexity than their TD peers. To compute complexity, we use an information theoretic measure of dynamic complexity, called the *sample entropy* [22, 23], which is often useful in the context of physiological time series analysis. In particular, we used a multiscale variant of the sample entropy method called the *refined composite multiscale entropy* (RCMSE) [23].

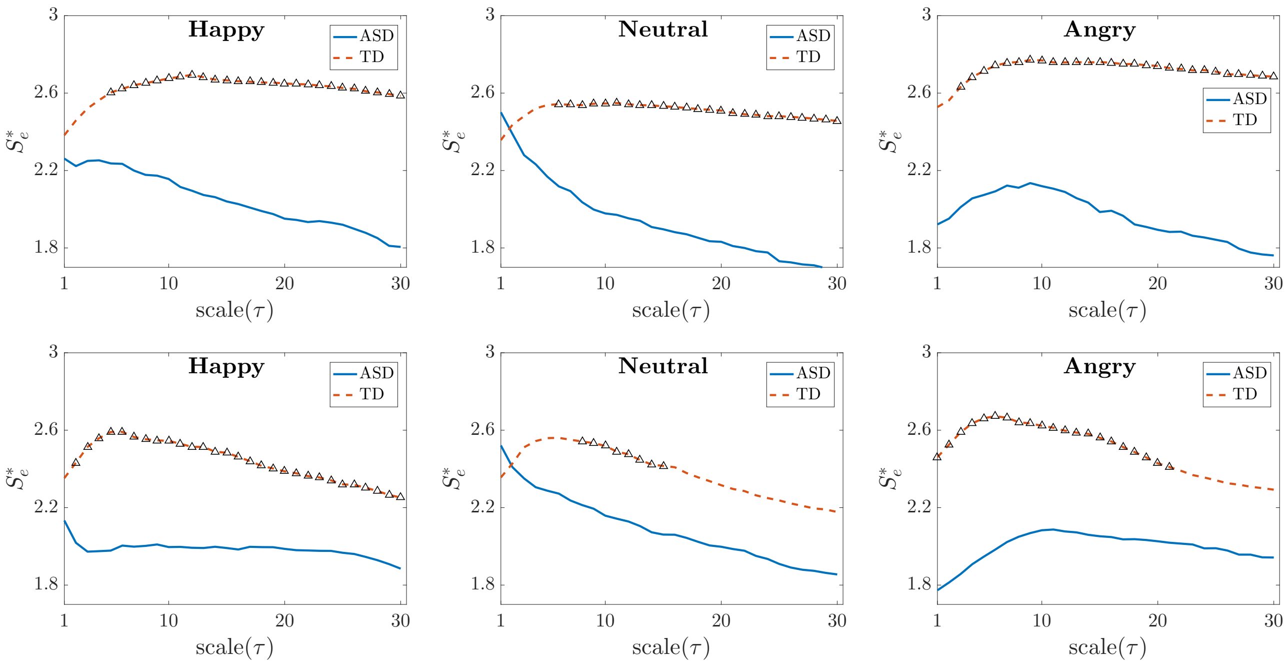

Fig. 3. Complexity in pupil response of ASD and TD groups pertaining to *face ROI* (top row) and *others ROI* (bottom row). The markers indicate the scales at which the group difference is statistically significant ($p \leq 0.05$).

Given a pupil response signal $\mathbf{P} = [p_1, p_2, ..., p_N]$, we construct a number of template vectors of dimension m as follows: $\mathbf{P}_i^m = [p_i, p_{i+1}, ..., p_{i+m-1}]$, where $i \in [1, N-m]$. Two template vectors $\mathbf{P}_i^m, \mathbf{P}_j^m$, $i \neq j$ are considered similar if $d(\mathbf{P}_i^m, \mathbf{P}_j^m) \leq r$, where $d(.)$ is the Chebyshev distance. Let the number of similar vector pairs in m-dimensional space be $\beta_r^{(m)}$. We repeat this exercise in $(m+1)$-dimensional space by constructing template vectors of dimension $(m+1)$ from $\mathbf{P}$. Let the number of similar vector pairs found in $(m+1)$-dimensional space be $\beta_r^{(m+1)}$. The sample entropy S_e is then computed as follows:

