An Affect-Sensitive Social Interaction Paradigm Utilizing Virtual Reality Environments for Autism Intervention

Karla Conn Welch[1], Uttama Lahiri[2], Changchun Liu[1], Rebecca Weller[2], Nilanjan Sarkar[2,1], and Zachary Warren[3]

[1] Department of Electrical Engineering and Computer Science
[2] Department of Mechanical Engineering
[3] Vanderbilt Kennedy Center and Department of Pediatrics
Vanderbilt University, Nashville, TN 37212
karla.conn-welch@vanderbilt.edu

Abstract. This paper describes the design and development of both software to create social interaction modules on a virtual reality (VR) platform and individualized affective models for affect recognition of children with autism spectrum disorders (ASD), which includes developing tasks for affect elicitation and using machine-learning mathematical tools for reliable affect recognition. A VR system will be formulated that can present realistic social communication tasks to the children with ASD and can monitor their affective response using physiological signals, such as cardiovascular activities including electrocardiogram, impedance cardiogram, photoplethysmogram, and phonocardiogram; electrodermal activities including tonic and phasic responses from galvanic skin response; electromyogram activities from corrugator supercilii, zygomaticus major, and upper trapezius muscles; and peripheral temperature. This affect-sensitive system will be capable of systematically manipulating aspects of social communication to more fully understand its salient components for children with ASD.

Keywords: Human-computer interaction, Physiological responses, Virtual Reality, Autism, Affective model.

1 Introduction

The development of technological tools that can make application of productive intensive treatment more readily accessible and cost effective is an important new direction for research on autism spectrum disorders (ASD) [1, 2]. A growing number of studies have been investigating the application of advanced interactive technologies to address core deficits related to autism, namely computer technology [3-5], robotic systems [6-8], and virtual reality environments [9-11]. There is increasing consensus in the autism community that development of assistive tools that exploit advanced technology will make application of intensive intervention for children with ASD more efficacious.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2009, LNCS 5612, pp. 703–712, 2009.

Virtual reality (VR) represents a medium well-suited for creating interactive intervention paradigms for skill training in the core areas of impairment for children with ASD (i.e., social interaction, social communication, and imagination). VR-based therapeutic tools can partially automate the time-consuming, routine behavioral therapy sessions and may allow intensive intervention to be conducted at home [6,10]. However, no existing VR-based system specifically addresses the issue of how to detect and flexibly respond to affective cues of children with ASD within an intervention paradigm. Affective cues are insights into the emotions and behaviors of children with ASD, and as such the ability to utilize the power of these cues is critical given the importance of human affective information in human-computer interaction [12], the significant impacts of the affective factors of children with ASD on intervention [13], and the core social and communicative vulnerabilities that limit individuals with ASD to accurately self-identify affective experiences [14].

There are several modalities such as facial expression, vocal intonation, gestures and postures, and physiology that can be utilized to evaluate the affective states of individuals interacting with a computer. In this work, the affective models are based on physiological data for several reasons. Children with ASD often have communicative impairments (both nonverbal and verbal), particularly regarding expression of affective states. These vulnerabilities place limits on traditional conversational and observational methodologies; however, physiological signals are continuously available and are arguably not directly impacted by these difficulties [15]. As such, physiological modeling may represent a methodology for gathering rich data despite potential communicative impairments. Furthermore, there is evidence for typical individuals that the physiological activity associated with various VR experiences is differentiated [16], and the transition from one affective state to another state is accompanied by dynamic shifts in indicators of autonomic nervous system activity [17].

The ultimate objectives of this work are to develop technologies capable of flexibly responding to subtle affective changes in individuals with ASD during social paradigms and to take current VR-based ASD intervention technology to a higher level such that it can present itself as a realistic and powerful intervention platform. This paper describes the current VR task design and affective modeling techniques that could eventually be used during closed-loop interactions in which the VR system autonomously responds to the affective cues of a child with ASD. The framework of the affect-sensitive VR system during closed-loop interaction with a child with ASD is presented in Fig. 1. The physiological signals from the children with ASD are recorded while they are interacting with the VR system. These signals are processed in real time to extract features, which are fed as input into the developed affective models. The models map the features to an intensity level of an affective state and return this information as an output. The affective information is used by a controller to decide the next course of action for the VR system. The child who engages with the system is then influenced by the system's behavior, and the closed-loop interaction cycle begins anew.

