



Design of an Assistive Avatar in Improving Eye Gaze Perception in Children with ASD During Virtual Interaction

Ashwaq Zaini Amat^{1(✉)}, Amy Swanson^{3,4}, Amy Weitlauf^{3,4}, Zachary Warren^{3,4},
and Nilanjan Sarkar^{1,2}

¹ Electrical Engineering and Computer Science,
Vanderbilt University, Nashville, TN 37212, USA

ashwaq.zaini.amat.haji.anwar@vanderbilt.edu

² Mechanical Engineering, Vanderbilt University, Nashville, TN 37212, USA

³ Treatment and Research Institute for Autism Spectrum Disorders (TRIAD),
Vanderbilt University, Nashville, TN 37212, USA

⁴ Pediatrics, Psychiatry and Special Education, Vanderbilt University, Nashville, TN 37212, USA

Abstract. Children diagnosed with autism spectrum disorder (ASD) usually experience impairment in social interaction and often display reduced gaze sharing when interacting with another person. The lack of gaze sharing or joint attention early in the children's developmental stage may create a delay in their ability to learn new things and share information with others. The presented study involved the design of a novel virtual reality (VR)-based training game with an avatar and eye tracker aimed to eventually address the joint attention impairment in children with ASD. The assistive avatar provides necessary cues and hints based on both the eye tracking data recorded by the VR system and the task performance of the participant. The system uses the task performance to adaptively change the difficulty level of the game. We believe that the training game will be able to improve participant's gaze following skill. A usability study was carried out to validate the system design. The result showed that the system was feasible and able to obtain the expected gaze performance from the participants. The details of the system architecture and result of system validation are presented in this paper.

Keywords: Virtual reality · Eye gaze · Joint attention · Autism spectrum disorder
Children with ASD

1 Introduction

Autism spectrum disorder (ASD) is a range of disorder that affects social communication and interaction. As of 2014, 1 in every 68 children in the USA are diagnosed with ASD [1]. Individuals within the ASD spectrum usually have atypical gaze patterns [2] and often display reduced gaze when interacting with another person [3]. Research shows that individuals with ASD spend less time looking at facial features especially the eye region compared to other non-facial areas. This atypical gaze behavior contributes to

impairment in language development, facial expression processing and sharing of information during social interaction [4].

The human gaze plays an important role as a mechanism for information sharing in addition to emotional cues. This form of information sharing is also known as joint attention, which can be defined as one's ability to coordinate one's attention with another person [5], which is a fundamental social skill that is impaired in many children with ASD [6–8].

Children with ASD, in general, show more affinity towards computer and machine interaction than human interaction [9]. In particular, advancement in virtual reality technology has contributed towards significant boost in the use of virtual systems for children with ASD. Virtual systems have the added advantage that they can record quantitative measures and track performance in real-time. As a result, several important virtual systems have been explored in the context of social games [10–12]. While there have been several studies on robot-assisted joint attention [13], studies on joint attention for individual with ASD using virtual systems are limited. Caruana et al. introduced an interactive VR-based joint attention social task that is focused on adult with ASD and found that joint attention difficulties are still present even in adult with ASD [14]. Courgen et al. conducted studies with adult and adolescent with ASD on gaze awareness and a pilot study on joint attention with adult with ASD participants only [15]. Researches have shown that early intervention of joint attention in children with ASD can significantly improve the children's ability to develop their communication skills [16–18].

The current study presents the design and development of a novel VR-based gaze training paradigm with an avatar. The aim of the design is to address the joint attention impairment and reduced eye contact in many children with ASD through the gaze-based interaction paradigm with the avatar. In this paradigm, a children with ASD participant and the avatar play a puzzle game in VR where the participant will be required to look at the avatar's eye region in order for the avatar to cue which puzzle piece the participant has to move. The system is capable of providing: (1) different gaze configurations for the avatar, (2) real-time computation of game performance, (3) game hints by the avatar when the participant is unable to get the correct piece and 4) an adaptive difficulty level adjustment based on the participant's performance. We present this new framework with system architecture, initial system validation and conclusion together with future works in this paper.

2 System Design

The system provides an environment that aims to improve gaze sharing and gaze perception in children with ASD. The virtual system is set up with an avatar at the center of the screen, seven pieces of tangram puzzle is spread around the avatar and a target image placed in front of the avatar at the lower part of the screen as shown in Fig. 1. An eye tracker is used to track the participant's gaze position on the screen and as an input device that interacts with the avatar and puzzle pieces in the virtual system. Participant is required to share their gaze with the avatar in order to know

which puzzle piece to move to complete the target image. The participant uses the mouse to move the puzzle pieces.

