The SARA Project: An Interactive Sandbox for Research on Autism

Diana Arellano[*]
Filmakademie
Baden-Württemberg
Ludwigsburg, Germany

Volker Helzle
Filmakademie
Baden-Württemberg
Ludwigsburg, Germany

Ulrich Max Schaller
University Medical Center
Freiburg, Germany

Reinhold Rauh
University Medical Center
Freiburg, Germany

Marc Spicker
University of Konstanz
Konstanz, Germany

Oliver Deussen
University of Konstanz
Konstanz, Germany

ABSTRACT
SARA is an ongoing research project that investigates in a novel and artistic way the causes for social communication and emotion recognition deficits in children and adolescents with high-functioning autism spectrum disorders (ASD). The novelty of our work resides in the real-time generation and parameterization of emotional facial expressions of virtual characters by means of speed, intensity and abstraction, the latter achieved by non-photorealistic rendering (NPR) techniques. Although the project is currently in an ongoing phase, it shows the potential of using virtual characters and real-time techniques for interactive experiments, which otherwise would be impossible using "linear stimuli" (e.g. pre-rendered animations).

Categories and Subject Descriptors
H.5 [**Information Interfaces and Presentation**]: Multimedia Information Systems; I.3 [**Computer Graphics**]: Applications; J.4 [**Social and Behavioral Sciences**]: [Psychology]

General Terms
Autism, Social Skills, Animation, Emotions

Keywords
Autism Spectrum Disorder, Non-Photorealistic Rendering, Real-time Animation, Facial Expressions, Interaction

1. INTRODUCTION
Autism spectrum disorder (ASD) is a developmental disability that tosses significant communication, behavioral and social challenges. According to the Center for Disease Control and Prevention(CDC), research on ASD has increased a great deal in recent years [1], as well as the number of children and adults with this disorder. In 2014 the CDC estimated that about one percent of the world population has ASD. In the United States the prevalence of ASD was in the same year one in 68 births [5]. In Spain, the "Confederación Autismo España" communicated that approximately 350.000 people in this country has a form of ASD [8].

Motivated by these numbers and the amount of research done in this area, we propose SARA (Stylized Animations for Research in Autism), an ongoing project that investigates the causes behind communication and emotion perception deficits in children and adolescents with high-functioning ASD. To achieve this, SARA combines psychology, real-time non-photorealistic rendering (NPR) and 3D computer animation, with the main goal of studying how abstracted faces, with different levels of details are categorized by children and adolescents with ASD.

The interest in using NPR is that it permits a variation in the level of abstraction and visuo-spatial information, adapting images to "focus the viewer's attention" [6]. Thus the information load in the characters' facial expressions can be reduced, conveying the emotional information more efficiently. In the end, a set of virtual characters are used in an interactive computer-based psychological test, where each character displays emotional facial animations generated in real-time. In the following, we will shortly refer to previous research in this area, we will explain the current status and first results of SARA, and finally, we will conclude with the ongoing and future work.

2. RELATED WORK
Many interactive applications created to develop or enhance the social skills of individuals with autism make use of virtual characters. Among the tools used for research in a lab environment we found the work of Whyte et al. [14], who used game components (e.g., storyline, long-term goals, rewards) to create engaging learning experiences, especially in computer-based interventions. Milne et al. [9] employed autonomous agents as social skills tutors for teaching children with ASD conversation skills and how to deal with bully-

---

[*]diana.arellano@filmakademie.de

ing. Grawemeyer et al. [7] developed an embodied pedagogical agent together with, and for young people with ASD. ECHOES VE [2] presents a virtual environment where children with ASD need to assist a virtual character in selecting objects by following the character's gaze and/or pointing at the object. In JeStiMulE [12] participants are taught to recognize emotions on the faces and gestures of virtual characters, while considering the context. LIFEisGAME [3] deploys a low cost real-time animation system embedded in a game engine to create a game that helps individuals with ASD to recognize emotions in an interactive way. FaceSay[TM] [11] aids children with ASD to recognize faces, facial expressions and emotions by offering students simulated practice with eye gaze, joint attention, and facial recognition skills. *Let's face it!* [13] is a program comprised of seven interactive computer games that target the specific face impairments associated with autism.

One thing all these applications have in common is the use of virtual characters to enhance or develop skills in subjects with ASD. In all the previous cases, the characters present a defined visual style, which can be either cartoony or realistic. In this sense, one of the assets of our project is the possibility to change the visual representation of the characters (from more realistic to more abstract) in real-time, opening up new possibilities for more personalized applications.