In particular, the affect-sensitive VR system will monitor the affective states of anxiety and engagement, measured by physiological signals which vary with respect

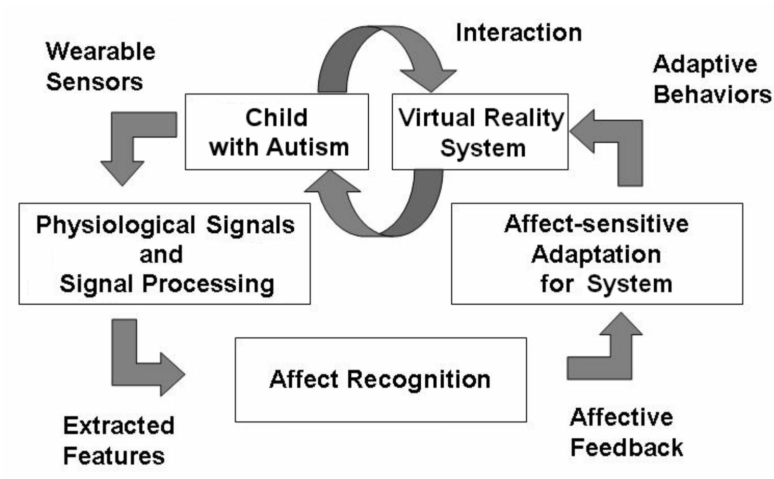

Fig. 1. Framework of the affect-sensitive VR system during closed-loop interaction

to the variation of specific communication factors (e.g., proximity and eye contact) presented in the VR environment. The discriminating capability of the physiological features will be measured to identify ones that have significant influence during social communication in the VR environment for the children with ASD. Assessment of the prediction accuracy of the affective models could be useful in developing intervention applications to help children with ASD explore social interaction dynamics in an adaptive and customized manner.

2 Physiological Features and Affective Modeling

More than one physiological signal, judged as a favorable approach [18], will be examined. The collected physiological signals are cardiovascular activities including electrocardiogram (ECG), impedance cardiogram (ICG), photoplethysmogram (PPG), and phonocardiogram (PCG)/heart sound; electrodermal activities (EDA) including tonic and phasic responses from galvanic skin response (GSR); electromyogram (EMG) activities from corrugator supercilii, zygomaticus major, and upper trapezius muscles; and peripheral temperature. These signals were selected because they are likely to demonstrate variability as a function of the targeted affective states, can be measured non-invasively, and are relatively resistant to movement artifacts [19].

Signal processing techniques will be used to derive the relevant features from the physiological signals. For example, inter beat interval (IBI) is the time interval between two "R" waves in the ECG waveform. Time-domain features of IBI, the mean and standard deviation (SD), are computed from the detected R peaks. Power spectral analysis is performed on the IBI data to localize the sympathetic and parasympathetic nervous system activities associated with different frequency bands. The high-frequency power (0.15-0.4 Hz) is associated with parasympathetic nervous system activity. The low-frequency power (0.04-0.15 Hz) provides an index of sympathetic effects on the heart. Very-low frequency power is associated with the frequency band less than 0.04 Hz. The ratios of different frequency components are also computed as the input features for affective modeling. The PPG signal measures changes in the volume of blood in the finger tip associated with the blood volume pulse (BVP) cycle. Pulse transit time (PTT) is estimated by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the BVP wave reaching the peripheral site. Besides PTT, the mean and SD values of BVP peak amplitudes are also extracted as features. Pre-ejection period (PEP), derived from ICG and ECG, measures the latency between the onset of electromechanical systole and the onset of left-ventricular ejection. The time intervals between the successive peaks of ICG time-derivative and ‘‘R’’ peaks of ECG are calculated to obtain the value of PEP. The features obtained are the mean of PEP and the average time interval between two peaks of the ICG time-derivative. The features extracted from the heart sound signal consist of the mean and SD of the third (138-275 Hz), fourth (69-138 Hz), and fifth (34-69 Hz) level coefficients of the Daubechies wavelet transform.

EDA consists of two main components - tonic response and phasic response. The phasic skin conductance detection algorithm uses the following heuristics for considering a particular peak as a valid skin conductance response: (i) the slope of the rise to the peak should be greater than 0.05 μSiemens/min; (ii) the amplitude should be greater than 0.05 μS; and (iii) the rise time should be greater than 0.25s. Once the phasic responses are identified, the rate of the responses and the mean and maximum amplitude are determined as features. All the signal points that are not included in the phasic response constitute the tonic part of the skin conductance signal. The slope of tonic activity is obtained using linear regression. Another feature derived from tonic response is the mean tonic amplitude.

