Design of a Virtual Reality Based Adaptive Response Technology for Children with Autism Spectrum Disorder

Uttama Lahiri[1], Esubalew Bekele[2], Elizabeth Dohrmann[3], Zachary Warren[4], and Nilanjan Sarkar[1,2]

[1] Mechanical Engineering Department, Vanderbilt University, Nashville, TN, USA
[2] Electrical Engineering and Computer Sc., Vanderbilt University, Nashville, TN, USA
[3] Vanderbilt Kennedy Center, Vanderbilt University, Nashville, TN, USA
[4] Psychiatry Department, Vanderbilt University, Nashville, TN, USA

Abstract. Impairments in social communication skills are thought to be core deficits in children with autism spectrum disorder (ASD). In recent years, several assistive technologies, particularly Virtual Reality (VR), have been investigated to promote social interactions in this population. It is well-known that these children demonstrate atypical viewing patterns during social interactions and thus monitoring eye-gaze can be valuable to design intervention strategies. However, presently available VR-based systems are designed to chain learning via aspects of one's performance only permitting limited degree of individualization. Given the promise of VR-based social interaction and the usefulness of monitoring eye-gaze in real-time, a novel VR-based dynamic eye-tracking system is developed in this work. The developed system was tested through a small usability study with four adolescents with ASD. The results indicate the potential of the system to promote improved social task performance along with socially-appropriate mechanisms during VR-based social conversation tasks.

Keywords: Autism Spectrum Disorder, Virtual Reality, eye tracking, social communication, engagement.

1 Introduction

Autism Spectrum Disorder (ASD) is characterized by core deficits in social interaction and communication [1]. Children with ASD demonstrate difficulties in social judgment and in using face as a channel for social communication [2]. Thus social attention and its measurement is considered a major aspect in ASD intervention [3], which could involve acquiring and analyzing eye-gaze data [4]. Also, given the limitation on the availability of trained professional resources in ASD intervention, it is likely that emerging technology will play an important role in providing more accessible intensive individualized intervention [5]. We chose Virtual Reality (VR) technology since this possesses several strengths in terms of potential application for children with ASD, namely, malleability, controllability, replicability, modifiable sensory stimulation, and an ability to pragmatically individualize intervention approaches and reinforcement strategies [6].

S. D´Mello et al. (Eds.): ACII 2011, Part I, LNCS 6974, pp. 165–174, 2011.

Despite potential advantages, current VR environments as applied to assistive intervention for children with ASD [7] [8] are designed to chain learning via aspects of performance alone. In fact, a number of VR applications have investigated social skill training for children with ASD [7] [8]. These works were pioneering in establishing the usefulness of VR in ASD intervention. However, they monitored the partipants' performance and did not measure how long and where the individuals were looking during interaction and thus were not able to assess their engagement to the task. Recently Wilms et al. [9] demonstrated the feasibility of linking the gaze behavior of a virtual character with a human observer's gaze position during joint-attention tasks. Thus, the current VR environments as applied to assistive intervention for children with ASD are designed to chain learning via aspects of performance alone (i.e., correct, or incorrect) thereby limiting individualization of application. Even these recent systems that measure eye gaze do not adaptively respond to engagement predicted from one's viewing pattern and eye physiological indices. The viewing patterns (e.g., Fixation Duration (FD)), and eye physiological indices (e.g., Blink Rate (BR) and Pupil Diameter (PD)) of an individual can be indicative of one's engagement [10-12], and if properly considered during a VR-based interaction, may lead to a substantially realistic intervention.

The objective of this paper is to present the design and development of a VR-based engagement-sensitive system with adaptive response technology that can be applied to social communication task for children with ASD. We design the system to present VR-based social tasks of varying difficulty levels along with bidirectional interaction in the form of social conversations with virtual avatars coupled with the ability to monitor one's behavioral viewing and eye physiological indices in real-time. Based on the participant's behavioral viewing, eye physiological indices, and performance metric, the system adaptively and socially responds by using a rule-governed strategy generator. While not an intervention study, here we present the results of a usability study as a proof-of-concept of our designed system.

2 System Design

The VR-based engagement-sensitive system with adaptive response technology has three main subsystems: (i) a VR-based social communication task module, (ii) a real-time eye-gaze monitoring module, and (iii) an individualized adaptive response module utilizing a rule-governed intelligent engagement prediction mechanism.