3. SARA

Similarly to previous research, SARA (Stylized Animation for Research in Autism) combines clinical psychology and computer animation to create a tool for assessing the categorization of dynamic emotional facial expressions by children and adolescents with high-functioning ASD. Moreover, our project's innovation includes real-time non-photorealistic rendering (NPR) algorithms to abstract the faces of the virtual characters used in the test. This will allow us to explore how a reduction in the level of details of facial expressions affect their categorization by individuals with ASD.

The core of SARA is the DECT (Dynamic Emotion Categorization Test) [10], an interactive computer-based tool created to assess the feasibility of using real-time animations by comparing virtual characters to video clips of human actors. A previous version of the test contained material of two human actors and two virtual characters displaying dynamic facial expressions of the basic emotions: anger, disgust, fear, happiness, sadness, and surprise, on three intensity levels: weak, medium, and strong. The results of this very first version of DECT showed that the three levels of intensity were equally categorized in the virtual characters and in the human actors. This motivated us to continue exploring the use of virtual characters in autism research.

The design and development of the DECT, as well as the implementation of the NPR algorithms and real-time animations is being done in the software development platform Frapper[1], in particular using the Agent Framework [4]. Frapper is a C++, Ogre3D and Qt based development environment consisting of a node-based scene model, a model-view-controller architecture, a panel-oriented user interface similar to commercial 3D packages. The Agent

[1]http://sourceforge.net/projects/frapper/

Framework is the set of functionalities (nodes and plugins) that allow users the rapid prototyping of applications that make use of virtual characters. Both Frapper and the Agent Framework are provided with two human-like characters distributed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

4. INTERACTING WITH DECT

DECT has been conceived as an interactive research tool, rather than as an intervention tool, where both experimenters and participants can work in a real-time environment. Currently, the interaction with the software has been designed in a way where the experimenter has more control over it. It is the experimenter who explains the child or adolescent what it needs to be done, and even aids him in the selection of the emotional choice, as seen in Figure 1.

Figure 1: Participant during a DECT session.

The test deliberately does not contain any GUI-centric terms, so participants with ASD do not focus on other elements than the facial expressions of the characters and the emotional answers. A session with the DECT consists of several trials. Figure 2 shows the basic screens that are sequentially displayed in each trial. It begins with a pink-colored screen (Fig. 2(0)), which serves as separator between trials. By pressing the Enter key the trial and the interaction is initiated. It begins with a fixation cross (Fig. 2(1)) that appears for 0.5 seconds and indicates where the participant should fixate his gaze. After this time, one of two characters (Hank, an old male; or Nikita, a young female) appear displaying a real-time generated animation of a facial expression with certain intensity (weak, medium, strong) and speed (very slow, slow, moderately slow, normal, moderately fast, fast, very fast) (Fig. 2(2)). Then, a screen with white noise ((Fig. 2(3)) appears for 0,5 seconds, loading participants' iconic memory with task-irrelevant information. Finally, a screen with the answers options represented by emotional labels is shown (Fig. 2(4)). Here the user needs to select the one corresponding to the expression that he just saw. In order to select an emotion, each of the basic emotions were mapped to a number between 1-3 and 7-9, which was then selected using the numeric pad of the keyboard. The reason for not using the row 4-6 was to allow space between the fingers and avoid experimental errors by inadvertently pressing the wrong key. The pairing emotion-number is done randomly each time the test is carried out.

For the SARA project, three versions of the DECT have been planned. The first version, R-DECT was implemented and executed in the context of a pilot study to assess "Rapid Social Cognition" of children and adolescents with ASD. The
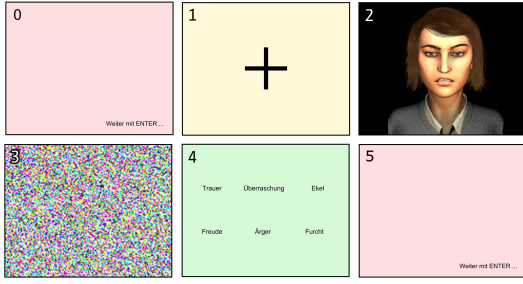
Figure 2: Screens of the DECT: (0) Initial screen, (1) Fixation Cross, (2) Character with an angry expression, (3) White noise, (4) Forced choices with emotion names, (5) Initial screen for next trial.

innovation of this test was the random presentation of real-time animations of facial expressions with different speeds (going from normal to very fast) and intensities (weak, medium, strong). The R-DECT served also to validate the improved facial animation from the first DECT, as well as to validate the test itself as a tool for the interactive categorization of emotional expressions.