$$S_e(\mathbf{P}, m, r) = -\ln \frac{\beta_r^{(m+1)}}{\beta_r^{(m)}}. \tag{1}$$

The above equation can be interpreted as the conditional probability of two template vectors within the signal being similar in an $(m+1)$-dimensional space, given that they are similar in the m-dimensional space. To extend this idea to multiple scales, we need to construct multiple coarse grained signals $\mathbf{Q}_\tau$ from the original signal $\mathbf{P}$ corresponding to different values of τ denoting different scales as $\mathbf{Q}_\tau = [q_1, q_2, ..., q_{\frac{N}{\tau}}]$, where $q_j = \frac{1}{\tau}\sum_{i=1+(j-1)\tau}^{j\tau} p_i$. At a given scale τ, we compute the number of similar vector pairs $\beta_{k,\tau}^m$ and $\beta_{k,\tau}^{m+1}$) for all k, $1 \leq k \leq \tau$ for m and $(m+1)$-dimensional space. The RCMSE, denoted by S_e^*, is then defined as

$$S_e^*(\mathbf{P}, \tau, m, r) = -\ln \frac{\bar{\beta}_{k,\tau}^{(m+1)}}{\bar{\beta}_{k,\tau}^{(m)}}. \tag{2}$$

where $\bar{\beta}_{k,\tau}^{(m)} = \frac{1}{\tau}\sum_{k=1}^{\tau} \beta_{k,\tau}^{(m)}$ and $\bar{\beta}_{k,\tau}^{(m+1)} = \frac{1}{\tau}\sum_{k=1}^{\tau} \beta_{k,\tau}^{(m+1)}$. Our experiments use $m = 2$, $\tau = 1, 2, \cdots, 30$ and $r = 0.2 \times \sigma$, where σ is the standard deviation of $\mathbf{P}$. These values of r and m are chosen based on the observations from previous studies on physiological signal analysis [24].

We compute S_e^* for each pupil response signal for each subject at 30 different scales. At each scale, the complexity values are averaged across all subjects within a group for each affective expression displayed by the avatar. Results of this analysis corresponding to the two ROIs are shown in Fig. 3. The markers in Fig. 3 indicate the scales at which the group difference is statistically significant ($p \leq 0.05$ using two sample t-test). In general, if a physiological signal shows higher S_e^* at the majority of scales compared to another, it is considered to be more complex. From Fig. 3 we observe significantly reduced complexity in the ASD group for all three emotions at the majority of the scales except for the case neutral-*others ROI*. Also note that the group differences are more pronounced for the pupil dynamics in *face ROI* than the *others ROI*, and for the happy and angry expressions as compared to the neutral expression.

3.3. Synchrony between pupil response and fixation duration

So far, we have analyzed the pupil response as an independent quantity. Since fixation duration data was obtained simultaneously during the same interaction, it is critical to investigate if there is any meaningful relationship between the two quantities. For this purpose, we use *mutual information* and *dynamic time warping* (DTW) to study synchrony (i.e. mutual dependence and coordination) between the two quantities.

Mutual information: Mutual information is a measure of mutual dependence between two random variables. It is often used to measure the dependence between different physiological signals [25, 26]. Mutual information quantifies the drop in uncertainty of one variable when the other is known. Given two random quantities, X and Y, $\mathcal{I}(X, Y)$ is computed as

$$\mathcal{I}(X, Y) = \sum_{x \in X}\sum_{y \in Y} p_{XY}(x, y) \log_2 \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \tag{3}$$

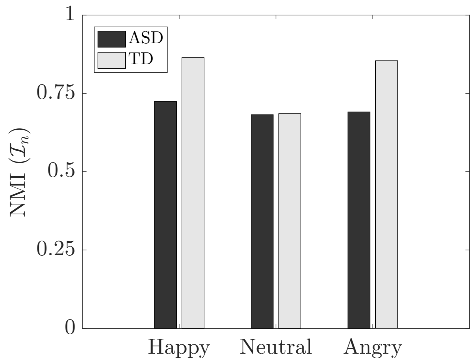

Fig. 4. Normalized mutual information (NMI) computed between pupil response and fixation duration for the ASD and TD groups.