2.1 VR-Based Social Communication Task Module

In this work, we use desktop VR applications. Vizard (www.worldviz.com), a commercially available VR design package, is used to develop the virtual environments and the assistive technology. In order to perform socially interactive tasks with children with ASD, we developed extensive social situations with context-relevant backgrounds, and avatars whose age and appearance resemble those of the participants' peers without trying to achieve exact similarities. Also, for effective bidirectional social communication between the avatars and the participants, we developed conversation threads. Our social communication task module comprises of task presentation and bidirectional conversation modules.

Task Presentation Module. We developed social task presentation modules with avatars narrating personal stories to the participants. These stories were based on diverse topics of interest to teenagers e.g., favorite sport, favorite film, etc. The voices for the avatars were gathered from teenagers from the regional area. The avatar can make pointing gestures and move dynamically in a context-relevant virtual environment. For example, when an avatar narrates his tour experience to a sea beach and introduces the participant to the rocky beach while narrating the rocks on the beach, the VR environment reflects the view of the beach (Fig. 1a). When the avatar narrates some of his favorite activities on the beach such as, tanning during the day, the VR world displays such a situation to the participant (Fig. 1b). Subsequently, when the avatar narrates his experience of the remarkable view of sunset he witnessed on the beach, the VR situation changes with a smooth transition of the background image to display such a situation to the participant (Fig. 1c). This helped us to expose the participants to realistic social situations relevant to the topic being narrated.

The avatar heads were created from 2D photographs of teenagers, which were then converted to 3D heads by '3DmeNow' software for compatibility with Vizard. One can view the avatars within our system from first-person perspective.

Fig. 1. Snapshot of avatar narrating his tour experience to a sea beach within VR environment

Bidirectional Conversation Module. The participant was asked to watch and listen to the avatar narrating a personal story during the VR-based task presentation. At the end of this task presentation, the participant was asked to extract a piece of information from the avatar using a bidirectional conversation module with varying levels of interaction difficulty (e.g., ‘Easy’, ‘Medium’, and ‘High’). The bidirectional conversation module followed a menu-driven structure used by the interactive fiction community. The degree of interaction difficulty was controlled by requiring the number of questions a participant needed to ask in order to obtain a desired piece of information from the avatar and the nature of the conversation. In order to control the interaction difficulty, we chose three levels of conversation tasks, namely, Type 1, Type 2, and Type 3 tasks. In Type 1 conversation task, the participant was required to obtain a piece of information from the avatar that was directly narrated in the story and could be obtained by asking a minimum of 3 questions if the participant chose the right sequence of questions. In Type 2 conversation task, the participant was required to obtain a piece of information from the avatar that was not directly narrated in the story but was hinted during the narration and could be obtained by asking a minimum of 5 questions if the participant chose the right sequence of questions. Finally, in Type 3 conversation task, the participant was required to obtain a sensitive piece of information (e.g., the avatar’s feeling about a situation in general or some personal details, etc.) from the avatar that was not discussed in the story and could be obtained by asking a minimum of 7 questions if the participant chose the right sequence of questions. If a participant could acquire the needed information in a particular level by asking the minimum number of questions for that interaction difficulty level, he/she achieved the highest performance score. On the other hand, if the participant could not choose the right questions in the right sequence, causing him/her to ask more questions to find the needed information, he/she would acquire a proportionately less performance score. After a certain number of attempts if the participant was still unable to obtain the right information, the system would terminate the task and provide a task of lesser challenge to the participant. In order to ensure consistency among the tasks, task in each level of difficulty was carefully designed in consultation with experienced clinicians such that the structure of conversation remains similar regardless of the topics.

Type 1 Conversation Task with Easy Level of Interaction Difficulty. For an easy scenario, a participant was required to ask 3 appropriate questions in right sequence to get the intended piece of information using the conversation-thread structure as shown in Fig. 2.

For example, after an avatar narrated his experience of a football game during the VR-based task presentation, the participant was asked to find the avatar's experience of his first football game that he played. If the participant first selected the choice 3 (from the top) of the menu (Fig. 3) to introduce (represented by block A in Fig. 2) himself to the avatar, the avatar responded by saying "Hi. I am Tom. Yes. I really love football, especially when I get to play!" If instead of selecting choice 3, the participant chose any other option, then the avatar responded as, "I'm sorry! Do I know you? Maybe we should introduce ourselves first." thereby serving the role of a facilitator. Then if the participant selected the choice 1 (from the top) of the menu (Fig. 3) to ask the avatar about the topic (represented by block B in Fig. 2) of the conversation (i.e., regarding the first time the avatar played a football game), the avatar responded by saying "Of course! I was in the second grade. Our P.E. teacher split our class into two small junior football teams." Finally, if the participant selected the choice 2 (from the top) of the menu to ask the avatar regarding his overall feeling (represented by block C in Fig. 2) of his first football game, the avatar ended the conversation by saying "Yes, it was a lot of fun to play with my classmates."