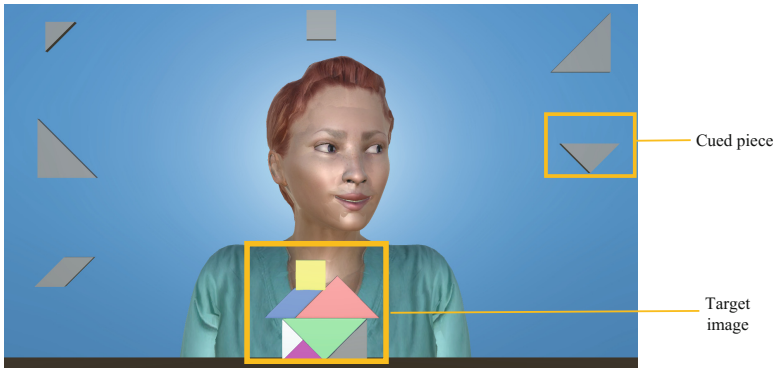


Fig. 1. The view of the training game. Avatar is at the center of the screen with tangram puzzle pieces spread around it. In this figure, the avatar is currently cueing the piece on the right. The target image is placed at the bottom in front of the avatar.

The system calculates the participant's performance based on the eye gaze inputs and response time. There are three different gaze configurations in the avatar that is used in the system. The different configuration is used to train gaze perception of the participant. An adaptive change and assistive hints are introduced in the system to optimize the system further. Figure 2 illustrates the block diagram of the eye tracking training system. Details of each of these configuration and game components are discussed in the subsequent sub-sections.

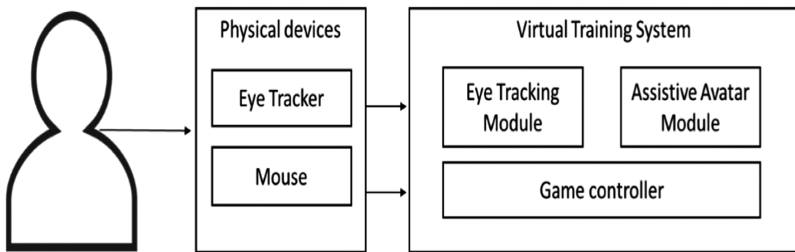


Fig. 2. Block diagram of the eye tracking training system

At the beginning of the training game, all the colors of the puzzle pieces were removed (zero color saturation). The avatar waits for the participant to make eye contact before it cues the piece to move. Participant will then look for the correct piece and when a correct piece is selected (through participant's gaze), the color of the puzzle piece is revealed and the participant is allowed to move the piece using the mouse to the corresponding slot on the target image. The interaction with the avatar is repeated for the remaining six puzzle pieces to complete the target image. The adaptive difficulty level

is applied at the beginning of every game. When a participant is not making eye contact or unable to select the correct piece in time, the avatar will provide necessary hints. Participants are provided with three attempts to select the correct piece before the avatar moves the piece when the participant failed all attempts.

2.1 Physical Inputs: Eye Tracker and Computer Mouse

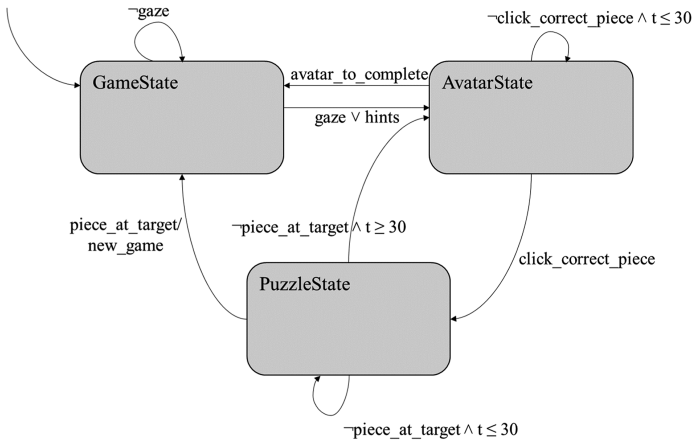
The eye tracker collects participant's eye gaze position on the screen during the game. A Tobii EyeX [19] is used in this study. The eye tracker is very lightweight and portable to use and can be easily attached to the lower edge of the monitor. The operating frequency is 50 Hz which is quite low, but since we are interested in at fixation data points rather than saccadic and fast-moving gaze, this sampling frequency is acceptable [20]. Tobii EyeX uses a USB 3.0 cable for data transfer with a rate of 20 MB/s. Other specifications of the eye tracker include an operating distance between 50–90 cm and a maximum monitor size of 27 in. The mouse input is used when selecting and moving puzzle pieces. A typical USB connected mouse is used in this experiment.