The EMG signal from corrugator supercilii muscle (eyebrow) detects the tension in that region, and the EMG signal from the zygomaticus major muscle (cheek) captures the muscle movements while smiling. Upper trapezius muscle EMG activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. Time-domain features, the mean, SD, and slope are calculated from the EMG signals after performing a bandpass filtering operation (10-500 Hz). The analysis of the EMG activities in the frequency domain involve applying fast Fourier transform on a given EMG signal, integrating the EMG spectrum, and normalizing it to [0,1] to calculate the two features of interest – the median frequency and mean frequency for each EMG signal. The blink-related features are determined from the corrugator supercilii EMG signals after being preprocessed by a low-pass filter (10 Hz). The peripheral temperature signal is down-sampled by 10 and filtered to remove high-frequency noise, from which the mean, SD, and the slope are calculated as features.

In order to have reliable reference points to link the physiological feature sets to the affective states, subjective reports on the affective states from a therapist, each participant's parent, and the child himself/herself are collected. Each participant has a dataset comprised of both the objective physiological features and corresponding subjective reports on intensity of the target affective states from the three reporters. Using support vector machines (SVM), we can build affective models to map between the physiological features and the intensity (i.e., high/low), as reported on questionnaires, of a particular affective state. As illustrated in Fig. 2, a therapist-like

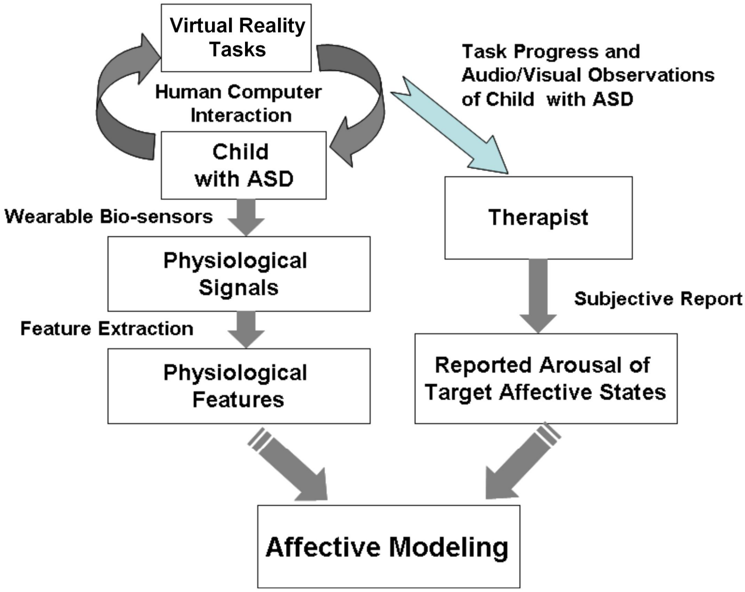

Fig. 2. Overview of affective modeling when the therapist's subjective reports are used

affective model (i.e., a model that captures the therapist's ability to assess affective states) can be developed when the therapist's reports are used. This process of differentiating high/low levels of the target affective states from physiological signals attempts to emulate present autism intervention practices and to experimentally demonstrate the feasibility of affective modeling for children with ASD via psychophysiological analysis.

In this work we chose anxiety and engagement to be the target affective states. Anxiety was selected for two primary reasons. First, anxiety plays an important role in various human-machine interaction tasks that can be related to task performance [20]. Second, anxiety is not simply a frequently co-occurring disorder; in some ways it is also a hallmark of autism [21]. Engagement has been regarded as one of the key factors for children with ASD to make substantial gains in academic, communication, and social domains [22].

3 Experimental Design

3.1 VR Task Design

We created realistic VR scenarios for social interaction with virtual human characters (i.e., avatars) using the Vizard Virtual Reality Toolkit (www.worldviz.com). To prevent possible "cybersickness" [23] and since in ASD intervention VR is often effectively experienced on a desktop system using standard computer input devices [2], our participants view the VR environment on a computer monitor from the first-person perspective. Within the controllable VR environment, components of the interaction are systematically manipulated to allow the participants to explore different social compositions.