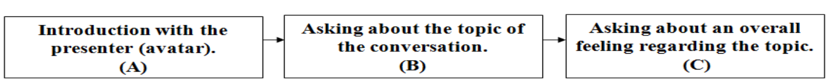

Fig. 2. Block Diagram of the Conversation Threads (Easy Level of Interaction Difficulty)

Type2 Conversation Task with Medium Level of Interaction Difficulty. For a scenario with a medium level of interaction difficulty, a participant was required to ask 5 questions in the right sequence to obtain the needed information from the avatar using a conversation-thread structure as shown in Fig. 4.

For example, after an avatar narrated her experience of her trip to a zoo with her friends during the VR-based task presentation, the participant was asked to find out some more details from the avatar about her experience at the zoo.

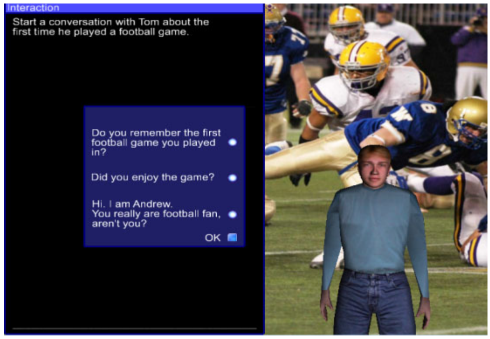

Fig. 3. Snapshot of a bidirectional conversation module (Easy Level of Interaction Difficulty)

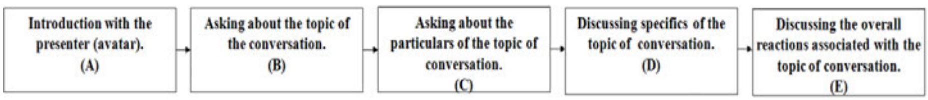

Fig. 4. Block Diagram of the Conversation Threads (Medium Level of Interaction Difficulty)

Type3 Conversation Task with High Level of Interaction Difficulty. For a scenario with a high level of interaction difficulty, a participant was required to ask 7 questions in the right sequence to obtain the needed information from the avatar using a conversation-thread structure as shown in Fig. 5.

For example, after an avatar narrated her experience of playing softball and her not liking the softball coach, the participant was asked to find out what the avatar actually felt for not liking the softball coach (i.e., 'sensitive' topic).

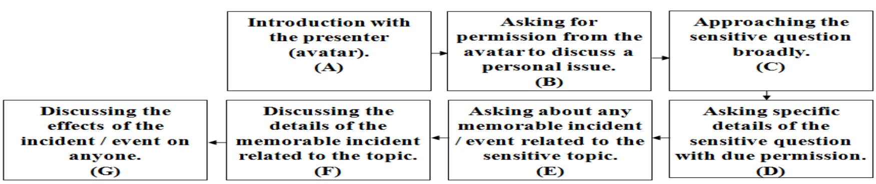

Fig. 5. Block Diagram of the Conversation Threads (High Level of Interaction Difficulty)

2.2 Real-Time Eye-Gaze Monitoring Module

The system captures eye data of a participant interacting with an avatar using eye-tracker goggles from Arrington Research Inc. This eye-tracker comes with some basic features acquiring capability for offline analysis and with a Video Capture Module with a refresh rate of 30 Hz to acquire a participant's gaze data using the 'Viewpoint' software. We designed the Viewpoint-Vizard handshake module, acquired the raw gaze data using Viewpoint, and transformed it to the Vizard compatible format using the handshake interface in a time synchronized manner. Subsequently, we applied signal processing techniques, such as windowing, thresholding, etc. to eliminate noise and extract the relevant features, such as, mean Pupil Diameter (PD_{Mean}), mean Blink Rate (BR_{Mean}), mean Fixation Duration (FD_{Mean}) at an interval of 33 ms.

