Virtual Reality-Based Facial Expressions Understanding for Teenagers with Autism

Esubalew Bekele[1], Zhi Zheng[1], Amy Swanson[3], Julie Davidson[2,3], Zachary Warren[2,3], and Nilanjan Sarkar[4,1]

[1] Electrical Engineering and Computer Science Department
[2] Pediatrics and Psychiatry Department
[3] Treatment and Research in Autism Spectrum Disorder (TRIAD)
[4] Mechanical Engineering Department,
Vanderbilt University, Nashville, TN, USA
{esubalew.bekele,nilanjan.sarkar}@vanderbilt.edu

Abstract. Technology-enabled intervention has the potential to individualize and improve outcomes of traditional intervention. Specifically, virtual reality (VR) technology has been proposed in the virtual training of core social and communication skills that are impaired in individuals with autism. Various studies have demonstrated that children with autism have slow and atypical processing of emotional faces, which could be due to their atypical underlying neural structure. Emotional face recognition is considered among the core building blocks of social communication and early impairment in this skill has consequence on later complex language and communication skills. This work proposed a VR-based facial emotion recognition mechanism in the presence of contextual storytelling. Results from a usability study support the idea that individuals with autism may employ different facial processing strategies. The results are discussed in the context of the applicability of multimodal processing to enable adaptive VR-based systems in delivering individualized intervention.

Keywords: Social interaction, virtual reality, multimodal system, adaptive interaction, eye tracking, physiological processing, autism intervention.

1 Introduction

Individuals with ASD are known to demonstrate impaired recognition and understanding of emotional facial expressions [1]. They have been shown to have impaired face discrimination, and slow and atypical face processing strategies accompanied by reduced attention to eyes and context relevant areas rather than context irrelevant areas of the face [2]. Neural studies have indicated that children with ASD used different regions in the brain and relied on different strategies than their control groups while performing image matching based on facial expression tasks [3]. Research has demonstrated that recognition of socially derived emotional

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2013, Part II, LNCS 8010, pp. 454–463, 2013.

expressions, which are not isolated facial expressions, are difficult for individuals with ASD. It was also shown that they required more prompt and were slow to respond in facial expression identification tasks [4]. In general, children with ASD have significant impairment of understanding complex facial emotional expressions, their relations and what they entail [5]. In light of these evidences of impaired or atypical facial emotional processing and identification in children with ASD, it is important to address this emotion recognition skill deficit since emotion recognition is one of the building blocks of social communication. However, traditional behavioral intervention requires intensive behavioral sessions and is not accessible to wider ASD population. In this context, emerging technology such as virtual reality (VR) [6,7] have the potential to offer useful technology-enabled intervention systems that promise alternative or assistive therapeutic paradigms in increasing intervention accessibility, decreasing assessment efforts, reducing the cost of treatment, and ultimately promoting skill generalization [8]. Lack of adaptability of current VR systems to internal and implicit cues expressed by the participant restricts interaction. Therefore, incorporating modalities that can capture implicit cues expressed by the participant during the interaction using sensors such as peripheral physiological signals [9] and eye tracking [6] within the VR environment may help facilitating individualization and adaptation, and possibly accelerate the learning of emotional cues. Implicit cues offer more understanding of underlying psychological states which are not be possible using performance-based systems.

The objective of this paper is to present an analysis of physiological as well as eye tracking data from a usability study aimed at evaluating the efficacy of an innovative VR-based system for facial emotional expression identification task. These insights will drive development of future adaptive VR-based systems for social interaction. The remainder of the paper is organized as follows. Section 2 describes the overall technical aspect of the developed system while Section 3 is devoted to the methods and procedures followed in the usability study together with the summary of subjects who participated in the study. Section 4 presents the results and their implications. Finally, Section 5 concludes the discussion and highlight future extensions of this preliminary work.

2 System Development

The VR-based system is composed of an eye tracking application, a peripheral physiological monitoring and sensing and VR world presentation engine. It was developed to study the emotional face processing pattern differences by adolescents with autism as compared to a typically developing (TD) control group. The three applications ran separately while communicating via a network interface in a distributed fashion. The VR environment was rendered in Unity (www.unity3d.com) by Unity Technologies. A wireless physiological signals acquisition device called BioNomadix by Biopac Inc. (www.biopac.com) with 8 channels was used to record the physiological signals. A remote desktop eye tracker by Tobii Technologies (www.tobii.com) called Tobii X120 was employed to record the eye gaze data.

