Analyzing Dynamic Components of Social Scene Parsing Strategy in Autism Spectrum Disorder

Chen Song[1], Aosen Wang[1], Kathy Doody[2], Michelle Hartley-McAndrew[3], Jana Mertz[3], Feng Lin[1], Wenyao Xu[1]

***Abstract*— Autism spectrum disorder (ASD) is a complex developmental disability characterized by deficits in social interaction. Gaze behavior is of great interest because it reveals the parsing strategy the participant uses to achieve social content. The legacy features in gaze fixation, such as time and area-of-interest, however, can not comprehensively reveal the way the participant cognizes the social scene. In this work, we investigate the dynamic components within the gaze behavior of children with ASD upon the carefully-selected social scene. A cohort 51 children (between 2 and 10 years) were recruited. The results show significant differences in the social scene parsing strategies of children with ASD, giving added insight into their ability to adequately decode and interpret the social scenarios.**

I. Introduction

Children with autism spectrum disorder (ASD) have impairments in social interaction, communication as well as atypical behaviors that include restricted interests and repetitive behaviors [1]. Impairments in the social domain include challenges in decoding and processing socially relevant information from faces and facial expressions. Studies have shown that children with ASD have different social parsing strategies and respond atypically when compared to same-aged peers with typical development, with respect to attention to faces, facial recognition and identification of different emotional expressions [2], [3], [4].

Eye-tracking technology has been explored to differentiate between the eye gaze patterns of children with ASD versus those with typical development. Advances in ability to track eye gaze have led to the growing popularity in assessing the eye gaze of children with ASD in performing various tasks. The corresponding hypothesis is that facial scanning strategies are abnormal in autism spectrum disorders [5].

Most of the existing work has made use of two major aspects of study: 1) exploring the fixation areas where gaze points are located and 2) calculating the corresponding fixation time that is spent in specific areas. Static images are frequently served as the visual stimuli. Some reports have suggested that individuals with ASD have poor face recognition and give decreased attention to internal features of faces such as eyes, nose and mouth compared to typical controls [6], [7], [8], [9]. Some others have stated that there is similarity in fixation pattern between individuals with ASD and with typical development [10]. Rutherford *et al.* found that the ASD group did not look less at the eyes overall, and did not look more at the mouth overall, when exploring how differently ASD participant processes the facial image [11]. Apart from that, dynamic images has also been employed. For example, Pierce *et al.* designed dynamic geometric images (DGI) and dynamic social images (DSI) [12]. By analyzing the distribution of the gaze points, they observed that individuals with ASD spent a greater amount of time examining DGI than DSI when compared to two other groups: participants with developmental delays and those with typical development.

One limitation of applying fixation patterns is that they cannot indicate how the brain actually combines all the visual information it receives and processes the visual scene it is observing. For example, even if an individual with ASD has a similar fixation pattern while gazing upon an image when compared to those without ASD, it does not necessarily mean that the individual has the correct and accurate perception of the image. Aside from the recognition of key objects in a visual scene, deficits in social interaction of the participants with ASD occur due to the lack of ability to understand the relationship between the key objects within the social scene. The corresponding **dynamic connection information** among the key objects is what the traditional fixation features such as time and area-of-interest (AOI) cannot reveal.

In this work, we demonstrate the salience of exploring the dynamic component in the gaze behavior of children with ASD. We hypothesized that children with ASD demonstrate noteworthy differences in eye gaze patterns when parsing social scenes compared to controls with typical development. We further hypothesized that the dynamic parsing of social situations can reveal the objective difference between children with ASD and with typical development.

II. Materials and Methods

A. Participants

This study was approved individually by the Institutional Review Boards at the Women and Childrens Hospital of Buffalo, University at Buffalo, and SUNY, Buffalo State. A total of 51 participants were enrolled in this study. Of the participants, 16 were female (31%) and 35 were male (69%). All children enrolled were between the ages of 2 and 10. Of the 51 children enrolled, 25 had a diagnosis of ASD (5 of the females and 20 of the males) and 26 had typical development (11 of the females and 15 of the males). In total, 8 of the children had received a diagnosis of ASD through assessment by administering the Autism Diagnostic Observation Schedule (ADOS) [13]; the remainder received a

[1]C.Song, A.Wang, F.Lin, W.Xu are with the Department of Computer Science and Engineering, SUNY at Buffalo
[2]K. Doody is with the Department of Exceptional Education, SUNY Buffalo State
[3]M.Hartley-McAndrew, J.Mertz are with Children's Guide Foudation Autism Spectrum Disorder Center

diagnosis via direct observation, parent report, and physician judgment closely following the criteria set forth in the Diagnostic Statistical Manual (DSM) for ASD.