2.3 Individualized Adaptive Response Module

Our goal was to switch tasks based on the composite effect of one's performance and engagement to the task. One's engagement to the VR-based social task was predicted based on objective metrics such as, dynamic viewing patterns characterized by Fixation Duration (FD) and two eye physiological indices - Blink Rate (BR), and Pupil Diameter (PD) – all of which have been shown to indicate engagement in other studies [10-12]. In order to discretize the engagement space, we assigned numeric values of 1, 2, and 3 (Tables 2a, 2b, and 2c) to these indices to quantify engagement in three levels – 'not engaged', 'moderately engaged' and 'highly engaged' respectively. If the cumulative sum of the engagement level obtained by real-time monitoring of FD, PD, and BR ≥ 6, then the engagement was considered as 'Good Enough', otherwise this was 'Not Good Enough'. If a participant scored ≥ 70% of the maximum score possible (e.g., 18, 30, and 42 for 'Type 1', 'Type 2', and 'Type 3' tasks representing the 'Easy', 'Medium', and 'High' difficulty level respectively with each relevant choice of bidirectional conversation options giving 6 points and an irrelevant choice causing a penalty of 3 points) in a task, then the performance was considered as 'Adequate', otherwise, this was considered as 'Inadequate'.

Table 2. Prediction of Engagement based on (a) Fixation Duration (FD), (b) Pupil Diameter (PD), and (c) Blink Rate (BR)

Fixation Duration	Value	Pupil Diameter	Value	Blink Rate	Value
$0\% \le T \le 50\%$	1	$PD_{Now} > PD_{Prev}$	1	$BR_{Now} > BR_{Prev}$	1
$50\% < T < 70\%$	2	$PD_{Prev} \ge PD_{Now} \ge 0.95PD_{Pre}$	2	$BR_{Prev} \ge BR_{Now} \ge 0.95BR_{Prev}$	2
$T \ge 70\%$	3	$PD_{Now} < 0.95PD_{Prev}$	3	$BR_{Now} < 0.95BR_{Prev}$	3

T : Percent FD towards Face region i.e., Face_ROI (during conversation) out of total FD.
PD_{Now}, PD_{Prev}:Pupil Diameter during present and previous situation respectively.
BR_{Now}, BR_{Prev}: Blink Rate during present and previous situation respectively.

This module fuses the information on the engagement (i.e., 'Good Enough' or, 'Not Good Enough') and the task performance (i.e., 'Adequate', or, 'Inadequate') to dynamically switch tasks of different difficulty levels by implementing an individualized task modification strategy (Table 3).

Table 3. Task Modification Strategy based on Composite Effect of Behavioral Viewing, Eye Physiology, and Performance

Case No.	Engagement	Task Performance	Overall Predicted Engagement
Case1	Good Enough	Adequate	Engaged
Case2	Good Enough	Inadequate	Not Engaged
Case3a/b	Not Good Enough	Adequate	Semi-Engaged
Case4	Not Good Enough	Inadequate	Not Engaged

In Case1, the strategy generator will increase task difficulty level. In Case2, task difficulty level will be reduced. In Case3a, the strategy generator will maintain tasks at the same level of difficulty and look out for an improvement in the next cycle. In case of no further improvement in the next cycle, task difficulty level will be reduced (Case 3b). In Case4, the task difficulty level will also be reduced. We implemented the dynamic task switching by a finite state machine representation [14] (Fig. 6).

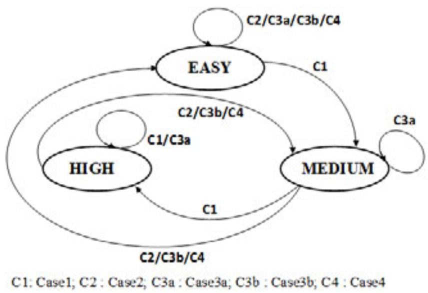

Fig. 6. State Machine Representation of Dynamic Decision Task Switching

3 Methods

3.1 Participants

Four adolescents with high-functioning ASD participated in this experiment. Their data on Peabody Picture Vocabulary Test (PPVT) [15], Social Responsive Scale (SRS) [16], Social Communication Questionnaire (SCQ) [17] Autism Diagnostic Observation Schedule (ADOS) [18], Autism Diagnostic Interview-Revised (ADI-R) [19] are presented in Table 4.

Table 4. Participant Characteristics

	Age (y)	PPVT	SRS	SCQ	ADOS	ADI-R
ASD1	17.58	134	80	12	13	49
ASD2	13.83	170	92	14	13	53
ASD3	18.25	97	63	17	9	49
ASD4	15.75	126	69	23	11	56

3.2 Procedure

Each participant participated in one session of VR-based social interaction task lasting for about 1 hour. During this visit, the participant sat comfortably on a height-adjustable chair and was asked to wear the eye-tracker goggles. They interacted with avatars narrating personal stories. They were asked to imagine that the avatars were their classmates at school giving presentations on several different topics. They were informed that after the presentations they would be required to interact with the avatar to find out some information from the avatar. They were also asked to try and make their classmate feel as comfortable as possible while listening to the presentation.