The second version called NPR-DECT comprises one of the novelties of our project: the use of NPR algorithms to abstract and manipulate visuo-spatial information in the faces of our virtual characters. It not only constitutes a way to reduce information load in the characters' facial expressions, but also a way to include more artistic approaches to investigate how these abstractions affect the recognition of the facial expressions of emotions, in comparison to their more realistic representations. Figure 3 shows different facial abstractions to be tested in the NPR-DECT.

Figure 3: NPR styles. Left to right: original, pencil drawing, watercolors, line drawing, loose & sketchy.

The third version called i-DECT is the one with the highest interactivity, stimulating a visual interaction between the participant and the virtual character. This test will study the differences in eye contact and mutual gaze between neurotypical subjects and subjects with ASD. It is worth mentioning that the interactivity to be achieved in this test is possible thanks to the real-time characteristic of our framework, providing flexibility and opportunity for more elaborated interactive experiments.

Another level of interaction that plays a main role is the one of the experimenter with the test. Having a tool that generates animations and visual representations in real-time allows the psychologists and experimenters to fine tune and parameterize the tests themselves according to their requirements, or the participant's needs. This flexibility makes them independent from the animator, an important aspect to consider when using a computer-based research tool.

5. EVALUATIONS

Until now we have evaluated the R-DECT, and partially the NPR-DECT with neurotypical participants.

5.1 R-DECT

Participants of the experiment were 39 adolescents with ages between 14.0 and 17.9 years and IQ $\geq$ 70. The group of neurotypically developed adolescents (NTD group: n=22) consisted of 18 males and 4 females. The group of individuals with high-functioning ASD (ASD group: n=17) consisted of 12 males and 5 females. The R-DECT consisted of 2 (characters) x 6 (basic emotions) x 3 (intensity levels), resulting in 36 animations. Regarding the speed variable, it was assigned according to a certain scheme to each of the 36 animations, ranging from 1 (normal speed) up to 2.25 times of normal speed. In total, six levels were used (1.00, 1.25, 1.50, 1.75, 2.00, 2.25).

In total, 62.2% of the animations were categorized correctly. Accuracy rate for the NTD group was 65.7 % whereas for the ASD group was 57.7%. A 2x6 MANOVA with repeated measurements showed no significant interaction between group and basic emotion($F < 1$). However, the two main effects were significant (*basic emotion*: $F(5,33) = 77.63$, p < .0001; *group*: $F(1,37) = 5.36$, p = .026), indicating that the ASD group performed significantly worse than the NTD group (Figure 4). The order of accuracy for basic emotions was the same for both groups, being "happiness" the one recognized with the highest accuracy and "fear" with the least. However, all post-hoc comparisons did not reach statistical significance. As for the intensity of facial emotions, we only considered the NTD group where typical facial emotion recognition is expected. The results showed that in general varying intensities from weak over medium to strong affected accuracy rates correspondingly: Weak: 59.5%, Medium: 65.5% and Strong: 72.0%.
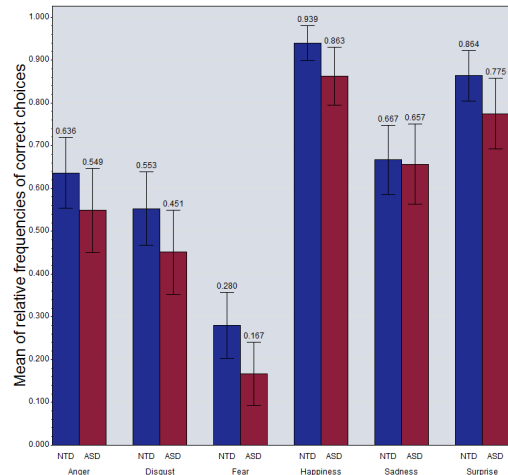

Figure 4: Mean of relative frequencies of correct categorizations of six basic emotions in ASD and NTD groups. Error bars represent 95% CI.

5.2 NPR-DECT

Participants of the experiment were 31 (9 male, 22 female) neurotypically developed psychology students with an age

range from 20 to 35 years. The NPR-DECT comprised 2 (characters) x 6 (basic emotions) x 13 (1 + 12 NPR style x abstraction combinations) = 156 trials, which were presented in a pseudorandomized order. The NPR styles: coherent line drawing (*CLD*), pencil drawing (*PD*), image abstraction (*IA*) and watercolors (*W*) were instantiated in one of three levels of abstraction (low, medium, high). During the whole NPR-DECT session participants' gaze was tracked by an RED-250 eye tracker (SMI), which API was integrated in Frapper to allow the communication between each other.

