Noname manuscript No.
(will be inserted by the editor)

Interactive Testbed for Research in Autism - The SARA Project.

Diana Arellano* · Reinhold Rauh* · Benjamin Krautheim · Marc Spicker · Ulrich Max Schaller · Volker Helzle · Oliver Deussen

Received: date / Accepted: date

Abstract The project SARA (Stylized Animations for Research on Autism) aims for a better understanding of the cognitive processes behind emotional categorization in children and adolescents with high-functioning autism spectrum disorder (ASD), in comparison to their neurotypically developed (NTD) peers. To this end, we combine novel real-time non-photorealistic rendering (NPR) algorithms, emotional facial animations, and eye tracking technologies in a framework that serves as an interactive testbed for empirical research. In this paper we focus on three experiments that: (1) validate real-time facial animations of virtual characters, (2) evaluate the NPR algorithms to create abstracted facial expressions, and (3) elucidate the relation between eye gaze behavior, ASD and alexithymia (i.e. difficulties in expressing ones emotions). The results show that our animations indeed can be used in the proposed experiments; however, more evaluation is needed regarding the NPR abstractions, especially with individuals with ASD. Finally, even though no correlation was found between gaze behavior, ASD and alexithymia, the study opened several questions that will be addressed in our future work.

D. Arellano · V. Helzle
Institute of Animation, Visual Effects and Digital Postproduction at Filmakademie Baden-Wuerttemberg, Germany
E-mail: diana.arellano@filmakademie.de

B. Krautheim · U.M. Schaller · R. Rauh
Dept. of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics; Medical Center University of Freiburg; Faculty of Medicine; University of Freiburg, Germany

M. Spicker · O. Deussen
Chair of Computer Graphics and Media Informatics, University of Konstanz, Germany

(*) These authors contributed equally to this work.

Keywords Autism Spectrum Disorder · Facial Animation · Eye tracking · Stylization

1 Introduction

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 [2], 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 had ASD. In the United States the prevalence of ASD was in the same year one in 68 births [8]. Currently the prevalence rate in Europe indicates that 1 in 100 people has this disorder [15] [31].

Motivated by these numbers and the amount of research done in this area [38] [12], we proposed SARA, a project that combines clinical psychology, NPR, 3D facial animation and eye tracking technology. The main objective is to investigate the causes behind communication and emotion perception deficits in children and adolescents with high-functioning ASD.

In this paper we present a compilation of three experiments carried out during the course of SARA. In all them a set of virtual characters displaying emotional facial animations are used in an interactive computer-based psychological test. All animations are generated in real-time. The first experiment, R-DECT, assesses one important faces of "rapid social cognition" of children and adolescents with ASD.

The second experiment, NPR-DECT, served as a pilot study in order to appraise how NTD individuals categorize abstracted faces with reduced levels of detail.

The third experiment is part of the NPR-DECT, but focused on determining the differences in eye gaze behavior between NTD and ASD participants. More specifically, we want to better ascertain the relation between ASD, gaze behavior, alexithymia and deficits in facial emotion perception. In this study, we only evaluated gaze behavior during the interaction with the realistic looking versions of the characters. The remaining NPR-DECT gaze data (i.e. when interacting with the abstracted characters) will be further analyzed but does not belong to the scope of this paper.

One of the main interests of our project is to better understand the cognitive processes behind the categorization of emotions and eye gaze behavior in ASD. This is crucial to designing and implementing methodologies and computer-based interventions (CBS) that can assist them in the improvement of social skills. Additionally, the use of a real-time interactive environment broadens the possibilities for experimentation and research that otherwise would not be possible (e.g. with pre-rendered animations).