2.2 Virtual Training System

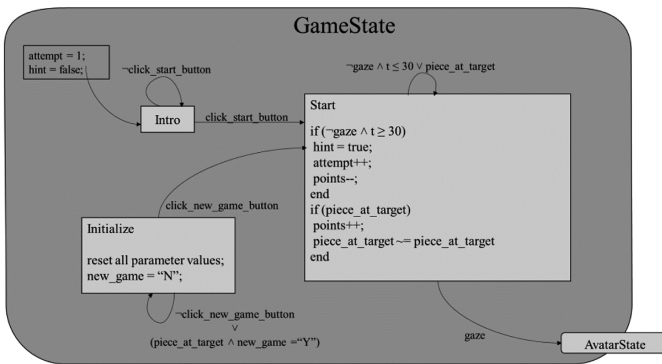
The virtual system is a platform for the participant to interact with the assistive avatar throughout the game. The system was developed using Unity v5.6.1f1 [21] and the games were modeled as finite state machines as shown in Fig. 3. The finite state machine provides a clean and organized way of tracking the state of the game. In each state, the corresponding adaptive response can be provided based on the performance level.

Eye Tracking Module. The eye tracking module is the interface between the physical world and the virtual world. A set of regions of interest (ROIs) were defined in the virtual system that included the avatar's facial features such as eyes, mouth, nose, forehead, and ears, and each of the puzzle pieces. The module uses a Tobii-Unity library [22] that has a gaze point API to collect gaze position from the eye tracker and another API to inform the system whenever participant's gaze is on any of the ROIs. The information collected from both APIs are used in the assistive avatar module to progress to the next sequence of the game or provide the necessary feedback to the participant.

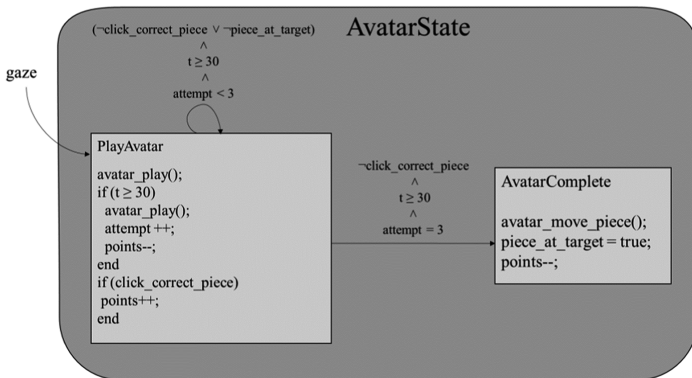
Assistive Avatar Module. The assistive avatar module consists of the avatar, controls of the animation and other configurations related to the avatar. The avatar and its different animations were created using Autodesk Maya [23]. Seven different gaze directions were created for the avatar for the different positions of the puzzle pieces on the screen. For each of the gaze directions, three different gaze configurations were created for the avatar to create different level of gaze perception in the cues: (1) head movement together with eye movement (HE), (2) only eye movement (E), and (3) minimal eye movement (ME). There were a total of 22 animations (21 moving animations and 1 static animation) stored in the system. Figure 4 shows the difference in the gaze configurations between HE and E. These different configurations were implemented to increase the difficulty



(a)



(b)



(c)

Fig. 3. (a) State machine of the training system (b) state machine for the puzzle pieces, PuzzleState (c) state machine for the virtual avatar, AvatarState

level of the gaze perception, where the region of gaze cue reduces from the whole head movement to very minimal eye movement.



Fig. 4. Comparison of avatar cue configurations. Both avatars are cueing the piece at the top left corner. The first image has the head and eye (HE) movement while the bottom image has eye only (E) movement.

