Eye Contact Conditioning in Autistic Children Using Virtual Reality Technology

Xi Wang[1(✉)], Nicholas Desalvo[1], Zhimin Gao[1], Xi Zhao[1], Dorothea C. Lerman[2], Omprakash Gnawali[1], and Weidong Shi[1]

[1] Department of Computer Science, University of Houston, Houston, TX 77004, USA
{xiwang,gnawali,larryshi}@cs.uh.edu, {nsdesalvo,zgao5,xzhao21}@uh.edu
[2] Center for Autism and Developmental Disabilities, University of Houston - Clear Lake, Houston, TX 77058, USA
lerman@uhcl.edu

Abstract. Children afflicted with developmental disabilities, namely autism, suffer from a natural aversion to dyadic (i.e., eye-to-eye) contact. Research has shown this aversion to be an early indicator of slower development of linguistic skills, a narrow vocabulary, as well as social issues later in life. In addition, this aversion may also result in the loss of already acquired abilities such as language and life skills. Consequently, manual prompt techniques have been adopted to address this issue. However, they are plagued with some inherent flaws: (i) the teacher must make unnatural movements when using a manual prompt such as gesturing towards the face; (ii) The child's attention will follow this prompt as it is removed from the face defeating the purpose as it detracts the child's attention from the teacher's eyes. To tackle these issues we have developed a system that can utilize effective prompt methodologies aimed at conditioning these children to establish and maintain dyadic contact. Our system not only reduces, but eliminates shortcomings present in the current manual method. This is accomplished through the use of a stereo camera and virtual reality headset to augment the child's vision when eye contact is not being established. The prompt is displayed in the child's vision over the eyes of the teacher to attract their attention. Once dyadic contact has been ascertained, the prompt is gradually fading leaving the child only to focus on the eyes of the teacher as is needed.

Keywords: Autism · Eye contact · Fading prompt · Virtual reality

1 Introduction

One of the earliest and most notable indicators of Autism Spectrum Disorder is a deficit in eye contact [1,2]. In early development, dyadic (i.e., eye-to-eye) contact directly relates to social interactions [5,6]. Research has suggested that eye contact serves an important social function for young children before vocal responses have begun to develop [3,4]. Dyadic contact has even been found to influence language acquisition skills [7].

P. Cipresso et al. (Eds.): MindCare 2014, LNICST 100, pp. 79–89, 2014.
DOI: 10.1007/978-3-319-11564-1_9

Previous studies suggest that diversity of prelinguistic pragmatic skills (e.g., eye contact and joint attention) act as a predictive element of subsequent vocabulary acquisition rates [8] which place these children at high risk of said effects. In addition, it has even been suggested that poor eye contact can negatively affect previous educational gains of children with autism. This is due to the direct relationship between dyadic contact and the ability to perceive and carry out teacher and instructional requests [9,10]. Since autism highly effects dyadic contact rates, it must be treated aggressively to limit the negative impact that it may have on other aspects of the child's life.

Current methods of teaching eye contact suggest using prompts. There are two types of prompts: A gesture prompt such as signaling towards the eye or putting a piece of a food that is of interest to the child [11]; and physical prompts such as guiding the child's head so that it is oriented towards the teacher. While effective in establishing dyadic contact there are some notable limitations in these approaches. One such limitation is that they are difficult to fade out or eliminate while continuing to hold the eye contact of the child. In addition, they are quite intrusive in that they interfere with natural social interactions. The problem in using a prompt without a way to fade out or unintrusively eliminate it, is that children with autism tend to exhibit stimulus overselectivity and inasmuch focus on the prompt itself rather than the teachers eyes [12,13]. In effect, once the prompt is physically removed there is a high probability that they follow the prompt and not focus on the eyes. Research suggests that when prompts are directly embedded in natural stimuli children perform better [14,15]. Hence, it is critical to increase their attention to aspects of the environment that normally command the response of the child if not affected by autism.