In the following we will refer to previous works on character facial animation and gaze behavior on autism research. Then we will explain the project SARA and the tools developed and used in it. The next sections unfold each of the three experiments, providing details on the participants, stimuli, procedure and measures. Finally a discussion of the results of each of these experiments is given, followed by our conclusions 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. For instance, Whyte et al. [40] used game components (e.g., storyline, long-term goals, rewards) to create engaging learning experiences. Milne et al. [27] employed autonomous agents as tutors for teaching children with ASD conversation skills, and how to deal with bullying. Grawemeyer et al. [18] developed a sort of non-photorealistic embodied pedagogical agent together with and for young people with ASD.

ECHOES VE [3] presented a virtual environment where children with ASD need to assist a cartoony virtual character in selecting objects by following the character's gaze and/or pointing at the object. JeStiMulE [34] attempted to teach participants to recognize emotions on the faces and gestures of virtual characters, while considering the context. LIFEisGAME [4] deployed 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} [30] aided 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!* [37] was a set of seven interactive computer games that target the specific face impairments associated with autism.

A common element in all of these applications is the use of virtual characters to enhance or develop skills in subjects with ASD. These characters presented a defined visual style, which was either cartoony or realistic. In this sense, one of the assets of our project is the wider spectrum of real-time stylizations it provides. They resemble more artistic styles, which maintain as far as possible the human-like features of the virtual characters' faces. By offering varied painterly representations with different levels of abstraction, the experimenters can change the visual representation of the characters in real-time, achieving more personalized applications.

Regarding the study of gaze behavior, normally these applications do not employ eye tracking (even though, it could be adapted) to assess irregularities in eye contact and gaze behavior in participants with ASD. The DSM-5 [5] and ICD-10 [1], the criteria used by clinicians and researchers to diagnose and classify mental disorders, lists abnormalities in eye contact as one possible sign of deficient non-verbal communication patterns in people with autism. However, the findings on ASD and gaze avoidance have been inconsistent so far [23].

Kirchner et al. [21] for instance found that people with ASD fixated less on the face than NTD participants, while taking the *Multifaceted Empathy Test*(MET).

Klin et al. [22] carried out a study where high-functioning ASD and NTDs had to watch naturalistic social scenes while eye gaze was tracked. They observed that a reduction in eye region fixation time serves as the best predictor of ASD. Apart from that, individuals with high-functioning ASD and improved social adjustment skills showed increased fixation duration on mouths.

Senju and Johnson [33] explored reduced eye contact in ASD, distinguishing four models of atypical eye contact: (1) Hyperarousal Model: ASDs perceive the eyes of others as aversive stimuli; (2) Hypoarousal Model: gaze behavior stems from a hypoactivation of the amygdala in early childhood; (3) Communicative Intention Detector Model: atypical eye contact in ASD reposes on the difficulty in reading others mental states due to impairments inferring mentalistic significance of the eyes; (4) Fast Track Modulator Model: ASDs are impaired in regards of the subcortical face detection pathway. Therefore, they perform more poorly when pictures or videos of faces are presented for a short time span, or when they contain low spatial frequency information.

In contrast, Sawyer et al. [32] demonstrated that individuals with Asperger's syndrome showed no gaze avoidance in comparison to participants with ASD.

A possible solution to this debate was offered by Bird et al. [11]. They hypothesized and proved that the degree of alexithymia predicts deviant gaze behavior in ASD patients. In a subsequent work, Bird and Cook [10] went a step further. They argued that emotional deficits in ASD might actually be due to alexithymia instead of ASD symptomatology per se.

Our study nurtures from these previous hypothesis, and attempts to shed new light on the connection between autism, gaze behavior and recognition of facial emotions.

3 SARA

In order to evaluate the categorization of dynamic emotional facial expressions by children and adolescents with high-functioning ASD, SARA combines clinical psychology, computer graphics and 3D animation.

One of the main innovation includes the use of real-time NPR algorithms to abstract the faces of the virtual characters used in the various test. This allows us to explore how a reduction in the levels of detail of facial expressions affect their categorization by individuals with ASD.

On its part, eye tracking technologies allow for the study of gaze behavior in ASD, giving the possibility to extend and corroborate the results of previous studies with an adolescent sample.