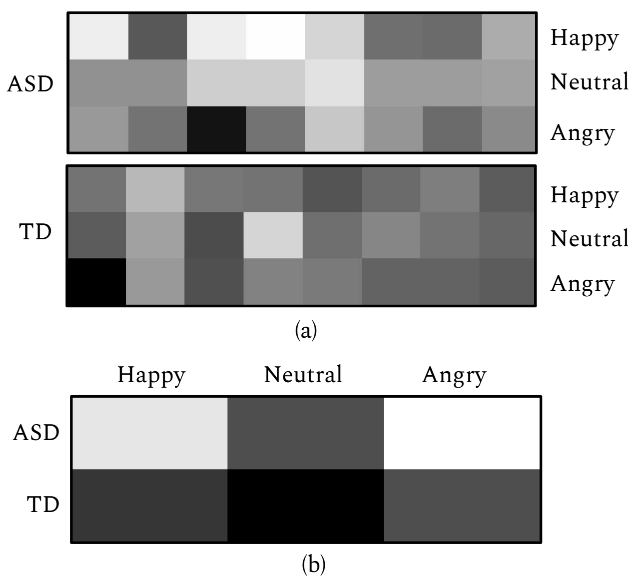

Fig. 5. (a) Warping distance for individual subjects for the three expressions displayed by avatars; (b) Warping distance averaged across all subjects. Brighter color indicates larger distance.

where p_X and p_Y are the marginal and p_{XY} is the joint probability density functions. Normalized mutual information (NMI) $\mathcal{I}_n(X,Y)$ is given by $\mathcal{I}_n(X,Y) = \mathcal{I}(X,Y)/\sqrt{H(X)H(Y)}$, where $H(.)$ is the entropy $H(X) = \sum_{x \in X} p_X(x) \log_2 p_X(x)$.

We compute NMI $\mathcal{I}_n(X,Y)$ between pupil response (X) and fixation duration (Y) for different context-relevant emotional expressions displayed by the avatar. From the results in Fig. 4, we observe that the ASD group has lower shared information between pupil response and fixation duration (for happy and angry) indicating lower coordination between them. A two sample t-test shows significant difference between the ASD and TD groups for happy ($p = 0.015$) and angry ($p = 0.005$) expressions. No group difference is observed when avatars display neutral expressions.

Dynamic time warping: To investigate the synchrony between the pupil response and fixation patterns, we employ the dynamical time warping (DTW) method. DTW directly compares two temporal sequences by finding the best alignment between them and computes a warping distance which can be interpreted as a measure of (dis)similarity or (lack of) synchrony.

Consider two time series of length N: $X \in \mathbb{R}^{N_1}$ and $Y \in \mathbb{R}^{N_2}$. DTW finds the best warping path by optimizing the distance between X and Y . We construct a distance matrix $D \in \mathbb{R}^{N_1 \times N_2}$, where the element $D_{i,j} = |\mathrm{x}_i - \mathrm{y}_j|_2$ measures the distance between x_i, the i^{th} point in X and the j^{th} point in Y i.e. y_j. A warping path $W = \{\mathrm{w}_1, \mathrm{w}_2, \cdots \mathrm{w}_K\}$ is a contiguous set of elements in D that defines a mapping between X and Y. The optimal warping path is the one that minimizes the warping distance $\sum_{k=1}^{K} D(\mathrm{w}_k)$, where w_k is the k^{th} element in W that defines mapping between x_i and x_j. $D(\mathrm{w}_k) = D_{i,j}$. The warping distance is determined by the optimal warping path $d_{DTW}(X,Y) = \min \sum_{k=1}^{K} D(\mathrm{w}_k)$. The optimal path can be found using dynamic programming with complexity $\mathcal{O}(N_1 N_2)$.

