Design of a Computer-Assisted System for Teaching Attentional Skills to Toddlers with ASD

Zhi Zheng[1(✉)], Qiang Fu[1], Huan Zhao[1], Amy Swanson[2], Amy Weitlauf[2], Zachary Warren[2], and Nilanjan Sarkar[3,1]

[1] Electrical Engineering and Computer Science Department, Nashville, TN, USA
zhi.zheng@vanderbilt.edu
[2] Vanderbilt Kennedy Center Treatment and Research Institute for Autism Spectrum Disorder, Nashville, TN, USA
[3] Mechanical Engineering Department, Vanderbilt University, Nashville, TN, USA

Abstract. Attentional skill, which is considered as one of the fundamental elements of social communication, is among the core areas of impairment among children with Autism Spectrum Disorder (ASD). In recent years, technology-assisted ASD intervention has gained momentum among researchers due its potential advantages in terms of flexibility, accessibility and cost. In this paper, we proposed a computer-assisted system for teaching attentional skills to toddlers with ASD, using the "response to name" skill as a specific example. The system was a fully closed-loop autonomous system capable of both providing name prompting from different locations of a room and detecting the child's attention in response to his name prompt. A preliminary user study was conducted to validate the proposed system and the protocol. The results showed that the proposed system and the protocol were well tolerated and were engaging for the participants, and were successful in eliciting the desired performance from the participants.

Keywords: Computer-mediated attention skills teaching · Toddlers with ASD

1 Introduction

Autism spectrum disorder (ASD) is a common disorder that impacts 1 in 68 children in the US [1]. Evidence suggests that early detection and intervention is critical to optimal treatment for ASD [2, 3]. Given the resource limitations in the healthcare system, technology-assisted approaches are being considered as potential intervention platforms due to their flexibility, controllability, duplicability, and cost effectiveness [4–9].

Previous work has examined technological interventions for older children with ASD. Feil-Seifer et al. [10] and Greczek et al. [11] proposed graded cueing mechanism for teaching imitation skills to children with ASD. Zheng et al. [12] studied adaptive gestures imitation training with fully autonomous robotic systems. Lahiri et al. [13], Wade et al. [8], and Herrera et al. [9] designed adaptive virtual reality response technologies for children with ASD in terms of conversation, a driving game, and symbolic game playing, respectively. While most of the existing technology-assisted intervention platforms target

M. Antona and C. Stephanidis (Eds.): UAHCI 2015, Part III, LNCS 9177, pp. 721–730, 2015.
DOI: 10.1007/978-3-319-20684-4_69

preschool or school-aged children, very few systems [14, 15] have been designed for potential applications to infants and toddlers, an age when the brain is still considered quite malleable. Robust systems capable of addressing meaningful core skill deficits during this time of neuroplasticity and prior to the full manifestation of ASD impairments could therefore have powerful impacts on children's long-term development.

We present a computer-assisted intervention system ultimately designed for use with infants and toddlers at risk for ASD. This system will address early social communication and social orienting skills, specifically response to name (RTN). Many children with ASD fail to orient to their names when called, nor do they share other core social attention bids with their caregivers as infants. These early orienting deficits often result in numerous lost social learning opportunities. The ultimate goal of the RTN protocol is to have the child successfully respond to caregiver's attempts to garner attention by calling his/her name from a variety of locations within the learning environment. The novel learning environment consists of:

- A spatially distributed name prompting system;
- A wide range attention detection system;
- An attractor that can help shift the child's gaze from the current location to the target location;
- A feedback mechanism to encourage the child. The integrated system works autonomously, and provides numerous practice sessions.

This paper is organized as follows. Section 2 introduces the system design. Section 3 presents the experimental setup. The experimental results are discussed in Sect. 4, and Sect. 5 draws conclusions about this work.

2 System Design

2.1 System Architecture

The proposed system was integrated as shown in Fig. 1. A supervisory controller (SC) controls the global execution logic. SC initializes the prompting sub-system and activates the name prompting or attractors as needed. It simultaneously receives the participant's gaze response from the attention tracking sub-system as feedback to make future decisions.