The system that we propose includes a novel solution which could overcome the intrusiveness of the prompt. The solution is an augmented reality system making use of a virtual reality headset with a stereo video feed. The child will wear the VR headset and see the natural world in the controlled treatment areas as to which they are accustomed to. The child and the teacher will interact with each other. When the child does not make eye contact and a prompt is needed, the teacher does not need to make unnatural actions such as a gesture towards the face or make use of food. The teacher need only press a key and the prompt will appear to grab the attention of the child facilitating eye contact. Then, the prompt will gradually fade away without the possibility that the child's attention will follow the prompt as if it were to be removed manually. In essence, the eye contact will remain after the prompt is no longer visible. This approach essentially holds the advantages of a prompt driven system in terms of facilitating eye contact but removes its disadvantages of being intrusive and the attention shift problem that the child will follow the prompt wherever it is moved.

2 Related Work

Virtual Reality technology applications have recently proliferated in autism therapy. The Cai research team implemented a Virtual Dolphinarium which allows

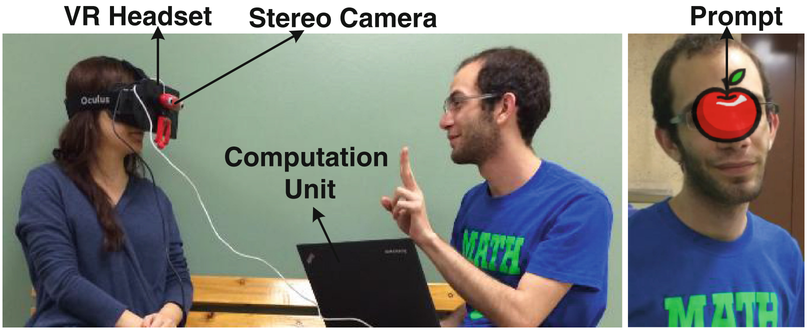


Fig. 1. System Overview: The stereo camera is connected to the computation unit through a USB 2.0 interface. This data, once processing and prompt placement (an apple in this case) has been applied, is then transferred to the VR Headset via an HDMI interface. Concurrently, required data, yet irrelevant in this context for VR Headset use is transferred back to the computation unit via another USB 2.0 interface. Note the VR Headset is connected bidirectionally with the computation unit.

Autistic children to interact with the virtual dolphins and to learn communication through hand gestures [18]. Developed by the University of North Carolina, the game "Astrojumper" allows autistic children to use their own physical movements to avoid virtual space-themed objects flying towards them. This game assists autistic children in developing dexterity and motor skills [16]. Researchers have also invented a VR social cognition mechanism aimed at training young adults with high-functioning autism. This was a substantial work in that it significantly increases social cognitive measures in theory, emotion recognition, as well as real life social and occupational functioning were found post-training [19]. While all of these aforementioned works contribute substantially to the treatment of autism in children, adolescents, and adults, none address the lack of eye contact in autistic children. Researchers from Vanderbilt University attempted to condition eye contact in Autistic children by creating a virtual storyteller to guide their focus on the communicator [17]. However, unlike our system, theirs did not adopt a fading prompt approach which suggests that there is still substantial gains to be made in autistic children's conditioning of eye contact in which we have seized the opportunity to address [20–23].

3 System Design

In this section, we present the hardware and software components of our system, as explained in Fig. 1.

3.1 Hardware

The proposed real-time system is only constrained by computational power. Inasmuch, it has been decided the best device to use would be a desktop system in that it will create a much more fluid experience for the user. The only required constraint inherent in this project is that the VR headset must not discomfort the user as to create a more immersive and beneficial experience where user acclimation time is minimized. The hardware system is comprised of three major entities: the computer, VR display headset, and stereo camera. Because both the stereo camera and the computer are standard they will not be addressed in much detail within this paper.