2.1 The VR Environment

The characters used in the VR environment were customized and rigged using an online animation service, mixamo (www.mixamo.com), and Autodesk Maya. All the facial expressions and lip-syncing for contextual stories narrated by the avatars were done in Maya. The characters were customized to suit the teenage age group targeted for the usability study, i.e., 13-17 years. A total of seven characters including 4 boys and 3 girls were selected and customized. Close to 20 of the facial bones were used for realistic facial expressions. The universally accepted 7 emotional expressions proposed by Ekman were animated for each character [10]. These are: enjoyment, surprise, contempt, sadness, fear, disgust, and anger. Each facial expression had four arousal levels (i.e. low, medium, high, and extreme).

2.2 Eye Tracking

The eye tracker recorded at 120 Hz frame rate allowing a free head movement of 30 x 22 x 30 cm (width x height x depth) at 70 cm distance. We used two applications connected to the eye tracker: one for diagnostic visualization as the experiment progresses and another one to record, pre-process and log the eye tracking data. The main eye tracker application computed eye physiological indices (PI) such as pupil diameter (PD) and blink rate (BR) and behavioral indices (BI) [11] such as fixation duration (FD) from raw gaze data. The velocity threshold identification (I-VT) algorithm [12] was implemented for fixation duration computation. The algorithm sets a velocity threshold to classify gaze points into saccade and fixation points. Generally, fixation points are characterized by low velocities (e.g., < 100 deg/sec) [12]. We used 35 pixels per sample (~ 60 deg/sec) as our velocity threshold. The blink rate was computed using condition code returned from the eye tracker whereas the pupil diameter was averaged for both eyes when both eyes data were available, and only using one eye data when the other eye was not in the field of view of the tracker. Five regions of interest (ROI) were defined on the face of each avatar including forehead, eyes (left and right), nose, and mouth. Facial regions outside of the 5 defined regions of interest were categorized as “other face regions” while all the background environment regions outside of the face regions were defined as “non-face regions”. This gave a total of seven regions into which all the gaze data points were clustered.

2.3 Physiological Processing

A total of 8 channels of physiological signals were acquired at 1000 Hz. The physiological signals monitored were: electrocardiogram (ECG), pulse plethesymogram (PPG), skin temperature (SKT), galvanic skin response (GSR), 3 electromyogram (EMG), and respiration (RSP). Due to social communication impairments in adolescents with autism, they are not usually expressive of their internal affective states and these

states often are not visible externally [13,14]. Physiological signals are, however, relatively less affected by these impairments and can be useful in understanding the internal psychological states and their pattern of children on the spectrum [15,14]. Among the signals we monitored, GSR, PPG, and ECG are directly related to the sympathetic response of the autonomic nervous system (ANS) [16]. When there is increased sympathetic activity due to external factors and pressures, the heart rate, the blood pressure, and sweating are all elevated [14]. The collected physiological data was analyzed to decipher any pattern differences between two situations: 1) when the subject correctly identified the emotion, and 2) when he/she did not correctly identified the emotion. A total of 51 features were extracted. These features were chosen because of their correlation with engagement and emotion recognition process as noted in psychophysiology literature [13,16,17]. The extracted features were mapped to a lower dimensional space using principal component analysis (PCA). Clustering analysis was performed using k-means clustering and Gaussian mixture clustering (GM).

3 Methods and Procedure

To study the behavioral and physiological pattern difference of adolescents with ASD as compared to those of typically developing children, a usability study involving 20 teenagers was completed.

Experimental Setup. The VR environment ran on Unity while eye tracking and peripheral physiological monitoring were performed in parallel using separate applications on separate machines that communicated with the Unity-based VR engine via a network interface. The VR task was presented using a 24'' flat LCD panel monitor. The experiment was performed in a laboratory with two rooms separated by one-way glass windows for parent observation. The parents sat in the outside room. In the inner room, the subject sat in front of the task computer. A therapist was present in the inner room at all times to monitor the process. The task computer display was also routed to the outer room for parent observation. The session was video recorded for the whole duration of the participation.