4 Results

We conducted a usability study to test whether the designed VR-based gaze-sensitive system with adaptive response technology functions as desired. Also, we tried to investigate whether the interaction with the system has the potential to impact the social task performance and behavioral viewing pattern of the participants with ASD.

4.1 Progression of VR-Based Social Communication Tasks

We analyzed the VR-based task progression for the participants across trials. ASD1 and ASD2 started at Medium difficulty level, and were able to achieve high difficulty level while interacting with our system (Fig. 7) with both of them interacting with 2 in Easy, 4 in Medium, and 3 in High difficulty levels. On the other hand, ASD3 started with Easy difficulty level and interacted with 7 VR-based social task trials with 2 in Easy, 2 in Medium, and 3 in High difficulty levels. ASD4 started with the High difficulty level and interacted with 4 VR-based social task trials in High difficulty level. Task switching ability of the system indicates the potential to individualize social communication training in the future.

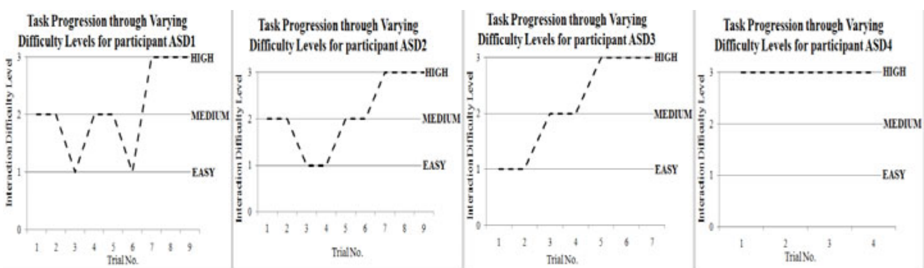

Fig. 7. VR-based Task Progression for ASD1, ASD2, ASD3, and ASD4 across trials. Interaction Difficulty Level=1: 'Easy'; Level=2: 'Medium'; Level=3: 'High'

4.2 Improvement in Behavioral Viewing Pattern

For effective social communication skills, one must not only achieve adequate task performance measures (e.g., obtaining the required information by asking the appropriate questions), but also carry out the conversation in a socially appropriate way (e.g., paying attention towards the face of the communicator during communication). We investigated to see whether there was any improvement in the behavioral viewing pattern of the participants while interacting with our VR-based engagement-sensitive system equipped with adaptive response technology.

Fig. 8 shows variation in the behavioral viewing pattern in terms of percent fixation duration towards the face region (Face_ROI) of the avatars of each participant across different trials. All participants demonstrated improvement in their behavioral viewing patterns in terms of greater attention towards the face of the communicator from trial1 to the last trial (trial9 for ASD1 and ASD2, trial7 for ASD3, and trial4 for ASD4). The mean percent fixation of all subsequent trials for each participant increased from their trial1 values by 26.58%. This data, although preliminary in nature, demonstrate that the feasibility of our system to impact the participants' looking patterns in a positive manner, i.e., they paid more attention to the avatars during conversation.

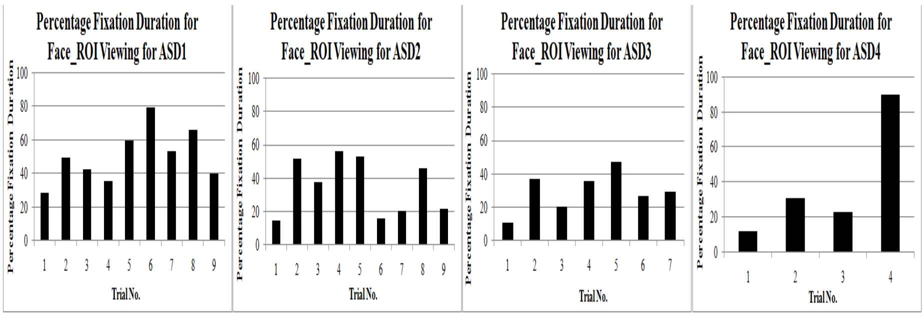

Fig. 8. Variation in Percent Fixation Duration while looking towards the face region (Face_ROI) of the avatar during the VR-based social communication task trials