The VR headset utilized is an Oculus Rift. It is minimized in terms of weight and also contains foam in parts that come in contact with the user's face. The headset technology, as seen in Fig. 2, is comprised of an LCD (liquid crystal display) unit that is placed behind a lens (lens explained in depth in next paragraph) in front of the user's eyes. This LCD measures seven inches diagonally, 1280×800 pixel resolution (640×800 pixel resolution for each eye since the display is split between both eyes), and 32 bit color depth [30]. The reason for hardware choices are as follows: The screen must be large enough to accommodate for minor visual overlap between the images viewed in each eye as well as the peripheral vision of each separate eye. The cone of vision for humans is not the same for the right and left eyes, inasmuch we must compensate for this in the hardware. The FOV (field of view) related to the right eye extends further to the right than the left eye and vice versa for the left. In essence, images for the right and left eye must be different but overlap slightly for the brain to correctly stitch and render the images in 3D [31]. Even more importantly without significant user eyestrain. This amount of overlap between the right and left visual input is controlled by the respective subjects interpupillary distance (see Fig. 2). The users interpupillary distance and the amount of overlap are inversely proportional, so a larger interpupillary distance would create a smaller overlap and vice versa. When the images are not correctly overlapped, besides causing eye strain and discomfort for the user it will also detract from the benefits the user can achieve from using this device making efficacy quite less, so this is a significant issue that must be addressed correctly [28,29].

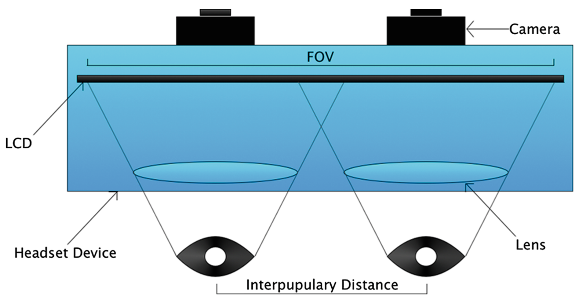


Fig. 2. Inside, top-down view of the VR Headset. As you can see distance between pupils is the interpupillary distance. Inside the device is both the LCD and Lenses. While attached on the outside of the device is the stereo camera. Please also note the FOV is the total cone of vision that the user can see which is expressed in degrees.

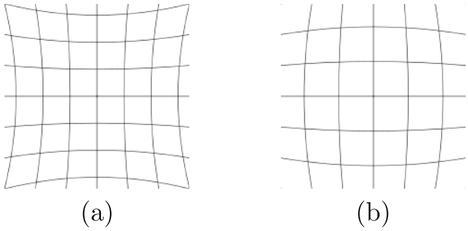


Fig. 3. (a): Pincushion distortion example; (b) Barrel distortion example

Now that this has been accomplished, a distortion by both lenses in front of the eyes as well as the video feed is imbued with a distortion to create a sense of depth perception for the user. The lenses in front of the user's eyes create a pincushion distortion as seen in Fig. 3(a), while the video feed placed in front of the user on the LCD has a barrel distortion applied as seen in Fig. 3(b). When these two distortions are used in conjunction with each other they will effectively cancel each other out in the perception of the user. However what the user does not notice is that the pincushion distortion creates a wider FOV so that for example when the user looks 40° the light is bent such that they see 30° to the left of the LCD panel. This is how the VR headset creates a more realistic experience.

OpenGL [24] and OpenCV [25] are both used in this software (greater scrutiny will be provided within the software section following). To offload some computations and speed up the process overall, it has been determined that the barrel distortion will be done in OpenGL in Preprocessing. This will migrate these actions from the CPU on to the GPU and create a more fluid experience within present real-time constraints. This also allows for a less powerful system to be used in settings that may otherwise be inhibited by this variable.