Subjects. A total of 10 high functioning subjects with ASD (M: n=8, F: n=2) of ages 13 – 17 (M=14.7, SD=1.1) and an age matched 10 TD (M: n=8, F: n=2) controls of ages 13 – 17 y (M=14.6, SD=1.2) were recruited and participated in the usability study. All ASD subjects were recruited through existing clinical research programs and had established clinical diagnosis of ASD at Vanderbilt University Kennedy Center. All subjects in the ASD group fell well above the clinical threshold (Table 1). The gold standard in clinical ASD diagnosis, the Autism Diagnostic Observation Schedule-Generic (ADOS-G) the new algorithm score and the severity score (ADOS-SS), were used to recruit the ASD subjects. IQ of the ASD subjects was obtained from existing clinical research database.

Table 1. Profile of subjects in the ASD group

Subjects		Age	ADOS-G (cutoff=7)	ADOS-CSS (cutoff=8)	SRS (cutoff=60)	SCQ (cutoff=15)	IQ
TD	Ave	14.5	N/A	N/A	39.3	1.9	111.1
	SD	1.38	N/A	N/A	3.44	1.97	11.14
ASD	Ave	14.7	12.0	7.4	79.9	17.78	118.7
	SD	1.1	2.1	1.11	5.34	7.36	9.55

The control group was recruited from the local community. To ensure control group subjects did not exhibit ASD related symptoms, we asked the parent to fill parent report of social responsiveness scale (SRS)[18] and social communication questionnaire (SCQ) [19]. Parents of both groups completed these forms. In addition, WASI [20] was used to measure IQ of the TD subjects. The IQ measures were used to potentially screen for intellectual competency to complete the tasks, mental retardation and group marching.

Tasks. The VR-based system presented a total of 28 trials corresponding to the 7 emotional expressions with each expression having 4 levels. Each trial was 12-15 s long. In each trial, first, the character narrated a context story that was linked to the emotional expression that followed for the next 5 s. The avatar exhibited a neutral emotional face during story telling. A typical laboratory visit was approximately one hour long. During the first 15 minutes, a trained therapist read approved ascent and consent documents to the subject and the parent, and explained the procedures. While the parent completed the SRS and SCQ forms, the subject wore the wearable physiological sensors with the help of a researcher. Before the task began the eye tracker was calibrated. The calibration was a fast 9 points calibration that took about 10-15 s. At the start of the task, a welcome screen greeted the subject and described what was about to happen and how the subject was to interact with the system. Immediately after the welcome screen, the trials started. At the end of each trial, questionnaires popped up asking the subject what emotion he/she thought the avatar displayed and how confident he/she was in his/her choice. The emotional expression presentations were randomized for each subject across trials to avoid ordering effects. To avoid other confounding factors arising from the context story, the story was recorded with monotonous tone and there was no facial expression displayed by the avatar during that context period.

4 Results and Discussions

The eye gaze and physiological data were separately analyzed. Results are also presented and discussed separately below.

4.1 Gaze Pattern Summary

We quantified gaze pattern as number of gaze points to a specific ROI as percentage of total number of gaze points. The ASD group was compared with TD group using these percentage gaze measures to ROIs. We have assessed statistical significance by using independent two sample unequal variance t-test to compare inter-group variations. Data were averaged across trials for each subject. Statistically significant gaze difference between the two groups was found for the mouth and the forehead ROIs as shown in Table 2.

Table 2. Gaze towards various ROIs as percentage of total gaze in a trial

	Nose	Left Eye	Right Eye	Eyes (Total)	Mouth*	Forehead*	Face	Non Face
ASD	10.92%	2.15%	5.36%	7.51%	11.91%	22.98%	80.37%	19.63%
TD	13.31%	3.78%	5.32%	9.10%	28.46%	9.31%	86.87%	13.13%

*p<0.05

The adolescents with ASD looked 11.32% ($p<0.05$) more towards the forehead area and 12.1% ($p<0.05$) less to the mouth area than the TD subjects. Note that both the forehead and the mouth areas primarily had the ROIs involved in most of the emotional expressions. Adolescents with ASD also looked 5.58% less to the eye area and 1.89% less to the nose area, albeit, not statistically significant manner.