All participants were recruited via existing research programs, a listserv maintained by two of the researchers, or a local parent group which met monthly. Many of the families who participated in the study were members of all three of these recruitment groups. Inclusionary criteria included falling between the ages of 2 and 10, with and without ASD. Informed parental consents were also required and were obtained at the time of the study by one of the authors. Children who had the ability to provide informed assent did so after listening to a description of the studys requirements by one of the authors. All children and parents who arrived at the study site opted to participate after receiving a thorough and comprehensive description of the study and its requirements. Children and families were reminded that their participation was voluntary. Participants were not provided any compensation for their involvement in the study.

B. Materials

1) Furniture: A child's size wooden, collapsible table with chairs was employed for use during the study. The table and chairs were adjustable to accommodate the varying heights of the participants.

2) Monitor: A Dell P2214H monitor ($476.64 * 267.78$ in mm/ $18.77 * 10.54$ in inch) was placed upon the table, and adjusted to be at eye level for each individual participant.

3) Reinforcers: Small items were available for the children upon completion of the study. Some children selected their preferred reinforcer prior to beginning the study, while others selected their chosen item at the conclusion. Reinforcers included items such as stickers, juice boxes, small snacks (granola bars or bags of chips), and sensory-friendly manipulatives, such as koosh balls.

4) Eye-tracking Device: A Tobii EyeX Controller [1] was used to measure the gaze behavior of participants in response to presented visual stimulus. The Tobii EyeX Controller is an eye tracking device which uses near-infrared light to track the eye movements of the participant. With the advanced eye tracking technology, Tobii EyeX Controller is able to record the X and Y coordinates of the participant's gaze point on the screen at the frequency of 120Hz. The recommended operating distance is from 45 to 75 cm (17.7 to 29.5). Figure 1(a) shows that a participant was sitting in front of the monitor and watching, while the Tobii EyeX Controller was tracking the eye gaze simultaneously.

5) Visual Stimulus: The visual stimulus used in this study was carefully prepared based on the following principles. First, there were no more than 5 key areas (such as faces and hands) included in order to comprehensively reveal the perception ability of the participant over the image. Second, the key areas of the photograph maintained a certain distance from each other to avoid the condition that the participant's eye gaze inadvertently lingered and coincidentally hit two or more key areas in a sequence. Third, the background needed to be visually clean to discourage the participant from being unintentionally distracted by colorful or irrelevant stimuli. Following the parameters above, Figure 1(b) demonstrates one such example of visual stimulus employed in our study. The blue circles are the gaze distribution from a participant with typical development (Some are filled in because the gaze points are dense in the areas).

(a) A participant is watching the visual stimulus. (b) An example of the gaze distribution on the visual stimulus.

Fig. 1. Demonstration of the experimental setting.

C. Procedure

A social narrative was first created with photographs of the study's new location, the vacant office suite where the study was conducted, and pictures of the furniture and computer equipment used. It is used to familiarize children with ASD to novel places or events.The social narrative also included photos of the researchers who would be conducting the study, along with pictures of the graduate assistants who were there in a supportive role. It was distributed to each participant via email prior to the study and hard copies were made available to parents attending the monthly meetings.

When the participants arrived at the study's location, the children were free to walk around the environment with one of the researchers while the graduate assistants completed the consent forms with the parents. At this time, assent was obtained from participants who were able to provide assent.

The participants were then ushered into a small room containing the table and computer equipment. Parents were invited to observe or were able to remain in the small waiting area directly adjacent to the study's location. One of the researchers accompanied the participants into the study area, while the graduate assistants calibrated the equipment.

18 visual stimuli in total were displayed to the child. Each one was shown for 5 seconds on the screen and the whole process took 90 seconds. One of the researchers, well-trained in behavior management and experienced in working with children with intensive behaviors, was next to the participant at all times, ready to intervene as necessary.

D. Dynamic Component Analysis

We explored the dynamic component within the gaze data to comprehensively evaluate the way the participant processed the visual stimuli.

1) Key-area Connecting Graph: We examined the sequential information found in a participant's gaze behavior. A Key-area Connecting Graph (KCG) was introduced before the specific data were analyzed. KCG is defined as a connected graph containing pre-defined key areas in regard to

[1] http://www.tobii.com/

(a) Key-area connecting graph. (b) Symbolic representation.

Fig. 2. Preprocessing upon visual stimuli.

the social scene within the photograph. As shown in Figure 1(b), there are three people engaged in a conversation; two of those three people are shaking hands with each other, while the third watches. The position of their bodies, coupled with the gesture of a handshake, indicates a social interaction. Therefore, we defined the corresponding key areas that can reflect the social information, and represented them by the blue circles (see Figure 2(a)). As the arrows depict, the inner connection of the key areas forms KCG.