3.2 Software

As seen in Fig. 4, the software consists of two main components: (1) a classifier training and configuration settings in the offline phase, and (2) image processing including prompt overlay in the online phase. The two parts are explained as follows:

Offline. The system adopts an object detection algorithm using Haar feature-based cascade classifiers [26]. In order to increase eye detection accuracy, the system detects the face from the image prior to detection of the eyes, which limits the search to the area detected as the face. The teacher can use default classifiers for both face and eye detection which are supplied by OpenCV, instead of training a new classifier.

The teacher may customize two settings which are the prompt image and its respective opacity. This allows the teacher to select one image out of the provided image bank that most attracts the interest and eye contact of the autistic child to

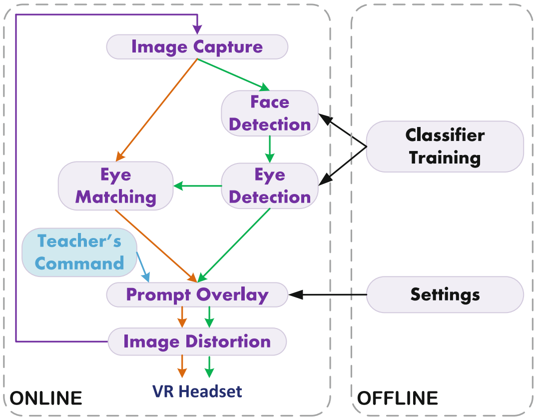


Fig. 4. The software consists of two parts: online and offline. The online phase includes two threads, the command thread (acting as a user interface) and the image processing thread. The green and orange arrows represent processing of both left and right images from the stereo camera. The processed images are later fed into the VR headset.

be overlain on the video feed. The provided images include an apple, orange, and flower. The teacher may find that none of these images are to the liking of the child. This case would render the system ineffective. To cope with this exception, the teacher may also upload an image not provided to be used as a prompt. The other adjustable setting is opacity, which translates to the transparency level of the overlain prompt. Opacity is expressed as a percentage in this context: 0 % being completely transparent (invisible) while 100 % is completely visible. The prompt gradually disappears at a constant rate. Every 100 ms, the opacity level will decrease by X %, X being an integer value as defined by the teacher. While the system is running, the teacher may decide when the overlain prompt should be faded by clicking the fade button. This triggers the prompt to fade away at the predefined rate. To clarify, the teacher may also choose default settings for the two aforementioned variables. The default prompt is set to be a red apple and the default fading rate is two. This means the prompt will fade at a rate of 2 % per 100 ms rendering the prompt completely transparent in five seconds.

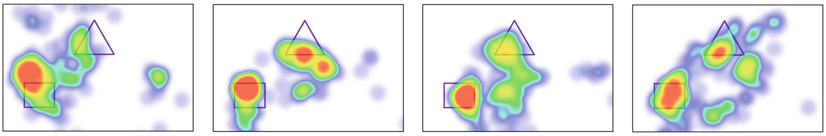
Online. During the online phase, the system is streamlined by utilizing two threads running in parallel (the command and image processing threads). The command thread waits for the teacher's prompt display command. If the teacher would like to start fading the prompt or overlay a 100 % opaque prompt again, she/he may click the fade or reset buttons respectively. A button click will trigger the command thread to send the teacher's request to the image processing thread where the commands are executed. The two 2D cameras send images continuously to the image processing thread. In order to form a proper 3D image, the eye areas of each image must be synced as accurately as possible. The system does not conduct autonomous eye detection inept of the other image fed at the same time (right and left images). It locates the eye area from one image and uses that data as template [27] to search the eye area of the other. Due to the small offset between the two images, when they are presented in front of a person, the brain renders the pictures and stitches them together accordingly in order to be perceived as a 3D image. The two prompts overlain on each image will then be merged into one prompt when the image is perceived by the user. In order to ensure highly accurate eye detection, the system detects the face first and then later detects eyes based upon the previously detected face area. Both face and eye detections need cascaded classifiers obtained from the offline phase. If no eyes are detected, the system continues detection for the following images. According to the teacher's requests, the system adjusts the opacity of the prompt, which could be from 0 % to 100 %, and then overlays the prompt on the two images next to the eyes. Before the two images are sent to the AR headset, a distortion is required to compensate for the VR headset. To accomplish this, OpenGL shaders are utilized. Because of constraints of the application as well as OpenGL itself, a vertex, geometric, and fragment shader are all required. All three of these shaders are loaded, compiled, and linked in one OpenGL program object. In this rendering process the program object is used on each image creating the distorted before being sent to the VR headset. The user will then physically see the image presented as two undistorted images, one in the right eye and one in the left eye. The user's brain then stitches these two 2D pictures together in order to perceive a 3D image.