3.1 DECT: Dynamic Emotion Categorization Test

The core of SARA is the DECT [29], an interactive computer-based tool created to determine the feasibility of using real-time animations, and to assess dynamic emotion categorization in facial expressions.

The original 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. These were presented on three intensity levels: weak, medium, and strong. The results showed that facial animations in virtual characters presented good concurrent validity with video clips of human actors in general. Furthermore, using parameterizable facial animations of virtual characters provided for an easier way to adjust emotion intensity levels in comparison to human actors. This motivated us to further explore the use of virtual characters in autism research.

Fig. 1 Participant during a DECT session.

Currently, the interaction with the software has been designed in a way where the experimenter has more control over it. Thanks to the real-time and parameterizable characteristics of the animations and NPR styles, the experimenter can generate different trials by direct manipulation of these elements. During the practice trials the experimenter explains the participant what needs to be done (Figure 1).

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 answer options with emotion names.

3.2 Frapper: Filmakademie Application Framework

The development of the current DECT versions (R-DECT and NPR-DECT), as well as the implementation of the NPR algorithms and real-time animations was done using our in-house software development platform Frapper[1], particularly the Agent Framework [6].

Frapper is a C++, Ogre3D and Qt based development environment consisting of a node-based scene model, a model-view-controller architecture, and a panel-oriented user interface similar to commercial 3D packages. The Agent Framework is the set of functionalities (nodes and plugins) that allow users the rapid prototyping of applications that make use of virtual characters.

New functionality can be simply added by creating node- or panel-plugins, which incorporate the desired features and allows linking to third-party libraries. In this way we bring complex technologies into Frapper like eye tracking, computer vision, synthetic speech, voice recognition, or artificial intelligence; as well as the integration of alternative input devices.

Frapper is available as open source and is released under the GNU Lesser General Public License 2.1.

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

3.3 Animated Virtual Characters

All characters were created to resemble a realistic-looking person. For the studies presented in this paper we used two characters: an elderly man, named Hank, and a young woman, named Nikita.

These were rigged using the Adaptable Facial Setup (AFS) [19], a tool-set that relies on a complete motion capture-based library of deformations, based on the Facial Animation Coding System (FACS) [14], generating high quality, natural and non-linear deformations. Once the characters were finished, they were exported into our animation framework. The facial movements are described in terms of Action Units (AUs), providing a parameterizable way of creating animated facial expressions, which could then be translated to any other character with a similar FACS-based rig.

Both Frapper and the Agent Framework are provided with the two human-like characters distributed under the Creative Commons Attribution Non Commercial Share Alike 3.0 Unported License.

3.4 Interactivity

A relevant aspect of our project is the level of interactivity it provides thanks to the real-time characteristics of our framework. Having a tool that generates animations and abstract visual representations in real-time allows the psychologists and experimenters to fine tune and parameterize the tests themselves. This results in more flexible and elaborated interactive experiments, which can be adapted to the experimenter's requirements, or participant's needs. It also makes the experimenters independent from the animator, an important aspect to consider when using a computer-based research tool.

4 DECT Session

A session with both versions of DECT (R-DECT and NPR-DECT) consists of several trials, which in turn are composed by a series of elements displayed in Frapper (Figure 2).

Frapper works in combination with the open source software PsychoPy [28]. PsychoPy is is a free, open-source application written in Python. It allows to design and implement the logic of the experiments, along with the collection of data.

A trial 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.

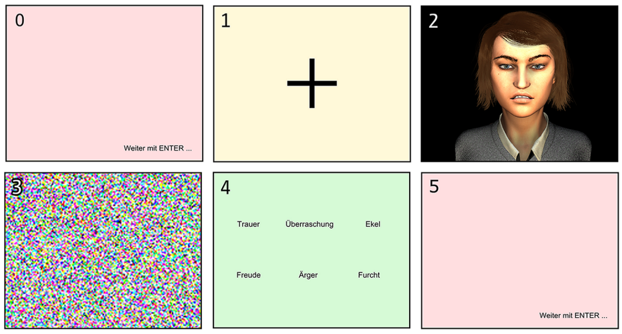

Fig. 2 Screens of the DECT: (0) Initial screen, (1) Fixation Cross, (2) Character with an angry expression, (3) White noise, (4) Forced choices with emotion names, (5) Initial screen for next trial.