We compute the warping distance between pupil response and fixation duration sequences of each subject (see Fig. 5(a)) for each expression. At the group level, we observe that ASD has significantly larger warping distance (indicating less synchrony between fixation and pupil response) for happy ($p = 0.022$) and angry ($p = 0.034$) as compared to their TD peers (see Fig. 5(b)).

3.4. Discussion

The major observations from our experiments are as follows:

- ASD subjects fixate less on the faces of the avatars and more on the other regions, especially when the avatars display expressions of happiness and anger.
- The pupil response dynamics of ASD group shows lower complexity as compared to the TD group. Complexity of a signal can be interpreted as a measure of how fast new patterns are generated by the underlying system. In the present context, lower complexity may be understood as a reduced capacity to adapt to the dynamic stimuli by generating new patterns. This observation is consistent with the *loss of complexity hypothesis* [27, 28], which suggests that physiological systems exhibit lower complexity under disease or disorder.
- We observe lower synchrony between the pupil response and fixation pattern in ASD subjects than their TD peers. This indicates higher variability and ambiguity in the overall gaze behavior of the ASD subjects.
- In all experiments, we note that the group difference is more prominent as the avatars display happy and angry expressions. Clearly, the gaze behavior of the ASD subjects is effected by the communicator's facial expressions.

4. CONCLUSION

This paper presented a systematic, computational study of the gaze behavior of individuals with Autism as they interacted with a VR-based social communication system. We observed atypical gaze patterns in ASD subjects, and quantified this atypicality further in terms of various statistical and mathematical components. While our experiments indeed present new observations and perspectives on eye gaze behavior in Autism, since our observations are made based on analyzing a small number of subjects with autism, we cannot generalize the findings to the entire spectrum population.

5. ACKNOWLEDGEMENT

The authors are grateful to Prof. Uttama Lahiri, IIT Gandhinagar for sharing their data.

6. REFERENCES

[1] G Dawson, S J Webb, and J McPartland, "Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies," *Developmental Neuropsychology*, vol. 27(3), pp. 403–424, 2005.

[2] A Klin, W Jones, R Schultz, F Volkmar, and D Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Archives of General Psychiatry*, vol. 59(9), pp. 809–816, 2002.

[3] R B Grossman, R Edelson, and H Tager-Flusberg, "Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism," *J Speech, Language, and Hearing Research*, vol. 56(3), pp. 1035–1044, 2013.

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

[5] C Trepagnier, M M Sebrechts, and R Peterson, "Atypical face gaze in autism," *Cyberpsychology & Behavior*, vol. 5(3), pp. 213–217, 2002.

[6] K A Pelphrey, J P Morris, and G McCarthy, "Neural basis of eye gaze processing deficits in autism," *Brain*, vol. 128(5), pp. 1038–1048, 2005.

[7] E Bal, E Harden, D Lamb, Amy V Van H, J W Denver, and S W Porges, "Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state," *J Autism and Developmental Disorders*, vol. 40, no. 3, pp. 358–370, 2010.

[8] L L Speer, A E Cook, W M McMahon, and E Clark, "Face processing in children with autism: Effects of stimulus contents and type," *Autism*, vol. 11, no. 3, pp. 265–277, 2007.

[9] K M Dalton, B M Nacewicz, T Johnstone, H S Schaefer, M A Gernsbacher, H H Goldsmith, A L Alexander, and R J Davidson, "Gaze fixation and the neural circuitry of face processing in autism," *Nature neuroscience*, vol. 8, no. 4, pp. 519, 2005.