4 Study

4.1 Setup

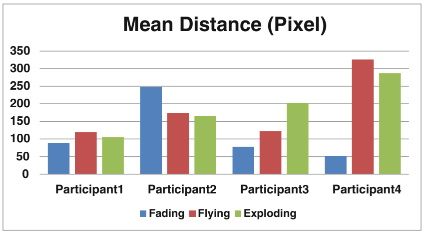
We made four black-background videos, each of which contains three objects: (1) A target which is a stationary non-filled, red-bordered triangle in the upper middle section, (2) A reference which is a stationary video window on the bottom left displaying an instruction on mechanical assembly, and (3) A prompt which is a moving window displaying a movie trailer. The target and reference stay visible through each video. In the first video, which lasts 10 s, the prompt appears in bottom right at six seconds and slowly moves towards the target. Once on the target it is stationary. In the rest of three movies, which all last 25 s, the prompt appears from bottom right corner, slowly moves towards the target, and at 21 s it disappears. The prompts in the three videos disappears in three different ways: (1) It stays on the target and fades out over time, (2) As is current conventions when used in eye contact training for autistic children, the prompt moves away from the target and out of the video at upper left corner, and (3) It stays on the target and explodes. After the prompt is no longer in the video, the reference and the target remain for five more seconds.

We designed a user study to answer two fundamental questions: (1) Whether the virtual moving prompt draws participant's attention when he/she focuses on the reference and not the target, and (2) Which of the three prompts holds the participant's eye contact the longest.

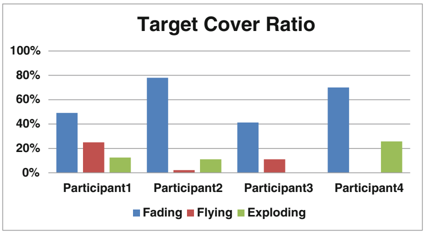


(a) Participant 1 (b) Participant 2 (c) Participant 3 (d) Participant 4

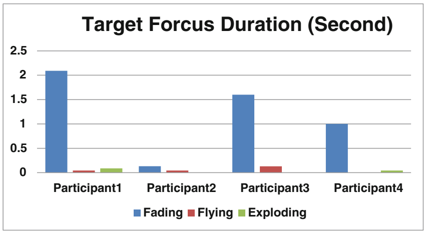
Fig. 5. Heatmaps of four participants watching the first video. The rectangle and triangle represent the positions of the reference and the target respectively.



(a)



(b)



(c)

Fig. 6. The three figures demonstrate the effectiveness of three prompts after the disappear beings. They depict Mean distance between gaze and the target center, Ratio of gaze within target, and duration of participants' focus on the target in respect.

We invited four participants (college students) and used Opengazer [32] to record participants' gaze positions in the video. In order to ensure the tests were unaffected by extrinsic factors the participants were instructed to rate the movie trailers and were not informed of why calibrations were needed prior to the four videos as well as the fact that their eyes were tracked in the process of the experiment.