Figure 2 shows the experiment room layout. Five monitors in the name prompting sub-system covered a large range of the room to generate name calling prompts and provide attention attractors. The participant's attention was tracked by the 4 cameras in the Attention tracking Sub-system. If the child did not look at the target monitor, then a bouncing ball appeared in a monitor closest to his/her gaze direction and bounced towards the target monitor. If the participant still looked away, then special motion and sound effects were added to the bouncing ball to enhance the attractor's effect. If the participant looked at the target screen (target hit) during this procedure, then a reward video was displayed with a firework animation. We hypothesized that the attractor would reorient the participant towards the name calling location since children with ASD were shown to be attracted to inanimate objects. Such repeated practice and reward could

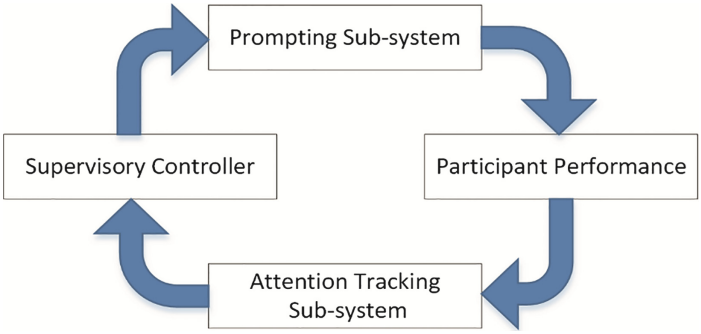

Fig. 1. Global system structure

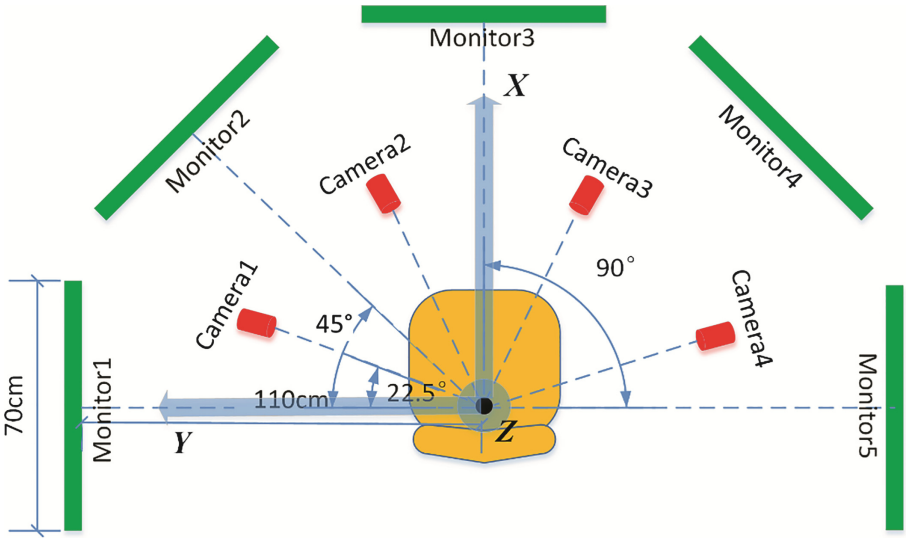

Fig. 2. Experiment room layout

teach children to develop an association between their names being called and looking towards the person, which may eventually lead to the development of RTN skills.