[10] K A Pelphrey, N J Sasson, J S Reznick, G Paul, B D Goldman, and J Piven, "Visual scanning of faces in autism," *Journal of autism and developmental disorders*, vol. 32, no. 4, pp. 249–261, 2002.

[11] M Bellani, L Fornasari, L Chittaro, and P Brambilla, "Virtual reality in autism: state of the art," *Epidemiology and Psychiatric Sciences*, vol. 20(3), pp. 235–238, 2011.

[12] S Kuriakose and U Lahiri, "Understanding the psycho-physiological implications of interaction with a virtual reality-based system in adolescents with autism: a feasibility study," *IEEE Trans Neural Systems and Rehab Engg*, vol. 23, no. 4, pp. 665–675, 2015.

[13] S Alves, A Marques, C Queirós, and journal=PsychNology Journal volume=11 number=3 year=2013 Orvalho, V, "Lifeisgame prototype: A serious game about emotions for children with autism spectrum disorders.," .

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

[15] P R KB, P Oza, and U Lahiri, "Gaze-sensitive virtual reality based social communication platform for individuals with autism," *IEEE Trans Affective Computing*, 2017.

[16] J S Richman and J R Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American J of Physiology-Heart and Circulatory Physiology*, vol. 278(6), pp. H2039–H2049, 2000.

[17] T Guha, Z Yang, R B Grossman, and S S Narayanan, "A computational study of expressive facial dynamics in children with autism," *IEEE Trans. Affective Computing*, vol. 9(1), pp. 14–20, 2016.

[18] F J Sansosti and K A Powell-Smith, "Using computer-presented social stories and video models to increase the social communication skills of children with high-functioning autism spectrum disorders," *J of Positive Behavior Interventions*, vol. 10(3), pp. 162–178, 2008.

[19] Hilary Coon, Michele E Villalobos, Reid J Robison, Nicola J Camp, Dale S Cannon, Kristina Allen-Brady, Judith S Miller, and William M McMahon, "Genome-wide linkage using the social responsiveness scale in utah autism pedigrees," *Molecular autism*, vol. 1, no. 1, pp. 8, 2010.

[20] Susie Chandler, Tony Charman, Gillian Baird, Emily Simonoff, TOM Loucas, David Meldrum, Mimi Scott, and Andrew Pickles, "Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 46, no. 10, pp. 1324–1332, 2007.

[21] A Catarino, O Churches, S Baron-Cohen, A Andrade, and H Ring, "Atypical eeg complexity in autism spectrum conditions: a multiscale entropy analysis," *Clinical Neurophysiology*, vol. 122(12), pp. 2375–2383, 2011.

[22] M Costa, A L Goldberger, and C-K Peng, "Multiscale entropy analysis of biological signals," *Physical review E*, vol. 71, no. 2, pp. 021906, 2005.

[23] S-D Wu, C-W Wu, S-G Lin, K-Y Lee, and C-K Peng, "Analysis of complex time series using refined composite multiscale entropy," *Physics Lett A*, vol. 378(20), pp. 1369–1374, 2014.

[24] D E Lake, Jo S Richman, M P Griffin, and J R Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789–R797, 2002.

[25] H Hinrichs, T Noesselt, and H-J Heinze, "Directed information flowa model free measure to analyze causal interactions in event related eeg-meg-experiments," *Human brain mapping*, vol. 29, no. 2, pp. 193–206, 2008.

[26] Di Hoyer, B Pompe, Ki H Chon, He Hardraht, C Wicher, and U Zwiener, "Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales," *IEEE Trans. Biomedical Engineering*, vol. 52, no. 4, pp. 584–592, 2005.

[27] L A Lipsitz, "Dynamics of stability: the physiologic basis of functional health and frailty," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 57, no. 3, pp. B115–B125, 2002.

[28] Le A Lipsitz and A L Goldberger, "Loss of'complexity'and aging: Potential applications of fractals and chaos theory to senescence," *Jama*, vol. 267, no. 13, pp. 1806–1809, 1992.