In total, 71.4% of the emotional expressions were correctly categorized. However, we found a considerable difference of accuracy between both characters (Hank: 66.5% vs. Nikita 76.3%). Differences in levels of accuracy seem to be dependent on the character in a certain style with a certain level of abstraction. The lowest percentage for Hank was obtained with the *CLD - high abstraction* (55.9%), whereas the highest percentage was with *W - medium abstraction* (73.7%). The lowest percentage for Nikita was obtained with the *PD - high abstraction* (71.5%), whereas the highest percentage was noted for both *IA - low abstraction* (81.7%).

As for the recognizability and likeability of the abstracted faces, they were measured through a computer-based questionnaire with 26 images of the two characters, stylized by the four NPR styles in three levels of abstraction, plus the original photorealistic representation. For each image, two questions were posed: (1) How good were you able to recognize the emotions from this representation? (2) How good did you like this representation? For each question, the answers were presented in a 7-point Likert scale ranging from very good ("1") to very bad ("7"). On average, recognition ratings were best for the original representation (mean = 2.02), followed by *IA* (mean = 2.44), *CLD* (mean = 3.69), *W* (mean = 3.73), and finally *PD* (mean = 4.17).

6. CONCLUSIONS

We have presented the ongoing research project SARA and the tests that have been implemented so far to assess the categorization of emotional facial expressions: R-DECT and NPR-DECT. In their current status, both are more a research tool than an intervention tool. Therefore, the interaction with the tool has not been exploited to its maximum. Results have shown that there is a significant difference in the categorization of emotions between the NTD and ASD groups, being the latter the one that performed the worst. However, it was not possible to say which emotion(s) contributed to this general difference. As for the NPR styles, no decisive conclusions were achieved. At the end of this project it is our goal to distribute all DECTs as a open-source tool. Moreover, based on the results obtained with SARA we will create new interactive applications or tests considering HCI elements, taking advantage of the NPR elements and artistic abstraction techniques.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Facts About ASD, 2014. Retrived April 06, 2015, from Center for Disease Control and Prevention: http://www.cdc.gov/ncbddd/autism/facts.html.

[2] A. Alcorn, H. Pain, G. Rajendran, T. Smith, O. Lemon, K. Porayska-Pomsta, M. E. Foster, K. Avramides, C. Frauenberger, and S. Bernardini. Social communication between virtual characters and children with autism. In *AIED*, 2011.

[3] S. Alves, A. Marques, C. Queirós, and V. Orvalho. LIFEisGAME prototype: A serious game about emotions for children with autism spectrum disorders. *PsychNology Journal*, 11(3):191–211, 2013.

[4] D. Arellano, V. Helzle, U. Schaller, and R. Rauh. Animated faces, abstractions and autism. In *Intelligent Virtual Agents 2014*, volume 8637 of *LNCS*, pages 22–25, 2014.

[5] J. Baio. Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, united states, 2010. In *Surveillance Summaries*, volume 63, pages 1–21. 2014.

[6] B. Gooch and A. Gooch. *Non-Photorealistic Rendering*. A.K. Peters, 2001.

[7] B. Grawemeyer, H. Johnson, M. Brosnan, E. Ashwin, and L. Benton. Developing an embodied pedagogical agent with and for young people with autism spectrum disorder. In *ITS 2012*, volume 7315 of *LNCS*, pages 262–267. Springer, Heidelberg, 2012.

[8] J. C. Hidalgo. Ciudades en azul en el día del autismo, April 2015. El País. Retrieved from http://goo.gl/Yzqdw6.

[9] M. Milne, M. Luerssen, T. Lewis, R. Leibbrandt, and D. Powers. Designing and evaluating interactive agents as social skills tutors for children with autism spectrum disorder. *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*, pages 23–48, 2011.

[10] R. Rauh and U. Schaller. Categorical perception of emotional facial expressions in video clips with natural and artificial actors: A pilot study. Technical report, University of Freiburg, 2009.

[11] L. Rice, C. Wall, A. Fogel, and F. Shic. Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with asd. *Journal of Autism and Developmental Disorders*, pages 1–11, 2015.

[12] S. Serret, S. Hun, G. Iakimova, J. Lozada, M. Anastassova, A. Santos, S. Vesperini, and F. Askenazy. Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new serious game: a pilot study. *Molecular Autism*, 5(37), 2014. http://www.jestimule.com.

[13] J. W. Tanaka, J. M. Wolf, C. Klaiman, K. Koenig, J. Cockburn, L. Herlihy, C. Brown, S. Stahl, M. D. Kaiser, and R. T. Schultz. Using computerized games to teach face recognition skills to children with autism spectrum disorder: the let's face it! program. *J Child Psychol Psychiatry*, 51(8):944–952, 2010.

[14] E. Whyte, J. Smyth, and S. Scherf. Designing serious game interventions for individuals with autism. *Journal of Autism and Developmental Disorders*, 2014.