Game Controller. Each puzzle pieces and the target image is configured through the game controller. The ROIs of the puzzle pieces from the eye tracking module are used as part of the logical sequence in the controller. The controller uses the input to enable or disable the movement of the puzzle pieces and also the color display settings of the pieces. The target location and angle of the puzzle pieces are determined by the controller using information of the target image in each game. Other game configuration parameters such as number of games, game progression and calculation of points are in the game controller.

Real-Time Game Points Calculation. Game controller calculates the game points at four different checkpoints in a single move as shown in Fig. 5. Points are gained when: (1) Participant makes eye contact with avatar; (2) Participant chooses the correct piece that was cued; (3) Participant moves the piece to the target; and (4) The piece is at the target within the response time. Based on this settings, the maximum points achievable is 4 per puzzle piece move, resulting in 28 points per game (as there are 7 pieces per game).

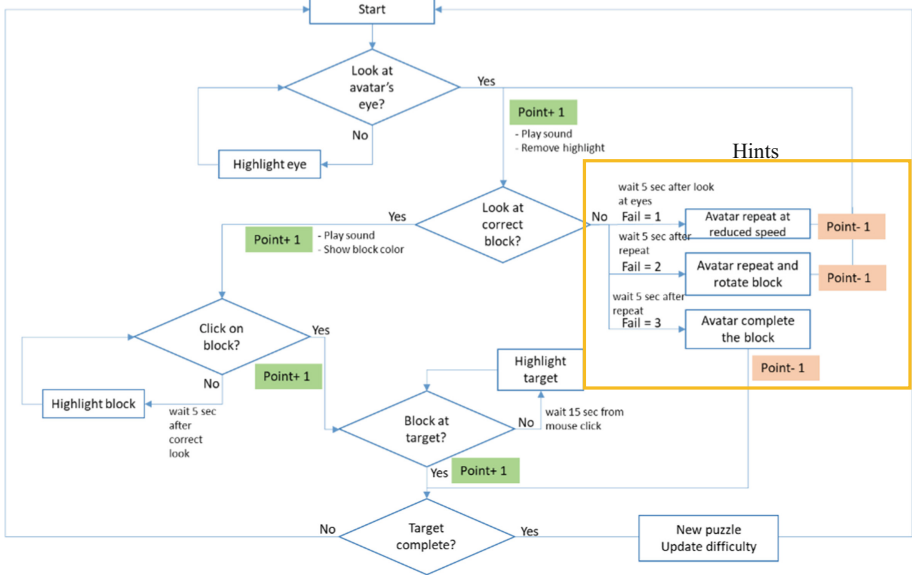


Fig. 5. Flow chart of game protocol

Assistive Hints. Figure 5 also shows the algorithm designed for point reduction and assistive hints provided by the avatar for the participant. Hints are system prompts that

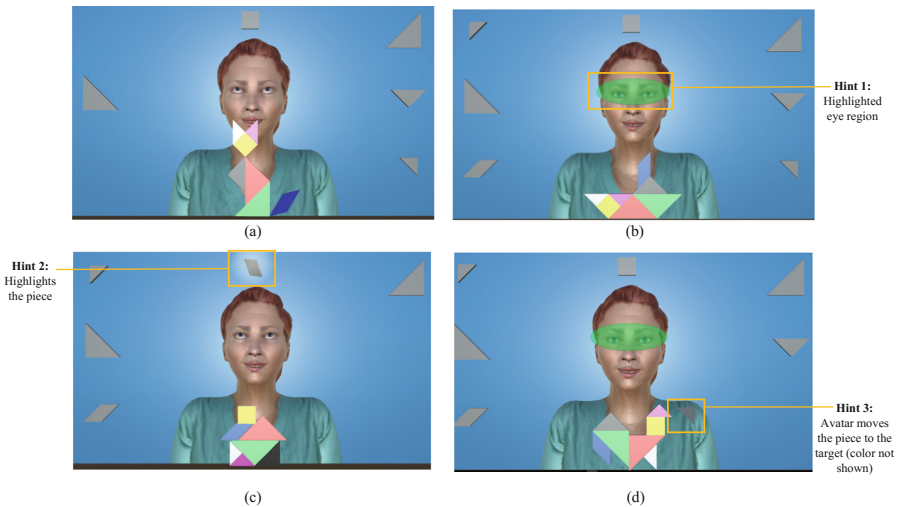


Fig. 6. (a) Avatar cues the piece at the top of the screen. (b) Highlighted eye region as a hint to make eye contact. (c) Avatar cues the piece at the top of the screen and a hint that highlights the piece it is looking at. (d) Avatar moves the piece to the target on its own when participant failed to select and move the correct piece three times.

are performed when participant fails to do certain tasks. In this training game, on the first failed attempt, the avatar will repeat the same gaze direction cue with its eye region highlighted as a hint. On the second failed attempt, the hint highlights and rotates the correct piece. On the last attempt, if the participant is still unable to choose the correct piece, the avatar will move the piece to the target on its own. Points are reduced in each failed attempt. Figure 6 shows an example of a hint where a piece is highlighted with a spotlight while the avatar cues it.