4.2 Results

We used the first video to test whether the virtual prompt is able to divert the participants' attention from the reference. Respectively four participants started focusing on the prompt 2.2, 0.8, 1.3, 1.5 s after the prompt appeared in the video. Heatmaps show the cumulative intensity with which a participant

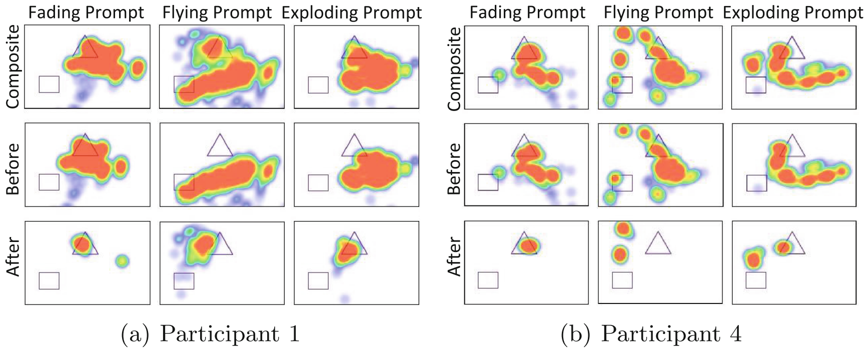


Fig. 7. Heatmaps of participant 1 and 4 using three different prompts. For each prompt, there are three heatmaps, one before the prompt disappears, one after, and one covering the entire trial.

viewed different parts of the video. As seen in Fig. 5, participants' focus was led to the target by the prompt.

We used the other three videos to evaluate how long the prompt can hold participants' eye contact point (gaze) at the target after the prompt begins to disappear. Mean distance between the gaze and target center represents how close a participant's gaze is to the target. In the Fig. 6(a), three participants have smaller mean distance with fading prompts than others. The mean distance is on average 25.5 % and 27.3 % smaller than when using flying and exploding prompts. Figure 6(b) demonstrates how many focal points fall within target. Fading prompts caught 16 and 3 times larger ratio over flying and exploding prompts. Figure 6(c) shows fading prompt is able to keep participants' attention 19 and 20 times longer than the other two. Figure 7 depicts, after the prompt begins to disappear, the participants' gaze concentrated on the target more with fading prompt, and the participants followed flying prompt as it was removed from target. In summary, fading prompt performed best among three prompts.

5 Conclusion and Future Work

We have proposed a novel system, which adopts VR technology to train children with autism to establish eye contact via a fading prompt approach. We conducted a study to demonstrate that a virtual prompt is able to draw user's attention to the target and the fading prompt is more effective than the traditional flying prompt and exploding prompt. Future plans include expanding the current system so that it is tolerant to scenarios in which multiple people are present in the child's view. Inasmuch, the system will be able to identify the individual the child is interacting with and display a prompt as is necessary.

Acknowledgments. We would like to thank Dr. Katherine Loveland for helpful discussions and equipment, and Tzu-Hua Liu for giving us helpful pointers to work related to this research.