5 Conclusion

In the present work we developed a VR-based engagement-sensitive system with adaptive response technology. In this paper we presented the system development and results of a small usability study to test the efficacy of the developed system as a first step to technology-assisted intervention. The developed system can detect subtle variations in one's eye-physiological features, and behavioral viewing pattern in real-time. Also it seamlessly integrates these pieces of information with the VR-platform to provide intelligent individualized adaptive response. Results of a usability study show the capability of the system to contribute to improving one's social task performance (e.g., ability of the participants to interact with VR-based social tasks with increased degree of difficulty) along with encouraging socially-appropriate mechanisms (e.g., improved attention towards the face of the communicator) to foster effective social communication skills among the participants with ASD. Such an integrated system has the potential to be incorporated into complex intervention paradigms aimed at improving functioning and quality of life for individuals with ASD.

Acknowledgments. The authors would like to acknowledge the National Science Foundation Grant 0967170 and the National Institute of Health Grant 1R01MH091102-01A1 that partially supported this research.

References

1. APA.: Diagnostic and statistical manual of mental disorders: DSM-IV-TR. American Psychiatric Association, Washington, DC (2000)
2. Kanner, L.: Autistic disturbances of affective contact. Nervous Child 2, 217–250 (1943)
3. Wieder, S., Greenspan, S.: Can Children with Autism Master the Core Deficits and Become Empathetic, Creative, and Reflective? The J. of Dev. and Learning Dis. 9 (2005)
4. Trepagnier, C.Y., Sebrechts, M.M., Finkelmeyer, A., Stewart, W., Woodford, J., Coleman, M.: Simulating social interaction to address deficits of autistic spectrum disorder in children. Cyberpsych. Behav. 9(2), 213–217 (2006)
5. Goodwin, M.S.: Enhancing and accelerating the pace of Autism Research and Treatment: The promise of developing Innovative Technology. Focus on Autism and other Developmental Disabilities 23, 125–128 (2008)
6. Strickland, D.: Virtual reality for the treatment of autism. In: Riva, G. (ed.) Virtual Reality in Neuropsychophysiology, pp. 81–86. IOS Press, Amsterdam (1997)
7. Parsons, S., Mitchell, P., Leonard, A.: The use and understanding of virtual environments by adolescents with autistic spectrum disorders. J. Autism. Dev. Disord. 34(4), 449–466 (2004)
8. Tartaro, A., Cassell, J.: Using Virtual Peer Technology as an Intervention for Children with Autism. In: Lazar, J. (ed.) Towards Universal Usability: Designing Computer Interfaces for Diverse User Populations, UK. John Wiley and Sons, Chichester (2007)
9. Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G.R., Vogeley, K.: It's in your eyes using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. Social Cognitive and Affective Neuroscience 5(1), 98–107 (2010)
10. NRC: Educating children with autism. National Academy Press, Washington, DC (2001)
11. Anderson, C.J., Colombo, J., Shaddy, D.J.: Visual Scanning and Pupillary Responses in Young Children with Autism Spectrum Disorder. J. of Clinical and Experimental Neuropsychology 28, 1238–1256 (2006)
12. Jensen, B., Keehn, B., Brenner, L., Marshall, S.P., Lincoln, A.J., Müller, R.A.: In-creased Eye-Blink Rate in Autism Spectrum Disorder May Reflect Dopaminergic Abnormalities. Intl. Society for Autism Research, Poster Presentation (2009)
13. Ruble, L.A., Robson, D.M.: Individual and environmental determinants of engagement in autism. J. of Aut. and Dev. Dis. 37(8), 1457–1468 (2006)
14. Booth, T.: Sequential Machines and Automata Theory. John Wiley and Sons, New York (1967)
15. Dunn, L.M., Dunn, L.M.: PPVT-III: Peabody Picture Vocabulary Test. Circle Pines. American Guidance Service, Minnesota (1997)
16. Constantino, J.N.: The Social Responsiveness Scale. W. Psych. Serv., California (2002)
17. Rutter, M., Bailey, A., Berument, S., Lord, C., Pickles, A.: Social Communication Questionnaire. Western Psychological Services, Los Angeles (2003a)
18. Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., DiLavore, P.C., Pickles, A., Rutter, M.: The Autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. J. of Autism and Dev. Dis. 30(3), 205–223 (2000)
19. Rutter, M., Couteur, A., Le, L.C.: Autism diagnostic interview revised WPS edition manual. W. Psych. Serv., Los Angeles (2003b)