2.2 Name Prompting Sub-System

As shown in Fig. 2, the name prompting system consisted of 5 monitors positioned in a half circle. The participant was seated in the chair in the center of the circle. The center of Monitor1 to the center of Monitor5 created a view of 180° in yaw angle from the perspective of the participant. The radius of the monitor circle was 110 cm and the length of the monitor was 70 cm, so the display region was about 34° across for each monitor. The origin of the whole system was the center of the monitor half circle with a height of 120 cm. The ***X, Y,*** and ***Z*** axes in Fig. 1 are marked with the half-transparent wide

arrows. ***Z*** axis points up. When the participant was seated in the chair, the head of the participant was on the origin of the system. Each monitor was 43 cm in height, and the participant's eye was around the lower edge of the monitor. Each monitor was linked with a speaker. The five speakers behind the five monitors were driven by the 5.1 surround sound speaker system to provide stereo sound effect. All the visual and audio displays were programed with the Unity [16] game engine. The displays included: (1) a prerecorded name calling video from an experienced therapist; (2) a bouncing ball as the attention attractor, with different sound and motion effects; (3) a reward video and animation; (4) welcome and instruction videos; and (5) warm-up cartoon videos. Those displays on each monitor were controlled individually as a client by the supervisory controller.

2.3 Attention Tracking Sub-System

The attention direction, or the gaze, was computed from the head pose. As shown in Fig. 2, four cameras were embedded in the attention tracking system. Each camera was driven by the IntraFace [17] software, which computed the head pose using a supervised descent method [18]. While there are other algorithms [19, 20] on head pose estimation from a single camera, we chose this algorithm due to its robustness, high precision of detection, and real-time computation ability. Each camera's detection range was around [-40°, 40°] in yaw, [-30°, 30°] in pitch and [-30°, 30°] in roll with respect to the camera frame. In other word, if the participant faced a camera within those ranges, the head pose was tractable. As the name prompting sub-system prompted from a much larger range than a single camera's detection range, we applied an array of 4 cameras to expand the detection range as needed. By arranging the cameras along a half circle concentric with the monitor half circle, the participant's frontal face could be captured by at least one of the cameras when he/she looked toward any of the monitors. The angle between the adjacent cameras' optical axes was 45°, and all the cameras were calibrated to make their optical axes intersect at the origin of the global frame. All four cameras were set up at the same height of 120 cm and the transformation matrix R_{GC} from the camera's frame and global frame was computed. By arranging the cameras as in Fig. 2, each individual camera's detection was transferred and seamlessly merged in the global frame. The actual detection range used for the system was [-17°, 197°] from the right to the left side in yaw, [-30°, 21°] from up to down in pitch and [-30°, 30°] from lower left to lower right in roll. This range covered the full display space.

If more than one camera detected the head pose at the same time, the result from the camera with the smallest detected yaw angle was chosen for transformation. A vector V_{face}^{camera}, which represents the frontal face orientation in the camera frame was then transformed as $V_{face}^{global} = R_{GC} V_{face}^{camera}$ in the global frame. We found that the participant's sidewise gaze direction was usually larger than the sidewise head turn, therefore the horizontal component of the angle between V_{face}^{global} and the X axis was amplified by 120 % to approximate the gaze to the side. The ratio of 120 % was found to be reasonable with a small study with 3 adults. If the extended line from V_{face}^{global} crossed with the target monitor's region, then a target hit was detected.

3 Experimental Setup

3.1 Task and Protocol

The flowchart for the user study is shown in Fig. 3. Each participant completed a single experimental session consisting of the following steps. First, a welcome video was played, followed by a fun video (Sesame Street) displayed on all five monitors, (one at a time). This helped participants understand that a video could be displayed on those positions. After 5 trials of RTN another fun video was shown to prevent the participants from getting bored. Another 5 trials followed the break and finally the whole experiments ended up with a "Good-bye" video.