The next element is a fixation cross (Fig. 2(1)), which appears for 0.5 seconds and indicates where the participant should fixate his gaze.

Afterwards, either Hank or Nikita appears, displaying a real-time generated animation of a facial expression. In the case of the R-DECT, different intensity and speed levels were used (Fig. 2(2)). For the NPR-DECT, the intensity of the expression is always strong. Moreover, it extends the visual representation to one of the following NPR styles: original (realistic-looking with no stylization), pencil drawing, coherent line drawing, image abstraction or watercolors. As for the level of abstraction, it can be: low, medium, or high.

Then, a screen with white noise ((Fig. 2(3)) is showed for 0.5 seconds, masking participants' iconic memory with task-irrelevant information.

Finally, another screen with the answer 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 constant within the experimental session, but counterbalanced across participants according to a sequentially counterbalanced latin square.

In the following we will present the R-DECT, NPR-DECT, and the study on gaze behavior.

5 Experiment 1: R-DECT

The R-DECT [7] is a DECT version that was part of a battery of tests for *rapid* social cognition and intuitive moral reasoning assessment in children and adolescents with ASD. It also served to validate the improved facial animation from the original DECT, as well as to validate the test itself as a tool for the interactive categorization of emotional expressions.

5.1 Participants

39 adolescents with ages between 14.0 and 17.9 years old, and IQ ≥ 70 took part in the experiment. We considered two groups:

- 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.

5.2 Stimuli

For each session, 36 animations were evaluated: 2 (characters) x 6 (basic emotions) x 3 (intensity levels).

The intensity levels were: weak, medium and strong.

The speed variable 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.

5.3 Procedure

The rapid social cognition experiment comprised two sessions, each lasting between 1.5 to 2 hours. However, the R-DECT took only about 15 minutes to be carried out, where each trial had a duration of approximately 2 seconds. The R-DECT was administered as the first of five tests in the corresponding session. The way each DECT session was carried out is explained in Section 4.

5.4 Results

In total, 62.2% of the animations were categorized correctly. Accuracy rate for the NTD group was 65.7%, whereas for the ASD group it 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*: $F(1,37) = 5.36$, $p = .026$). This indicated that

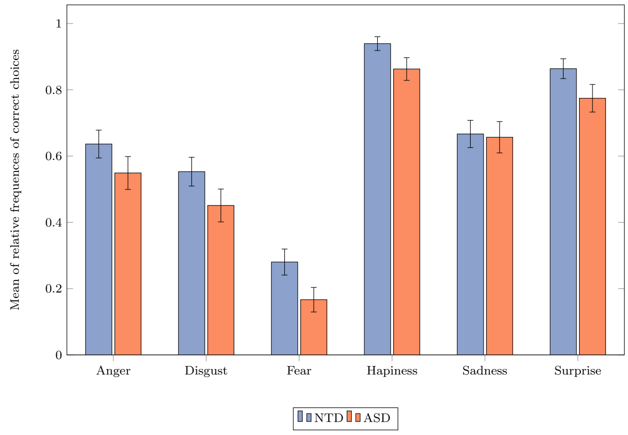

Fig. 3 Mean of relative frequencies of correct categorizations of six basic emotions in ASD and NTD groups. Error bars represent 95% CI.

the ASD group performed significantly worse than the NTD group (Figure 3).

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. Similar results concerning order of accuracy rates of basic emotions were obtained in [25] [26].

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% [7].

6 Experiment 2: NPR-DECT

The second version called NPR-DECT comprises one of the novelties of our project: the use of NPR algorithms to abstract and manipulate visual information in the faces of our virtual characters.