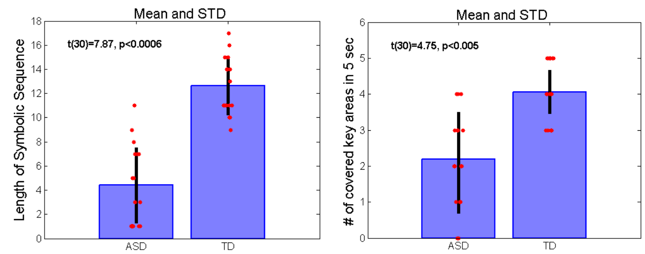
2) Symbolic Representation: This method was used to symbolize the gaze path with a symbol sequence for the convenience of eventual processing. The five key areas were labeled from A to E, as shown in Figure 2(b). Additionally, the background was defined as any area other than the key areas and was labeled as X. In practice, it is possible for an individual with ASD to scan the visual stimulus without purpose. Therefore, it is important to only extract the meaningful gaze point in the whole gaze sequence. Fixation threshold is the minimum fixation time of a gaze point that we applied to determine whether a gaze point is a real fixation or a saccade (a rapid eye movement that occurs when eyes fixate on one point after another on the visual stimulus).

3) Dynamic Component: We proposed four unique components from the participant's gaze pattern in the dynamic domain. The components are highly related to the way the gaze behavior performs on the KCG. Additionally, they will supplement our understanding of the participant's perception of the visual stimulus. The followings are the four components and detailed descriptions:

- **Length of the symbolic sequence:** This component infers the way the participant processes the social scene. Within the limited display time, the length of the symbolic sequence should not be too long or too short. An extremely long length of symbolic sequence is most likely caused by the repetitive local gaze between certain areas. An extremely short length can be the result of long-time staring at specific area.
- **The number of covered key areas within 5 seconds:** This component is highly related to the ability of the participant to grasp the global content on the scene. Normally, participants without ASD are expected to quickly gaze upon all the key areas regardless of the order of presentation on the image.
- **Number of the key-area pair:** The cognitive ability lays in the understanding of relationship between the key areas. Individuals with typical development will tend to consecutively look at two areas if they are related. Therefore, the total amount of times the participant switched fixation between two key areas was counted for each image. Participants without ASD are expected to have more fixation switches between the key areas than participants with ASD.
- **Effective gaze coefficient:** This coefficient is denoted as the effective gaze behavior in the gaze sequence. It is calculated as the percentage of key-area pairs in the total gaze sequence, inferring how many gaze fixations are involved to scan one key-area pair. If the participant frequently watches the areas other than the key areas (such as the background), the number of key-area pairs will be relatively small even if the whole symbolic sequence is long.

III. Results and Discussion

Four unique components from the participant's gaze pattern in the dynamic domain were proposed and analyzed.



(a) Length of the symbolic sequence. (b) # of covered key areas within 5s.

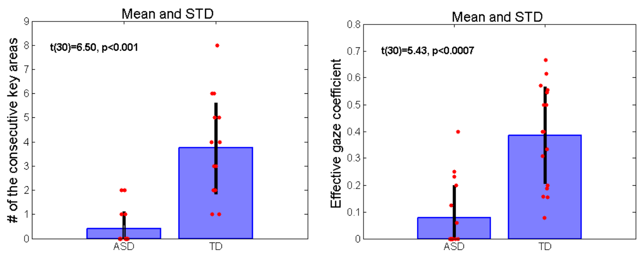
Fig. 3. Dynamic component analysis.

A. Length of the symbolic sequence

As shown in Figure 3(a), the length of the symbolic sequence was much shorter in participants with ASD than in the participants with typical development ($t(30) = 7.87$, $P < 0.0006$). Specifically, the participants with ASD had an average length of 4.44 ($std. = 3.38$), while the participants with typical development have a significantly larger average length of 12.63 ($std. = 2.42$). This result indicates that children with typical development moved their eye gaze a discrete distance between areas of interest and back again, presumably in order to sufficiently perceive the context of the social situation. Meanwhile, the group of children with ASD had a shorter amount of distance seen moving their eye gaze about the image, inferring that the participants with ASD tend to keep focusing on a limited local part of the visual stimulus and ignore the global information.

B. # of covered key areas within 5s

We analyzed the number of covered key areas within 5 seconds. Our hypothesis was that the participants with typical development can scan most of the key areas (such as the individuals face and shaking hands) of visual stimulus in 5 seconds. As shown in Figure 3(b), we observed the corresponding difference between the participants with and without ASD ($t(30) = 4.75$, $P < 0.005$). All the participants with typical development gazed more than 3 key areas. In contrast, only half the participants with ASD were able to do so. The typical perception strategy the participants with typical development applied on the visual stimulus was to quickly browse the key areas in order to grasp general information regarding the stimulus, before focusing on the specific interesting areas. However, the participants with ASD lacked the global strategy over visual stimulus and were more likely to focus on local areas.