The social parameters of interest, eye gaze and social proximity, are examined in a 4x2 experimental design. These parameters were chosen because they play significant roles in social communication and interaction [24], and manipulation of these factors may elicit variations in physiological responses [25]. The eye gaze parameter dictates the percentage of time an avatar looks at the participant (i.e., staring straight out of the computer monitor) or away from the participant. Four types of eye gaze will be examined. These are tagged as "straight," "averted," "normal," and "flip of normal," which correspond to the avatar staring straight ahead 100%, 0%, 30%, and 70% of the time during the interaction, respectively. The social proximity parameter is characterized by the distance between the avatar and the user. Two types of social proximity, termed "invasive" and "decorum," will be examined. For invasive proximity, the avatar stands approximately 1.5ft from the main view of the scene. This invasive distance is characterized by eliciting uncomfortable feelings and attempts to increase the distance to achieve a social equilibrium consistent with comfortable social interaction [26]. Decorum proximity means the avatar stands approximately 4.5 ft from the main view of the scene, and research indicates this distance results in a more comfortable conversation experience than the invasive distance [26]. Using Vizard software we will project avatars who display different eye gaze patterns at different distances (two examples shown in Fig. 3).

Each social interaction situation is represented three times, which creates 24 trials/epochs in the experiment. Each epoch includes one avatar for one-on-one interaction with the participant. In each epoch, participants are instructed to watch and listen as the avatar tells a 2min. first-person story. At the end of the story, the avatar asks the participant a question about the story. The questions were designed to facilitate interaction and to serve as a possible objective measure of engagement. Thus, the task can be likened to having different people introduce themselves to the user, which is comparable to research on social anxiety and social conventions [26][27]. Other social parameters, such as facial expression and vocal tone have been kept as neutral as possible. However, we also attempt to make the task interesting

Fig. 3. At *left* an avatar displays straight gaze at the invasive distance, while on the *right* an avatar stands at the decorum distance and looks to her right in an averted gaze

enough so that participants do not become excessively detached based on habituation or dull content. Our design is currently being evaluated by lab members and a pool of undergraduate students to determine to what extent these objectives are being met.

The Virtual Human Interaction Lab at Stanford University has provided distinct avatar heads created from front and side 2D photographs of college-age students using 3DMeNow software. The stories the avatars share were adapted from DIBELS (Dynamic Indicators of Basic Early Literacy Skills) reading assessments (dibels.uoregon.edu/measures/). The voices for the avatars were gathered from teenagers and college-age students from the regional area. Their ages (range = 13-22 years, mean = 18.5 years, SD = 2.3 years) are similar to the age of people used for the avatar heads and our participant pool.

3.2 Experimental Procedure

The experiment will include 24 social interaction epochs, broken up over two 1-h sessions. Each session will take place on a different day to avoid bias in data due to habituation. Participants will relax in a seated position and read age-appropriate leisure material during a 3-min baseline recording at the beginning of each session to offset day-variability. A second audio-only baseline including components of the subsequent epochs (e.g., story and corresponding question) except for the appearance of an avatar will also be recorded for comparison. Subsequently, emotional stimuli induced by the VR tasks will be applied in epochs of 2min. in length.

While interacting with the VR environment, each participant's physiological signals will be recorded using the Biopac MP150 data acquisition system (www.biopac.com). The physiological sensors are small, lightweight, non-invasive, and FDA approved. ECG will be measured from the chest using a two-electrode configuration. ICG will be measured by four pairs of surface electrodes that are longitudinally configured on both sides of the body along the neck and torso. A microphone specially designed to detect heart sound waves will be placed on the chest to measure PCG. PPG, peripheral temperature, and EDA will be measured from the middle finger, the thumb, the index finger, and the ring finger of the non-dominant hand, respectively. EMG will be measured by placing surface electrodes on two facial muscles (corrugator supercilii and zygomaticus major) and an upper back muscle (upper trapezius).

As shown in Fig. 4, the equipment setup for collecting physiological data for psychophysiological analysis includes a computer dedicated to the social interaction tasks where the participants interact with the VR environment (Task Computer C1), biological feedback equipment – labeled Biopac System, and another PC that is dedicated to acquiring signals from the biological feedback equipment (Biopac Computer C2). The Vizard software runs on computer C1 that is connected to the Biopac System via a parallel port to transmit task related event-markers. The physiological signals along with the event markers are acquired by the Biopac System and sent over an Ethernet link to the Biopac computer C2.