Adaptive Difficulty Level. The algorithm for adaptive difficulty level is designed by integrating the components of the system and evaluating the overall task performance of the participant. The adaptive changes in the avatar's speed and the participant's time to respond are shown in Table 1. When a participant earns between 7 to 14 points per game, it is considered as low performance category and no changes are made to the avatar's speed and time to respond in the consecutive game. For points earned between 15 to 21 points, medium performance category, only the avatar's speed is increased and time to respond remain the same in the next game. For highest points ranging between 22 and 28 points, the avatar's speed is increased and time to respond is reduced. High performers are challenged to respond in shorter time and at higher avatar's speed.

Table 1. Adaptive difficulty level matrix

Points earn in each puzzle game	Performance category	Avatar's speed	Time to respond
7–14	Low	Remain the same	Remain the same
15–21	Medium	Increase speed	Remain the same
22–28	High	Increase speed	Reduce time

By configuring the avatar to wait for the participant to make eye contact before it cues any puzzle piece, it is hoped that it will encourage the participant to make more frequent eye contact to progress in the game. Overall, the data collected from the system can be used to provide a comprehensive view of the participant's training performance.

3 System Validation

Usability study and system validation were conducted with the training. The system was tested for its feasibility, validity and reliability of the data collected and reliability of the algorithms. Three typically developing (TD) volunteers were recruited and tested the system.

The volunteers provided positive feedback after completing the test. They commonly agreed that: (1) the objective of the game was easy to understand; (2) all the gaze directions from the different gaze configurations correspond correctly to the location of the puzzle pieces; and (3) the eye tracker was responsive and sensitive to the gaze direction even when volunteers' head moved around a little bit.

To calculate the validity and accuracy of the eye tracker, we used eye gaze data and the ROIs that were defined in the training game. We selected three ROIs, highlighted in

Fig. 7, to be analyzed with gaze data from all three volunteers. The position of the ROIs on the screen were known, and data collected by the eye tracker includes gaze location on the screen and also identify if the eyes are on the ROIs. Based on these information, we calculated the distance between the gaze position to the actual position of the ROIs. We found that the accuracy were 0.88 cm in the y-direction (0.95° angle deviation in vertical direction) and 1.23 cm in the x-direction (1.33° angle deviation in the horizontal direction). These angles are acceptable for the application in this training game since the pieces and their ROIs are arranged far apart from each other.