References

1. Wimpory, D.C., Hobson, R.P., Williams, M.G., Nash, S.: Are infants with autism socially engaged? A study of recent retrospective parental reports. J. Autism Dev. Disord. **30**, 525–536 (2000)
2. Woods, J.J., Wetherby, A.M.: Early identification of and intervention for infants and toddlers who are at risk for autism spectrum disorder. Lang. Speech Hear. Serv. Schools **34**, 180–193 (2003)
3. Mirenda, P., Donnellan, A., Yoder, D.: Gaze behavior: a new look at an old problem. J. Autism Dev. Disord. **13**, 397–409 (1983)
4. Stern, D.: The Interpersonal World of the Infant. Basic Books, New York (1985)
5. Tiegerman, E., Primavera, L.H.: Imitating the autistic child: facilitating communicative gaze behavior. J. Autism Dev. Disord. **14**, 27–38 (1984)
6. Lee, K., Eskritt, M., Symons, L.A., Muir, D.: Childrens use of triadic eye gaze information for 'mind reading'. Dev. Psychol. **34**, 525–539 (1998)
7. Podrouzek, W., Furrow, D.: Preschoolers' use of eye contact while speaking: the influence of sex, age, and conversational partner. J. Psycholinguist. Res. **17**, 89–98 (1988)
8. Kleinke, C.L.: Gaze and eye contact: a research review. Psychol. Bull. **100**(1), 78–100 (1986)
9. Greer, D.R., Ross, D.E.: Verbal Behavior Analysis. Pearson Education, New York (2007)
10. Lovaas, O.I.: The Autistic Child: Language Development through Behavior Modification. Irvington, New York (1977)
11. Lovaas, O.I.: Teaching Developmentally Disabled Children: The Me Book. University Park Press, Baltimore (1981)
12. Dube, W.V., McIlvane, W.J.: Reduction of stimulus selectivity with nonverbal differential observing responses. J. Appl. Behav. Anal. **32**, 25–34 (1999)
13. Lovaas, O.I., Schreibman, L.: Stimulus overselectivity of autistic children in a two-stimulus situation. Behav. Res. Ther. **9**, 305–310 (1971)
14. Smeets, P.M., Lancioni, G.E., Striefel, S.: Discrimination training through time delay of multistimulus prompts: the shapes and locations of the prompts. Psychol. Rec. **37**, 507–521 (1987)
15. Zygmont, D.M., Lazar, R.M., Dube, W.V., McIlvane, W.J.: Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals: a methodological note. J. Exp. Anal. Behav. **57**, 109–117 (1992)
16. Finkelstein, S.L., Nickel, A., Barnes, T., Suma, E.A.: Astrojumper: designing a virtual reality exergame to motivate children with autism to exercise. 2010 IEEE Presented at the Virtual Reality Conference (VR), pp. 267–268 (2010)
17. Lahiri, U., Warren, Z., Sarkar, N.: Design of a gaze-sensitive virtual social interactive system for children with autism. IEEE Trans. Neural Syst. Rehabil. Eng. **19**(4), 443–452 (2011)
18. Cai, Y., Chia, N.K.H., Thalmann, D., Kee, N.K.N., Zheng, J., Thalmann, N.M.: Design and development of a virtual dolphinarium for children with autism. IEEE Trans. Neural Syst. Rehabil. Eng. **21**(2), 208–217 (2013)

19. Kandalaft, M., Didehbani, N., Krawczyk, D., Allen, T., Chapman, S.: Virtual reality social cognition training for young adults with high-functioning autism. J. Autism Dev. Disord. **43**, 34–44 (2013)
20. Ferster, C.B., Nurnberger, J.I., Levitt, E.B.: The control of eating. Obes. Res. Clin. Pract. **4**, 401–410 (1962)
21. Lovaas, O.I., Berberich, J.P., Perloff, B.F., Schaeffer, B.: Acquisition of imitative speech by schizophrenic children. Science **151**, 705–707 (1966)
22. Lovaas, O.I., Freitas, L., Nelson, K., Whalen, C.: The establishment of imitation and its use for the development of complex behavior in schizophrenic children. Behav. Res. Ther. **5**, 171–181 (1967)
23. Metz, J.R.: Conditioning generalized imitation in autistic children. J. Exp. Child Psychol. **4**, 389–399 (1965)
24. OpenGL (Open Graphics Library). http://www.opengl.org/
25. OpenCV (Open Source Computer Vision). http://opencv.org/
26. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
27. Brunelli, R.: Template Matching Techniques in Computer Vision: Theory and Practice. Wiley, Chichester (2009)
28. Lambooij, M.T.M., IJsselsteijn, W.A., Heynderickx, I.: Visual discomfort in stereoscopic displays: a review. Electronic 64900I–64900I-13 (2007)
29. Kooi, F.L., Toet, A.: Visual comfort of binocular and 3D displays. Displays **25**, 99–108 (2004)
30. Oculus Rift. http://www.oculusvr.com/
31. Lacrama, D.L., Fera, D.: Virtual Reality (2009)
32. Opengazer. http://www.inference.phy.cam.ac.uk/opengazer/