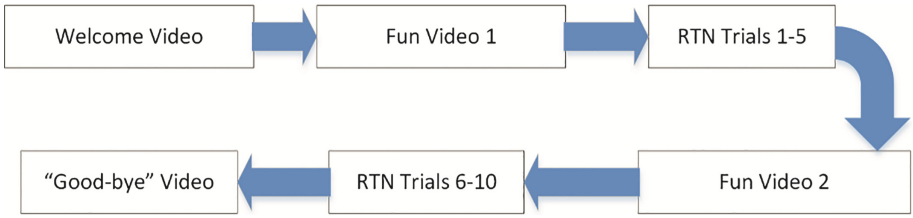

Fig. 3. Experimental procedure

The prompting levels of a RTN trial are listed in Table 1. The first level of prompt was a prerecorded name calling video displayed on the target monitor, where a therapist called the name of the participant twice. From prompt Level 2, a red bouncing ball appeared as the attention attractor. If the participant did not look at the target monitor within their name being called two times, the red ball started bouncing, starting at the direction of the participant's gaze and moving towards the target monitor. At the same time, the name calling video display was repeated at the target monitor. If the participant still did not look at the target monitor, the prompt Level 3 was introduced. Level 3 was the same as Level 2, except that the red ball bounced at the participant's gaze direction for 2 s before repeating the steps of Level 2. This initialized bouncing on the fixed position was designed to help the participant notice and focus on the attention attractor. If the participant still did not find the target monitor successfully, Level 4 was provided as the last level of prompting. Level 4 added a sound effect to the Level 3 prompts to make the bouncing ball more attractive. At any time during the 4 levels of prompts, if the participant looked at the target at any time, the prompting would be stopped and a reward video was displayed with a firework animation. The reward video was prerecorded by the same therapist in the name calling video saying, "Good job! You found me!" Even if the participant never looked at the target monitor, the rewards were still displayed for encouragement.

Table 1. Name Prompting Levels

Prompt level	Content
Level 1	Name calling on the target monitor
Level 2	Level 1 + Attractor bouncing from the gaze direction to the target monitor.
Level 3	Level 1 + Attractor ***first bouncing at the gaze direction for 2 s***, and then bouncing from the gaze direction to the target monitor.
Level 4	Level 1 + Attractor first bouncing at the gaze direction, and then bouncing from the gaze direction to the target monitor ***with special sound effect***.

3.2 Participants

The participants for the user study consisted of 5 typically developing (TD) children with the average age of 1.38 years and 5 children with ASD with an average age of 2.26 years. The children were recruited from a research registry of the Vanderbilt Kennedy Center. Children with ASD had confirmed diagnoses by a clinician based on DSM-IV-TR [21] criteria. They met the spectrum cut-off on the Autism Diagnostic Observation Schedule (ADOS) [22], and had existing data regarding cognitive abilities (Intelligent Quotient, or IQ) in the registry. Parents of participants in both groups completed the Social Responsiveness Scale– Second Edition (SRS-2) to index current ASD symptoms. The study was approved by the Vanderbilt University Institutional Review Board. The characteristics of the participants are listed in Table 2.

Table 2. Participant Characteristics

Mean (SD)	ADOS Raw Score	ADOS C2. Imagination/Creativity	IQ	SRS-2 Raw Score	SRS-2 T score	Age (Years)
ASD	22.40	2.88	55.00	92.80	69.60	*2.26*
	(3.65)	(0.45)	(8.49)	(22.69)	(8.85)	*(0.24)*
TD	NA	NA	NA	41.40	49.80	*1.38*
				(23.04)	(8.93)	*(0.44)*

The age range of the participants with ASD in this study was consistent with the age that many young children are able to be reliably diagnosed [23]. Therefore, this group provided the data regarding practicality of the proposed system for early intervention for children with ASD. Given that research has highlighted the importance of early screening and intervention for ASD in children, particularly those in high risk groups

(such as infant siblings) [3], we recruited the TD group with an average age lower than the typical age of ASD diagnosis to test its use with younger infants. Note that since the TD group was younger than the group with ASD, the results from each group were not directly comparable due to developmental differences. Instead, the results of each group were analyzed separately.

4 Experimental Results

All participants successfully completed all parts of the experiment. Their parents, the supervisory therapist as well as the engineers qualitatively noted that the participants were engaged in the tasks and seemed to enjoy the procedures. We investigated 2 important aspects of the participants' responses:

- The prompt level that the participants needed to hit the target. This signified how well the participant performed in the RTN tasks.
- How long the participants took from the start of the name calling until they looked at the target monitor.