We use NPR to stylize the faces of the characters because it provides variation in the level of abstraction and visual information, adapting images to "focus the viewer's attention" [16]. Thus the information load in the characters' facial expressions can be reduced, conveying the emotional information more efficiently [13] [17].

NPR-DECT was the 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. In addition, this test served to study the link between alexithymia, ASD and deviant gaze behavior.

Fig. 4 NPR styles (medium abstraction level) applied to Nikita (upper row) and Hank (lower row). Left to right: image abstraction, pencil drawing, watercolor, coherent line drawing.

6.1 Participants

31 NTD adults with ages between 20 and 35 years old took part in this (pilot) experiment. 9 were male and 22 female.

No subjects with ASD were considered, because we wanted to validate the developed NPR styles. The final experiment is described in [36].

6.2 Stimuli

The NPR-DECT comprised 2 (characters) x 6 (basic emotions) x 13 (1 + 12 NPR style x abstraction combinations). This resulted in 156 trials, which were presented in a pseudorandomized order.

The NPR styles used (Figure 4), apart from the original realistic-looking, were:

- Coherent Line Drawing (*CLD*)
- Pencil Drawing (*PD*)
- Image Abstraction (*IA*)
- Watercolors (*W*)

These were instantiated in one of three levels of abstraction: low, medium, and high.

6.3 Procedure

The way each session was carried out is similar to the R-DECT. The NPR-DECT had a duration of approximately 20 minutes. At the end of each session the participants needed to fill a computer-based questionnaire to rate the likability and recognizability of the faces abstracted using NPR styles.

6.3.1 Likeability and Recognizability Questionnaire

The questionnaire comprised 26 images: 2 (characters) x 4 (stylized faces by NPR) x 3 (levels of abstraction) + 2 (realistic-looking Hank and realistic-looking Nikita). For each image, two questions were posed:

1. How good were you able to recognize the emotions from this representation?
2. How 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”).

Additionally, during the whole NPR-DECT session participants’ gaze was tracked by an RED-250 eye tracker (SMI). The eye tracker was integrated in Frapper to allow the communication between each other.

6.4 Results

In total, 71.4% of the emotional expressions were correctly categorized. Due to technical problems in one session, rating data are based on n=30 participants only.

We found a considerable difference of accuracy between both characters: Hank: 66.5% vs. Nikita 76.3% (Figure 5).

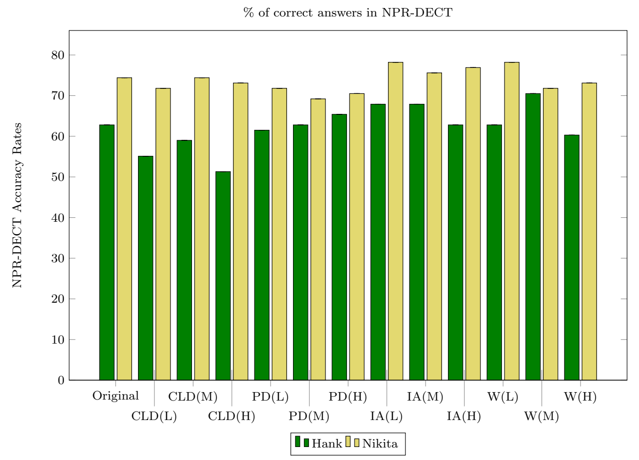

Fig. 5 Percentage (%) of correct answers in NPR-DECT. (*Note.* CLD = Line Drawing; PD = Pencil Drawing; W = Watercolor; IA = Image Abstraction; (L) Low abstraction; (M) Medium abstraction; (H) High abstraction)

Differences in levels of accuracy seem to be dependent on the character in a certain style with a certain level of abstraction. For Hank, the lowest rate was obtained with *CLD - high abstraction* (55.9%), whereas the highest rate was with *W - medium abstraction* (73.7%). For Nikita, the lowest percentage was obtained with *PD - medium abstraction* (71.5%), whereas the highest percentage was noted for *IA - low abstraction* (81.7%).