We video record the sessions to cross-reference observations made during the experiment. The signal from the video camera is routed to a television, and the signal from the participant's computer screen where the task is presented is routed to a

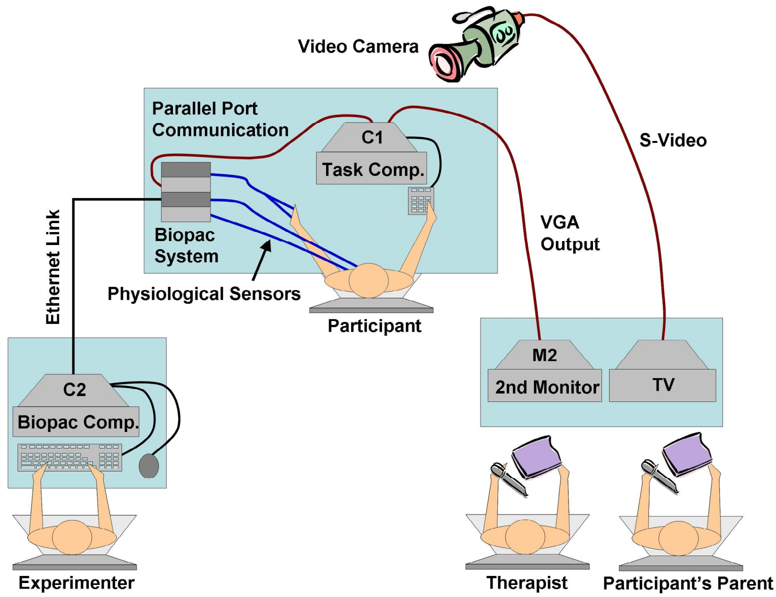

Fig. 4. Experimental setup for collecting physiological data and subjective reports

separate computer monitor (2nd Monitor, M2). The therapist and the participant's parent, seated at the back of the experiment room, will be fully exposed to the experiment by watching the participant on the TV from the view of the video camera and observing how the task progresses on the separate monitor.

Self-reports from each participant on their perceived affective states will be collected after each epoch via dialog windows on C1 and automatically stored in corresponding data-logger files. However, children with ASD may have deficits in identifying and describing their own emotions on a self-report [14]. Therefore, a therapist and a parent will observe the experiment and provide subjective reports about how they think the participant is feeling. Given the fact that a therapist's judgment based on his/her expertise is the state-of-the-art in most autism intervention approaches and the results about the reliability of the subjective reports in our completed studies [28], the reports from the therapist may provide the most reliable reference points to link the objective physiological data to each child's subjective affective states. Since the remarks of parents based on their every-day experience are also sought-after in the autism community, reports from each participant's parent will also be collected to compute any correlation with the therapist and child. The three reports regarding the intensity level (i.e., high/low) of the target affective states will be collected after each of the 24 social interaction epochs.

4 Measurement and Analysis

Due to the phenomena of person stereotypy [19], an individual-specific approach will be applied when creating the affective models. However, patterns of physiological response that may be related to presumed core impairments of ASD may be detectable. Post-experiment analysis will identify physiological reactions to the social communication parameters and any correlations to the subjective reports. These techniques have provided highly reliable results for typical adults [20] and children with ASD during performance-based computer and robot tasks [28]. We will further explore the discriminating capability of those features during social communication in a VR environment. For example, GSR is generally accepted as an indicator of anxiety. Therefore, if GSR increases significantly when an avatar uses direct eye contact, the child could likely have a communication deficit related to this social situation, and the subjective reports may reveal high anxiety.

5 Conclusions

The proposed design of integrating biofeedback sensor technology and VR social interaction tasks is novel, yet relevant to the current priorities of computer-assisted ASD intervention. This research is expected to result in the development of a physiological-based VR system for assessing physiological response during social interaction. Plans for user studies will make comparisons between typically-developing children and children with ASD to define social vulnerabilities within the interaction scenarios. The experimental design and methods could potentially produce a valuable tool for clinical application using this technology.

Acknowledgements. This work is supported by an Autism Speaks Pilot Study award.