4.1 Results of Participants with ASD

All participants with ASD eventually hit the target across all trials. 41 target hits were on prompt Level 1 and 9 target hits were on prompt Level 2. Table 3 lists the mean and standard deviation of the target hit prompt level and the time spent from the start of the name prompting to the target hit. On average, this group needed a prompt level of 1.18 to hit the target. The average time spent to hit the target was 1.85 s. Given that one name calling period from the name prompt video was around 2 s, we can see that the participants with ASD generally responded to their names the first time it was called.

Table 3. Results of Participants with ASD

	Prompt level needed	Time spent to hit the target
Mean	1.18	1.85
SD	0.39	0.95

4.2 Results of TD Participants

The TD participants hit the target in 49 trials in the whole experiment. 40 target hits were on prompt Level 1 and 9 target hits were on prompt Level 2. Table 4 lists the average and standard deviation of the target hit prompt level and the time spent from the start of the name prompting to the target hit of TD group. We can see that on average, this group needed a prompt level of 1.26 to hit the target. The average time spent to hit the target was 2.87 s. Overall, TD participants responded to their names on the second call.

Table 4. Results of TD participants

	Prompt level needed	Time spent to hit the target
Mean	1.26	2.87
SD	0.66	2.54

5 Discussion and Conclusion

In this paper, we proposed a computer-assisted system to help infants and toddlers with, and at risk of, ASD learn response to name skills. The system consisted of a name prompting sub-system and an attention tracking sub-system. The name prompting sub-system covered a wide range (over 180° yaw angle) in space to simulate a free and natural name calling environment. To detect the participant's response within this broad environment, we formed a closed loop system by introducing a wide range attention tracking sub-system that collaborated with the name prompting system.

The proposed system prompted name calling in a reinforced pattern and with an assistive attention attractor. During the name prompting training trials, if the participants could not respond to their names being called at a low prompting level, a higher prompting level was introduced with a red bouncing ball attractor to help shift the participant's gaze toward the target.

A user study was conducted to test the effectiveness of the proposed system on toddlers with ASD and infants. This allowed us to test the system on children with existing diagnoses as well as children who could be considered at risk of later diagnosis. The experimental results showed that the system and task protocol were well tolerated by the participants. They showed engagement during the intervention and performed well in response to name calling. Meanwhile, we found that the participants mainly utilized prompt Level 1 and Level 2. This revealed that the difficulty level of the name calling stimuli can be increased to scaffold the learning procedure and take advantage of advanced attention attractors. Based on these preliminary results, we will further investigate the children's response to more difficult name calling stimuli, such as audio only prompts from different directions. Furthermore, we will study the effect of the advanced attention attractor (Level 3 and Level 4) in the upgraded prompting environment.

Acknowledgement. The authors are grateful to the participants and their parents for their participation in this study. This study was supported in part by a grant from the Vanderbilt Kennedy Center (Hobbs Grant), the National Science Foundation under Grant 1264462, and the National Institute of Health under Grant 1R01MH091102-01A1 and 1R21MH103518-01. Work also includes core support from NICHD (P30HD15052) and NCATS (UL1TR000445-06).