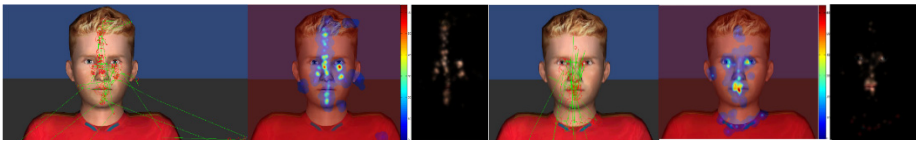

Fig. 1. Individual gaze comparison: ASD subject (left), TD subject (right). From left to right: fixation-saccade gaze plot, heat map and masked gaze visualizations.

Fig. 1 shows comparison of representative individual comparisons of gaze patterns between ASD subject and a control group subject. It can be seen, in this example, that the ASD face scanning pattern is different from that of the TD subject. The ASD pattern seem a little more irregular and have a randomized distribution tendency while the TD subject's gaze is concentrated in to crucial ROIs of the face such as eyes and mouth.

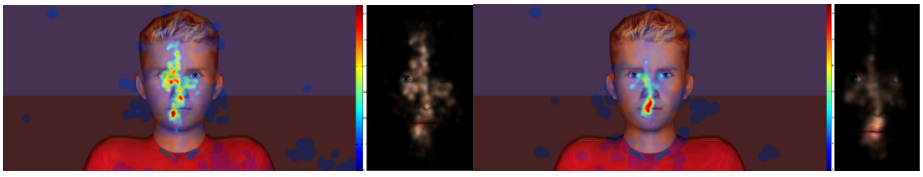

Fig. 2. Combined (group) gaze comparison: ASD group (left), TD group (right). From left to right: heat map and masked gaze visualizations.

Although less random than the individual representative pattern shown in Fig. 1, the combined group comparison (Fig. 2) between the ASD and TD groups also follows similar differences in the gaze patterns between the two groups. The TD group gaze seems to be more concentrated while that of the ASD group is more spread out.

4.2 Eye Behavioral Indices (BI) and Eye Physiological Indices (PI)

Average fixation duration (FDave), the average saccade path length (SPLave) and sum of fixation counts (SFC) were used as behavioral viewing pattern measures in this study. These behavioral indices are indicative of engagement to particular stimuli and are correlated with social functioning for individuals with autism [21]. Generally, adolescents in the ASD group had lower FD and SFC than the control group (Table 3). However, none of the differences were statistically significant.

Table 3. Measures of behavioral viewing pattern

	PDave (mm)	BRave (bpm)	FDave (ms)	SPLave (pix)	SFC (no unit)
ASD	3.21	5.34	414.84	116.08	26.90
TD	3.61	12.26	471.59	128.27	32.26

mm: millimetres, bpm: blinks per minute, pix: pixels, *$p<0.05$

Unlike reported high blink conditioning for individual with ASD in general [22], these low blinks (Table 3) could be attributed to 'sticky attention' that this population sometimes exhibits [23].

4.3 Physiological Analysis

Identifying physiological pattern differences when there were external factors such as stress (e.g., the fear of being incorrect) within a social task such as emotion identification is important for the development of an affective state detection system to enable adaptive VR social interactive task. To investigate how much separable these data are, we used unsupervised clustering and compared it to ground truth clusters. The physiological data of the adolescents in both groups was separated into data from trials when the adolescents were correct and that of trials when they were incorrect in identifying the emotion displayed by the avatar in those specific trials. The combinations of these four dataset were clustered using k-means and Gaussian mixture (GM) clustering methods taking two datasets at a time. These resulted in four comparisons. Accuracy here represents cluster quality as compared to the ground truth. Fig. 3 (Right) shows one of the comparisons, i.e., data clusters from ASD subjects when they were correctly identifying the emotions vs. when they were not able to identify the emotions.