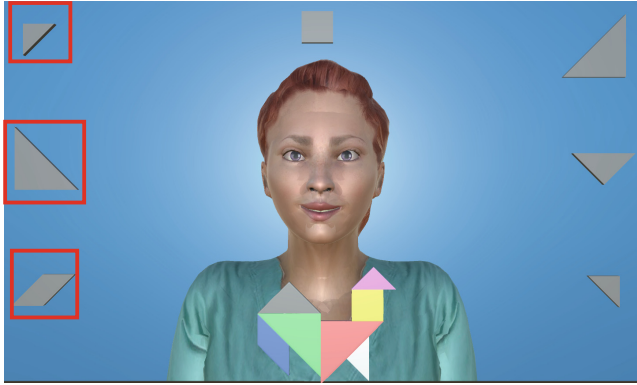
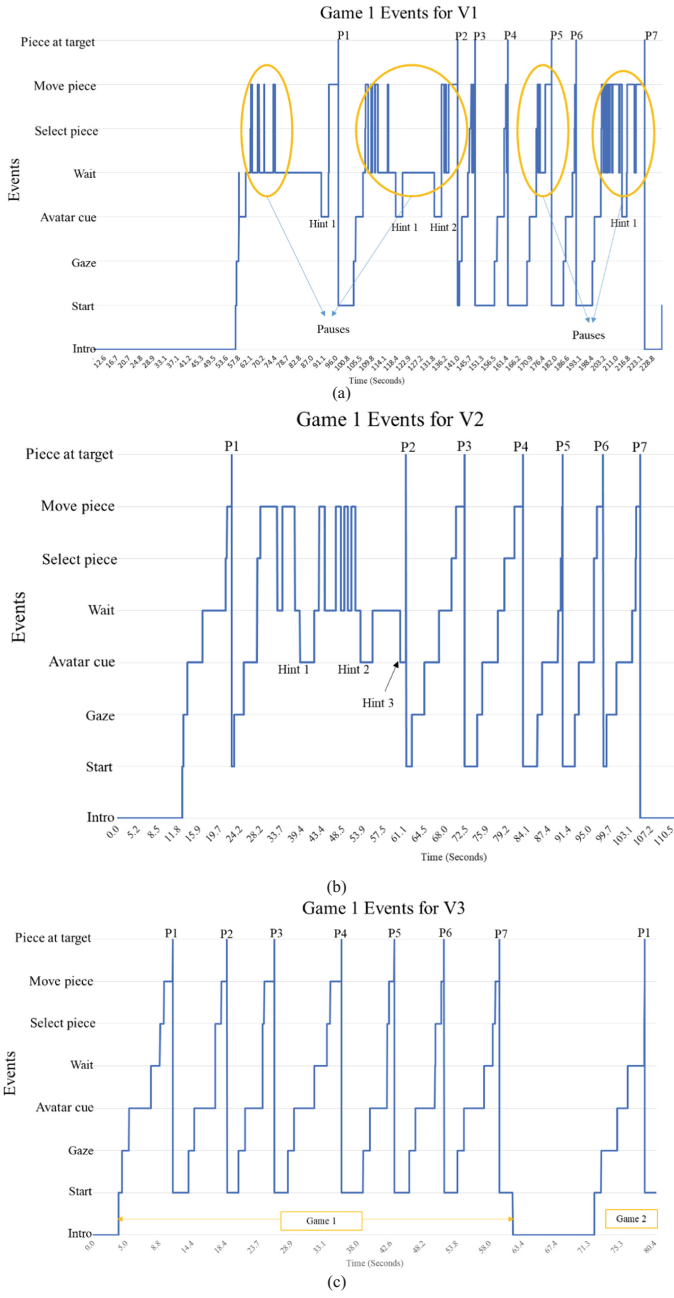


Fig. 7. The ROIs marked in the red boxes used for the accuracy calculation. (Color figure online)

In order to validate the algorithm used in the training game, a graph of events against game time progression was created for the volunteers for a single game. Based on the graphs in Fig. 8, each graph was unique and represents the performance of the volunteers. In Fig. 8(a), data from volunteer 1 (V1) showed multiple pauses in between movement of pieces to the target due to the volunteer having motor skill issues with using the mouse and had to take short breaks for each piece. But as the game progressed, the volunteer's performance improved but was not consistent as can be seen in the time interval for piece 3 which is shortest but then increased again for piece 4. The total time to complete one game for V1 was 226 s. As for volunteer 2 (V2), as can be seen in Fig. 8(b), it shows that the volunteer was able to progress quite well on average, except for the second puzzle piece where the avatar had to provide all three hints to the volunteer and completed the piece for the volunteer. The rest of the game progressed well with a smooth progression and shorter interval. The total time to complete one game for V2 was 106 s. For volunteer 3 (V3) as shown in Fig. 8(c), the graph showed a steady progression from the first piece to the seventh piece and also the first piece of the second game. The volunteer did not use any hints and was able to proceed with each trial successfully. The total time to complete one game for V3 was 59 s. These data show that the algorithm designed for the training game is working and that data analysis for this training game can show the level of performance for each participant and provide a good comparison of performance progress for the participant over the period of training sessions.



4 Conclusion and Future Work

Autism spectrum disorder (ASD) affects the social and communication skills of 1 in 68 children in the USA. Therapies and intervention sessions are known to be financially costly and time-consuming. Advancement in virtual reality and human-computer technology have provided a platform to further explore the application of virtual systems in the intervention and training for children with ASD.

One area of interest for intervention for children with ASD is in gaze sharing and joint attention. Children with ASD are known to have reduced gaze during social interaction and have poor joint attention skills. The training game discussed in this paper introduced a novel virtual reality-based game designed for joint attention training for children with ASD. The design and architecture were explained in detail. Results from system validation showed that the system is feasible, reliable and can provide the sufficient view of the performance of each participant.