Regarding the recognizability and likability rates (Table 1), these were the highest for the original representation ($M = 2.02, SD = 1.11$), followed by *IA* ($M = 2.44, SD = 1.05$), *CLD*($M = 3.69, SD = 1.70$), *W* ($M = 3.73, SD = 1.39$), and finally *PD* ($M = 4.17, SD = 1.33$).

Table 1 Mean of (a) recognizability and (b) likeability ratings of the rendering styles and virtual characters (VC) from 1 (very good) to 7 (very bad). Mean values are color-coded from white to red, more intense red meaning worse. $M(H)$: Hank's mean values. $SD(H)$: standard deviation of Hank's mean values. $M(N)$: Hank's mean values. $SD(N)$: standard deviation of Nikita's mean values

(a) Recognizability

Style	$M(H)$	$SD(H)$	$M(N)$	$SD(N)$
Ori.	2.27	1.34	1.77	0.77
CLD (L)	3.67	1.35	3.07	1.6
CLD (M)	3.83	1.64	3.07	1.74
CLD (H)	4.73	1.74	3.8	1.67
PD (L)	4.53	1.28	3.8	1.24
PD (M)	4	1.14	4.07	1.55
PD (H)	4.27	1.23	4.37	1.45
IA (L)	2.1	1.03	1.77	0.5
IA (M)	2.5	0.97	2.1	0.76
IA (H)	3.1	1.12	3.07	1.05
W (L)	3.53	1.36	3	1.23
W (M)	3.67	1.30	3.17	1.23
W (H)	4.63	1.22	4.37	1.33

(b) Likeability

Style	$M(H)$	$SD(H)$	$M(N)$	$SD(N)$
Ori.	2.7	1.18	2.07	1.14
CLD (L)	4	1.29	3.23	1.38
CLD (M)	4.07	1.31	3.2	1.47
CLD (H)	4.83	1.39	3.9	1.49
PD (L)	4.47	1.22	4.7	1.37
PD (M)	4.2	1.27	4.37	1.67
PD (H)	4.2	1.21	4.93	1.46
IA (L)	2.57	1.17	2.03	0.81
IA (M)	2.9	1.16	2.47	1.22
IA (H)	3.33	1.37	3.6	1.33
W (L)	3.87	1.43	3.6	1.43
W (M)	4.2	1.37	3.77	1.38
W (H)	4.9	1.32	4.87	1.38

7 Experiment 3: Gaze Behavior

The current experiment addressed the question whether deviant gaze behavior and higher scores of alexithymia are more common in individuals with ASD when compared to NTD controls. Alexithymia describes patients' difficulties in communicating due to the "inability to find appropriate words to describe their feelings" [35].

Our study is based on the work of Bird et al. [11], where they proposed a possible solution for the ongoing debate about gaze behavior and ASD. They conducted an eye tracking study with adults with and without autism. Bird et al. interpret their results as it is the degree of alexithymia what influences a person's eye fixation rather than the severity degree of autistic symptomatology. This means that individuals with autism only display anomalous eye contact when they suffer from co-morbid alexithymia.

The novelty of our work is to find out if Bird et al.'s findings in an adult sample could be replicated using an adolescent sample. We also attempt to improve their scientific and methodological approach. This study represents another use case of the multimodal potential of the SARA testbed.

7.1 Participants

35 male adolescents between 14 and 17 years old took part in the experiment. From this group, 14 were diagnosed with ASD and 21 were NTD. The mean age in the ASD group was 15.33 years ($SD = 0.99$), and in the NTD group 15.64 years ($SD = 1.15$).

The mean IQ, measured by the CFT 20-R [39], was 103.71 ($SD = 14.31$) for the ASD group, and 103.48 ($SD = 8.94$) for NTDs.