We used both the first two components (2D) in one case and the first three (3D) components in another case of the PCA output as both contained more than 90% of the information in the original feature set. Overall, the k-means and the GM achieved

accuracies of more than 55% using only the first two components and more than 60% using the first three components on average for 30 runs in each comparison. Interesting results were seen for within group comparison of ASD. The GM was able to cluster the correct and incorrect trials of the ASD group with average accuracy of 72.9% and k-means resulted in 75% for same comparison as shown in Fig. 3 (Left). This is important as it shows that underlying physiological patterns of children with ASD can be separable by a machine learning algorithm when they are under stress (in this case when they were incorrect) and when they were not (in this case when they were correct). The two clustering methods were able to achieve maximum accuracies of 94% and 91%, respectively.

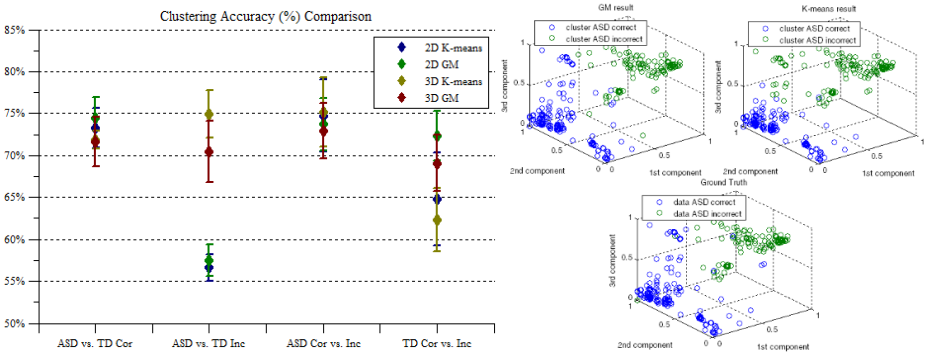

Fig. 3. Left: Results showing clustering quality of the two clustering methods using two different sets of PCA components for all the four sets of comparisons. Cor: correct and Inc: incorrect. Right: (a) Top left: result of the GM clustering, (b) Top right: result of the k-means clustering, and (c) Bottom: ground truth of clusters of data of from trials of ASD subjects when they were correctly identifying the emotions vs. when they were incorrect in identifying the emotions.

In general, these results indicate that there are clear pattern differences in physiological responses of adolescents with ASD while performing social tasks such as identifying emotional faces. Given enough training data, these pattern differences can be learned using supervised non-linear classifiers to enable adaptive VR-based social interaction in the future. Liu et al. [9], for instance, showed that it is possible to use such physiological measures to create an adaptive closed-loop robotic interaction in real-time using support vector machines (SVM) with Gaussian kernels.

5 Conclusion and Future Works

The system successfully presented the facial emotional expressions and collected the synchronized eye gaze and physiological data. Results indicated interesting differences in performance regarding identification of certain emotions as well as differences in how individuals with ASD often processed emotional faces. Individuals with ASD recognized expressions with greater accuracy than the TD group. However,

we did in fact find interesting differences in how facial expressions were processed and decoded between these groups. Specifically, there were significant eye gaze differences in the mouth and the forehead areas with adolescents with ASD paying significant attention to context irrelevant area such as the forehead while the TD group focused more on the context relevant ROI such as the mouth. Adolescents in the ASD group also paid less attention to the eye area than the TD group on average, although these differences were not statistically significant. Although our system was not designed to map specific physiological responses, our offline analysis demonstrated that meaningful physiological pattern differences could be detected during system performance. Such differences may be potentially modeled onto constructs of stress and engagement over time in order to further enhance and endow our system with the ability to understand processing and performance. There were several important limitations to note. First, this was a static performance driven system, and physiological indices were not incorporated into online performance for adaptation. Further, our design of the emotional expressions, while based on decades of research and theory, was not adequately able to push for accurate identification of certain emotions across groups. Finally, while the system created some sense of a social scenario, interactions were limited in the application. Despite limitations, this initial study demonstrates the value of future work subtly adjusting emotional expressions, integrating this platform into more relevant social paradigms, and embedding online physiological and gaze data to guide interactions with potential relevance toward fundamentally altering and improving how individuals with ASD process nonverbal communication within and hopefully outside of VR environments.

Acknowledgement. This work was supported in part by National Science Foundation Grant [award number 0967170] and National Institute of Health Grant [award number 1R01MH091102-01A1].