(a) # of the consecutive key-area pair. (b) Effective gaze coefficient.

Fig. 4. Dynamic component analysis.

C. The number of the consecutive key-area pair

Perception is essentially the ability to understand the relationship between the key areas in the scene. For example, eye gaze from a handshake to a face, might indicate an understanding of the meaning, while the gaze from a handshake to the sky in the image, might show that the meaning of that handshake is lost. We investigated the number of consecutive key-area pairs in the gaze sequence and hypothesized that the participants with ASD would notice far fewer key-area connection relationships compared to the participants with typical development. As shown in Figure 4(a), the significant difference has been revealed in this aspect ($t(30) = 6.5$, $p < 0.001$). By analyzing the number of consecutive key-area pairs, we found that the way participants with ASD processed the visual stimulus was much more detached than that of the participants with typical development. The participants with ASD tended to discretely perceive the area instead of noticing the relationship between the key areas, which eventually causes social communication deficit.

D. Effective gaze coefficient

As described above, the number of key-area pair provides information regarding how the participant perceived the visual stimulus. However, if the participant keeps switching gaze between two key areas, it will also result in a large number of key-area pairs. Therefore, to objectively analyze the gaze dynamic pattern, we introduce the effective gaze coefficient, which represents the percentage of key-area pairs in the total gaze sequence. As shown in Figure 4(b), the coefficient differs between two groups ($t(30) = 5.43$, $P < 0.0007$). On average, the participants with typical development used 38.6% ($std. = 18.8\%$) of the total time in processing the connecting relationship between the key areas. However, this percentage dramatically dropped to 7.9% ($std. = 12.5\%$) in the participants with ASD, inferring that the perception strategy used by individuals with ASD was not area-connection based.

IV. Conclusion

We explored a novel approach to determining differences in the eye gaze patterns of individuals with ASD. To our knowledge, ours is the first study to look at the dynamic processing and movement of eye gaze in a social scenario. By using a social scene of interacting subjects and comparing dynamic components of analysis, this study gives added insight on the individual with ASD's ability to adequately decode and interpret a given social scenario. This study indicates a significant difference in the way individuals with ASD process a given social situation through measuring dynamic gaze patterns. We propose that this observation is an important insight into the way ASD may affect social perception and may give added benefit to future diagnostic and therapeutic modalities.

References

[1] American Psychiatric Association, "Diagnostic and statistical manual of mental disorders," *Arlington: American Psychiatric Publishing*, 2013.

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

[3] S. J. Grice, H. Halit, T. Farroni, S. Baron-Cohen, P. Bolton, and M. H. Johnson, "Neural correlates of eye-gaze detection in young children with autism," *Cortex*, vol. 41, no. 3, pp. 342–353, 2005.

[4] C. Ashwin, S. Wheelwright, and S. Baron-Cohen, "Finding a face in the crowd: Testing the anger superiority effect in asperger syndrome," *Brain and Cognition*, vol. 61, no. 1, pp. 78–95, 2006.

[5] K. Chawarska and F. Shic, "Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 39, no. 12, pp. 1663–1672, 2009.

[6] 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.

[7] Z. L. Boraston, B. Corden, L. K. Miles, D. H. Skuse, and S.-J. Blakemore, "Brief report: Perception of genuine and posed smiles by individuals with autism," *Journal of Autism and Developmental Disorders*, vol. 38, no. 3, pp. 574–580, 2008.

[8] C. Katarzyna, V. Fred, and K. Ami, "Limited attentional bias for faces in toddlers with autism spectrum disorders," *Archives of general psychiatry*, vol. 67, no. 2, pp. 178–185, 2010.

[9] J. C. McPartland, S. J. Webb, B. Keehn, and G. Dawson, "Patterns of visual attention to faces and objects in autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 41, no. 2, pp. 148–157, 2011.

[10] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Defining and quantifying the social phenotype in autism," *American Journal of Psychiatry*, 2002.

[11] M. Rutherford and A. M. Towns, "Scan path differences and similarities during emotion perception in those with and without autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 38, no. 7, pp. 1371–1381, 2008.

[12] K. Pierce, D. Conant, R. Hazin, R. Stoner, and J. Desmond, "Preference for geometric patterns early in life as a risk factor for autism," *Archives of general psychiatry*, vol. 68, no. 1, pp. 101–109, 2011.

[13] K. Gotham, A. Pickles, and C. Lord, "Standardizing ados scores for a measure of severity in autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 39, no. 5, pp. 693–705, 2009.