References

1. Rogers, S.J.: Interventions That Facilitate Socialization in Children with Autism. Journal of Autism and Developmental Disorders 30(5), 399–409 (2000)
2. Parsons, S., Mitchell, P.: The potential of virtual reality in social skills training for people with autistic spectrum disorders. J. Intellect. Disabil. Res. 46(Pt 5), 430–443 (2002)
3. Bernard-Opitz, V., Sriram, N., Nakhoda-Sapuan, S.: Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. J. Autism. Dev. Disord. 31(4), 377–384 (2001)
4. Blocher, K., Picard, R.W.: Affective social quest: emotion recognition therapy for autistic children. In: Dautenhahn, K., Bond, A.H., Canamero, L., Edmonds, B. (eds.) Socially Intelligent Agents: Creating Relationships with Computers and Robots. Kluwer Academic Publishers, Dordrecht (2002)
5. Swettenham, J.: Can children with autism be taught to understand false belief using computers? Journal of child psychology and psychiatry, and allied disciplines 37(2) (1996)
6. Dautenhahn, K., Werry, I.: Towards interactive robots in autism therapy: background, motivation and challenges. Pragmatics & Cognition 12(1), 1–35 (2004)
7. Kozima, H., Nakagawa, C., Yasuda, Y.: Interactive robots for communication-care: A case-study in autism therapy. In: IEEE International Workshop on Robot and Human Interactive Communication, Nashville, TN, pp. 341–346 (2005)
8. Michaud, F., Theberge-Turmel, C.: Mobile robotic toys and autism. In: Dautenhahn, K., Bond, A.H., Canamero, L., Edmonds, B. (eds.) Socially Intelligent Agents: Creating Relationships with Computers and Robots, pp. 125–132. Kluwer Academic Publishers, Dordrecht (2002)

9. 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)
10. Strickland, D., Marcus, L.M., Mesibov, G.B., Hogan, K.: Brief report: two case studies using virtual reality as a learning tool for autistic children. J. Autism. Dev. Disord. 26(6), 651–659 (1996)
11. 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. John Wiley and Sons, Chichester (2007)
12. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
13. Seip, J.A.: Teaching the autistic and developmentally delayed: A guide for staff training and development, Delta, British Columbia (1996)
14. Hill, E., Berthoz, S., Frith, U.: Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. J. Autism. Dev. Disord. 34(2) (2004)
15. Ben Shalom, D., Mostofsky, S.H., Hazlett, R.L., Goldberg, M.C., Landa, R.J., Faran, Y., McLeod, D.R., Hoehn-Saric, R.: Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism. J. Autism. Dev. Disord. 36(3), 395–400 (2006)
16. Meehan, M., Razzaque, S., Insko, B., Whitton, M., Brooks Jr., F.P.: Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. Applied psychophysiology and biofeedback 30(3), 239–258 (2005)
17. Bradley, M.: Emotion and motivation. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G. (eds.) Handbook of Psychophysiology. Cambridge University Press, New York (2000)
18. Bethel, C., Salomon, K., Murphy, R., Burke, J.: Survey of psychophysiology measurements applied to human-robot interaction. In: IEEE International Symposium on Robot and Human Interactive Communication, Jeju, Korea (2007)
19. Lacey, J.I., Lacey, B.C.: Verification and extension of the principle of autonomic response-stereotypy. The American Journal of Psychology 71(1), 50–73 (1958)
20. Rani, P., Sims, J., Brackin, R., Sarkar, N.: Online stress detection using psychophysiological signal for implicit human-robot cooperation. Robotica 20(6), 673–686 (2002)
21. Gillott, A., Furniss, F., Walter, A.: Anxiety in high-functioning children with autism. Autism 5(3), 277–286 (2001)
22. Ruble, L.A., Robson, D.M.: Individual and environmental determinants of engagement in autism. Journal of Autism and Developmental Disorders 37(8) (2006)
23. Cobb, S.V.G., Nichols, S., Ramsey, A., Wilson, J.R.: Virtual reality-induced symptoms and effects. Presence 8, 169–186 (1999)
24. Bancroft, W.J.: Research in Nonverbal Communication and Its Relationship to Pedagogy and Suggestopedia. ERIC (1995)
25. Groden, J., Diller, A., Bausman, M., Velicer, W., Norman, G., Cautela, J.: The development of a stress survey schedule for persons with autism and other developmental disabilities. J. Autism and Dev. Disord. 31, 207–217 (2001)
26. Argyle, M., Deal, J.: Eye-contact, distance and Affiliation. Sociometry 28(3) (1965)
27. Schneiderman, M.H., Ewens, W.L.: The cognitive effects of spatial invasion. The Pacific Sociological Review 14(4), 469–486 (1971)
28. Conn, K., Liu, C., Sarkar, N., Stone, W., Warren, Z.: Towards affect-sensitive assistive intervention technologies for children with autism. In: Jimmy, O. (ed.) Affective Computing: Focus on Emotion Expression, Synthesis and Recognition, pp. 365–390. I-Tech (2008)