References

1. Adolphs, R., Sears, L., Piven, J.: Abnormal processing of social information from faces in autism. Journal of Cognitive Neuroscience 13(2), 232–240 (2001)
2. Dawson, G., Webb, S.J., McPartland, J.: Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. Developmental Neuropsychology 27(3), 403–424 (2005)
3. Wang, A.T., Dapretto, M., Hariri, A.R., Sigman, M., Bookheimer, S.Y.: Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder. Journal of the American Academy of Child & Adolescent Psychiatry 43(4), 481–490 (2004)
4. Capps, L., Yirmiya, N., Sigman, M.: Understanding of Simple and Complex Emotions in Non‐retarded Children with Autism. Journal of Child Psychology and Psychiatry 33(7), 1169–1182 (1992)
5. Weeks, S.J., Hobson, R.P.: The salience of facial expression for autistic children. Journal of Child Psychology and Psychiatry 28(1), 137–152 (1987)

6. Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z., Sarkar, N.: Design of a Virtual Reality based Adaptive Response Technology for Children with Autism. IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society PP (early access) (99), 1 (2012)
7. Standen, P.J., Brown, D.J.: Virtual reality in the rehabilitation of people with intellectual disabilities: review. Cyberpsychology & Behavior 8(3), 272–282 (2005)
8. Goodwin, M.S.: Enhancing and Accelerating the Pace of Autism Research and Treatment. Focus on Autism and Other Developmental Disabilities 23(2), 125–128 (2008)
9. Liu, C., Conn, K., Sarkar, N., Stone, W.: Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. International Journal of Human-Computer Studies 66(9), 662–677 (2008)
10. Ekman, P.: Facial expression and emotion. American Psychologist 48(4), 384 (1993)
11. Lahiri, U., Warren, Z., Sarkar, N.: Design of a Gaze-Sensitive Virtual Social Interactive System for Children With Autism. IEEE Transactions on Neural Systems and Rehabilitation Engineering (99), 1 (2012)
12. Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. Paper Presented at the Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (2000)
13. Liu, C., Conn, K., Sarkar, N., Stone, W.: Online affect detection and robot behavior adaptation for intervention of children with autism. IEEE Transactions on Robotics 24(4), 883–896 (2008)
14. Picard, R.W.: Future affective technology for autism and emotion communication. Philosophical Transactions of the Royal Society B: Biological Sciences 364(1535), 3575–3584 (2009)
15. Groden, J., Goodwin, M.S., Baron, M.G., Groden, G., Velicer, W.F., Lipsitt, L.P., Hofmann, S.G., Plummer, B.: Assessing cardiovascular responses to stressors in individuals with autism spectrum disorders. Focus on Autism and Other Developmental Disabilities 20(4), 244–252 (2005)
16. Cacioppo, J.T., Tassinary, L.G., Berntson, G.G.: Handbook of psychophysiology. Cambridge Univ. Pr. (2007)
17. Welch, K.C., Lahiri, U., Liu, C., Weller, R., Sarkar, N., Warren, Z.: An Affect-Sensitive Social Interaction Paradigm Utilizing Virtual Reality Environments for Autism Intervention. In: Jacko, J.A. (ed.) HCI International 2009, Part III. LNCS, vol. 5612, pp. 703–712. Springer, Heidelberg (2009)
18. Constantino, J., Gruber, C.: The social responsiveness scale. Western Psychological Services, Los Angeles (2002)
19. Rutter, M., Bailey, A., Lord, C., Berument, S.: Social communication questionnaire. Western Psychological Services, Los Angeles (2003)
20. Wechsler, D.: Wechsler Abbreviated Scale of Intelligence® (WASI®-IV), 4th edn. Harcourt Assessment, The Psychological Corporation, San Antonio (2008)
21. Klin, A., Jones, W., Schultz, R., Volkmar, F., Cohen, D.: Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Archives of General Psychiatry, Am. Med. Assoc. 59(9), 809–816 (2002)
22. Sears, L.L., Finn, P.R., Steinmetz, J.E.: Abnormal classical eye-blink conditioning in autism. Journal of Autism and Developmental Disorders 24(6), 737–751 (1994)
23. Landry, R., Bryson, S.E.: Impaired disengagement of attention in young children with autism. Journal of Child Psychology and Psychiatry 45(6), 1115–1122 (2004)