References

1. Autism Spectrum Disorders Prevalence Rate. Autism Speaks and Center for Disease Control (CDC) (2011)
2. Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., et al.: Randomized, controlled trial of an intervention for toddlers with Autism: the early start denver model. Pediatrics **125**(1), e17–e23 (2010)
3. Warren, Z.E., Stone, W.L.: Best practices: Early diagnosis and psychological assessment. In: Amaral, D., Geschwind, D., Dawson, G. (eds.) Autism Spectrum Disorders, pp. 1271–1282. Oxford University Press, New York (2011)
4. Bekele, E., Lahiri, U., Swanson, A., Crittendon, J.A., Crittendon, Z., Sarkar, N.: A step towards developing adaptive robot-mediated intervention architecture (aria) for children with Autism. IEEE Trans. Neural Sys. Rehabil Eng. **21**(2), 289–299 (2013)
5. Greczek, J., Atrash, A., Matarić, M.: A computational model of graded cueing: robots encouraging behavior change. In: Stephanidis, C. (ed.) HCII 2013, Part II. CCIS, vol. 374, pp. 582–586. Springer, Heidelberg (2013)
6. Zheng, Z., Zhang, L., Bekele, E., Swanson, A., Crittendon, J., Warren, Z., et al.: Impact of Robot-mediated interaction system on joint attention skills for children with Autism In: Presented at the International Conference on Rehabilitation Robotics, Seattle (2013)
7. Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z., Sarkar, N.: Design of a virtual reality based adaptive response technology for children with Autism spectrum disorder. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011, Part I. LNCS, vol. 6974, pp. 165–174. Springer, Heidelberg (2011)
8. Wade, J., Bian, D., Zhang, L., Swanson, A., Sarkar, M., Warren, Z., Sarkar, N.: Design of a virtual reality driving environment to assess performance of teenagers with ASD. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014, Part II. LNCS, vol. 8514, pp. 466–474. Springer, Heidelberg (2014)
9. Herrera, G., Alcantud, F., Jordan, R., Blanquer, A., Labajo, G., De Pablo, C.: Development of symbolic play through the use of virtual reality tools in children with autistic spectrum disorders Two case studies. Autism **12**, 143–157 (2008)
10. Feil-Seifer, D., Matarić, M.: A simon-says robot providing autonomous imitation feedback using graded cueing. In: Poster paper in International Meeting for Autism Research (IMFAR), Toronto, May 2012
11. Greczek, J., Kaszubksi, E., Atrash, A., Matarić, M.J.: Graded Cueing Feedback in Robot-Mediated Imitation Practice for Children with Autism Spectrum Disorders. In: Proceedings of 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014), Edinburgh, August 2014
12. Zheng, Z., Das, S., Young, E.M., Swanson, A., Warren, Z., Sarkar, N.: Autonomous robot-mediated imitation learning for children with Autism. In: IEEE International Conference on in Robotics and Automation (ICRA), pp. 2707–2712 (2014)
13. Lahiri, U., Warren, Z., Sarkar, N.: Dynamic gaze measurement with adaptive response technology. In: Virtual Reality Based Social Communication for Autism. In: presented at the International Conference on Virtual Rehabilitation (ICVR) (2011)
14. Shic, F., Chawarska, K., Bradshaw, J., Scassellati, B.: Autism, eye-tracking, entropy. In: 7th IEEE International Conference on Development and Learning (ICDL), pp. 73–78 (2008)
15. Feil-Seifer, D., Mataric, M.: Robot-assisted therapy for children with Autism spectrum disorders. In: Proceedings of the 7th International Conference on Interaction Design and Children, pp. 49–52 (2008)
16. Unity Game Engine. Unity Game Engine-Official Site. http://unity3d.com

17. Xiong, X., De la Torre, F.: Supervised Descent Method for Solving Nonlinear Least Squares Problems in Computer Vision. arXiv preprint arXiv:1405.0601 (2014)
18. Xiong, X., De la Torre, F.: Supervised descent method and its applications to face alignment. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 532–539 (2013)
19. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**, 607–626 (2009)
20. Fanelli, G., Gall, J., Van Gool, L.: Real time head pose estimation with random regression forests. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 617–624 (2011)
21. Diagnostic and Statistical Manual of Mental Disorders: Quick reference to the Diagnostic Criteria from DSM-IV-TR, Fourth ed. Washington D.C.: American Psychiatric Association (2000)
22. Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., Bishop, S.: Autism Diagnostic Observation Schedule. Torrance, Western Psychological Services (2012). (ADOS-2)
23. Johnson, C.P., Myers, S.M.: Identification and evaluation of children with autism spectrum disorders. Pediatrics **120**, 1183–1215 (2007)