Future work will include running clinical experiment with children with ASD and typically developing (TD) children and compare the difference in gaze pattern and visual information processing during such social interaction.

Acknowledgment. We are grateful for the support provided by NIH grants 1R21MH11548-01 for this research. The authors are solely responsible for the contents and opinions expressed in this manuscript.

References

1. Centers for Disease Control (CDC): Facts About ASD (2014)
2. Van Der Geest, J.N., Kemner, C., Verbaten, M.N., Van Engeland, H.: Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. *J. Child Psychol. Psychiatry* **43**, 669–678 (2002). <https://doi.org/10.1111/1469-7610.00055>
3. Pelphrey, K.A., Sasson, N.J., Reznick, J.S., et al.: *J. Autism Dev. Disord.* **32**, 249 (2002). <https://doi.org/10.1023/A:1016374617369>
4. Dawson, G., et al.: Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Dev. Psychol.* **40**(2), 271 (2004)
5. Mundy, P., Newell, L.: Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* **16**(5), 269–274 (2007). <https://doi.org/10.1111/j.1467-8721.2007.00518.x>
6. Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., Drew, A.: Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. *Dev. Psychol.* **33**(5), 781–789 (1997). <https://doi.org/10.1037/0012-1649.33.5.781>
7. Mundy, P., Sigman, M., Kasari, C.: A longitudinal study of joint attention and language development in autistic children. *J. Autism Dev. Disord.* **20**(1), 115–128 (1990)
8. Dawson, G., et al.: Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Dev. Psychol.* **40**(2), 271 (1990)
9. Moore, M., Calvert, S.: *J. Autism Dev. Disord.* **30**, 359 (2000). <https://doi.org/10.1023/A:1005535602064>

10. Bian, D., Wade, J.W., Zhang, L., Bekele, E., Swanson, A., Crittendon, J.A., Sarkar, M., Warren, Z., Sarkar, N.: A novel virtual reality driving environment for autism intervention. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013 Part II. LNCS, vol. 8010, pp. 474–483. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39191-0_52
11. Zhang, L., et al.: Design of a mobile collaborative virtual environment for autism intervention. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2016 Part III. LNCS, vol. 9739, pp. 265–275. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40238-3_26
12. Bekele, E., et al.: Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *IEEE Trans. Vis. Comput. Graph.* **19**(4), 711–720 (2013)
13. Zheng, Z., Zhao, H., Swanson, A.R., Weitlauf, A.S., Warren, Z.E., Sarkar, N.: Design, development, and evaluation of a noninvasive autonomous robot-mediated joint attention intervention system for young children with ASD. *IEEE Trans. Hum.-Mach. Syst.* **48**, 125–135 (2017)
14. Caruana, N., et al.: Joint attention difficulties in autistic adults: an interactive eye-tracking study. *Autism* (2017). <https://doi.org/10.1177/1362361316676204>
15. Courgeon, M., et al.: Joint attention simulation using eye-tracking and virtual humans. *IEEE Trans. Affect. Comput.* **5**(3), 238–250 (2014)
16. Mundy, P., Crowson, M.: Joint attention and early social communication: implications for research on intervention with autism. *J. Autism Dev. Dis.* **27**, 653 (1997)
17. Kasari, C., Freeman, S., Paparella, T.: Joint attention and symbolic play in young children with autism: a randomized controlled intervention study. *J. Child Psychol. Psychiatry* **47**, 611–620 (2006)
18. Zheng, Z., Nie, G., Swanson, A., Weitlauf, A., Warren, Z., Sarkar, N.: Longitudinal impact of autonomous robot-mediated joint attention intervention for young children with ASD. In: Agah, A., Cabibihan, J.-J., Howard, A.M., Salichs, M.A., He, H. (eds.) ICSR 2016. LNCS (LNAI), vol. 9979, pp. 581–590. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47437-3_57
19. Tobii Eyetracking EyeX. <https://help.tobii.com/hc/en-us/categories/201185405-EyeX>
20. Gibaldi, A., Vanegas, M., Bex, P.J., Maiello, G.: Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. *Behav. Res. Methods* **49**(3), 923–946 (2017)
21. Unity. <https://unity3d.com/unity>
22. Tobii Unity SDK. <http://developer.tobii.com/tobii-unity-sdk/>
23. Maya, 3D animation software. <https://www.autodesk.com/education/free-software/maya>