Individuals with ASD who also met criteria for ADHD were not excluded from the study. Apart from that, participants who needed a correction to the visual acuity higher than 1.5 diopters, and therefore need to wear glasses or soft contact lenses, were excluded since the eye tracking would be inaccurate for higher degrees of diopters. This however was not necessary for participants who wore rigid gas-permeable lenses. Additionally, former participation in any experiments with one of the DECT versions was one of the exclusion criteria of the study.

7.2 Stimuli

The stimuli for this study corresponded to the NPR-DECT. As such, the test presented: 2 (characters) x 6 (basic emotions) x 13 (1 + 12 NPR style x abstraction combinations) = 156 trials..

However, owing to the amount of data to be evaluated, in this paper we present some preliminary results that correspond to the evaluation of only 12 animations of the characters in the realistic-looking style (2 characters x 6 basic emotions).

7.3 Procedure

The participants underwent other tasks that included: emotional categorization, assessment of their IQ and different questionnaires. The whole experimental session took approximately 90 minutes per participant.

To collect gaze behavior data during emotional categorization, we used the RED250 remote eye tracking system (SensoMotoric Instruments - SMI, Teltow, Germany). Among its advantages are its ease of use, high data quality with high speed capabilities, and calibration times, which are of the order of few seconds. Moreover, it offers the possibility to work with subjects that wear most glasses and contact lenses (these are specified by the manufacturer). Nevertheless, for the present study the participants either wore no glasses or wore rigid contact lenses.

For our study, it was necessary to communicate Frapper with the eye tracker. To that end we implemented a plugin in Frapper that established a network communication between the test (R-DECT or NPR-DECT) and the eye tracker SDK.

The procedure to perform the eye tracking required the participant to seat in front of the RED250 and computer screen. Then chair, chin rest, table, and monitor were properly adjusted according to the person's height. Subsequently, the calibration was conducted and the actual experiment began. The participant was asked to move as little as possible in order to minimize artifacts in the eye tracking data.

The emotional categorization was carried out using the NPR-DECT, which took approximately 15 to 20 minutes. The answers to the categorization were typed in by the experimenter to avoid distortions in the eye tracking data.

Additionally, a number of instruments were used to measure intelligence quotient and degree of alexithymia (Table 2).

7.3.1 Dependent Variables

We considered three dependent variables: gaze behavior, correctness of the answers, and reaction times.

The gaze behavior was assessed by considering the "total dwell time" (TDT), which is the sum of sample durations for all gaze data samples that hit the area of interest (AOI) [20]. The relevant AOIs were eyes, mouth and face (Figure 6).

The TDTs on these specific AOIs were further used to calculate: face-non-face difference (FNFD) and eye-mouth difference (EMD). According to Bird et al. [11], FNFD serves as an index of social attention. EMD is the total fixation time on the eyes compared to the mouth.

A value of 0 indicates equal attention to facial and non-facial stimuli, as well as equal fixations on the eyes and mouth. Values *greater than* 0 imply attentional preference towards facial stimuli (or the eyes) compared to non-facial stimuli (or the mouth). Values *lower than* 0 signify the opposite case.

The methodological difference between our study and Bird et al.'s is that they used ratios between face-non-face fixations and mouth-eyes fixations, and we are using differences because there were cases where the denominator value was 0, causing division by zero.

The second variable, correctness of answers, regarded the 12 relevant NPR-DECT items.

The third dependent measure, reaction times, was obtained through the data recording program PsychoPy in combination with Frapper.

Table 2 Overview and description of instruments used in the gaze behavior study. *Note*: NPR-DECT = Non-Photorealistic Rendering Dynamic Emotion Categorization Test, CFT 20-R = Culture Fair Intelligence Test, TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Questionnaire).

instrument	performer	type of instrument	approximate duration
NPR-DECT	participant	computer test (ET)	20 minutes
CFT 20-R	participant	paper-pencil IQ test	30 minutes
TAS-26	participant	paper-pencil questionnaire	10 minutes
BVAQ-AB	participant	paper-pencil questionnaire	10 minutes

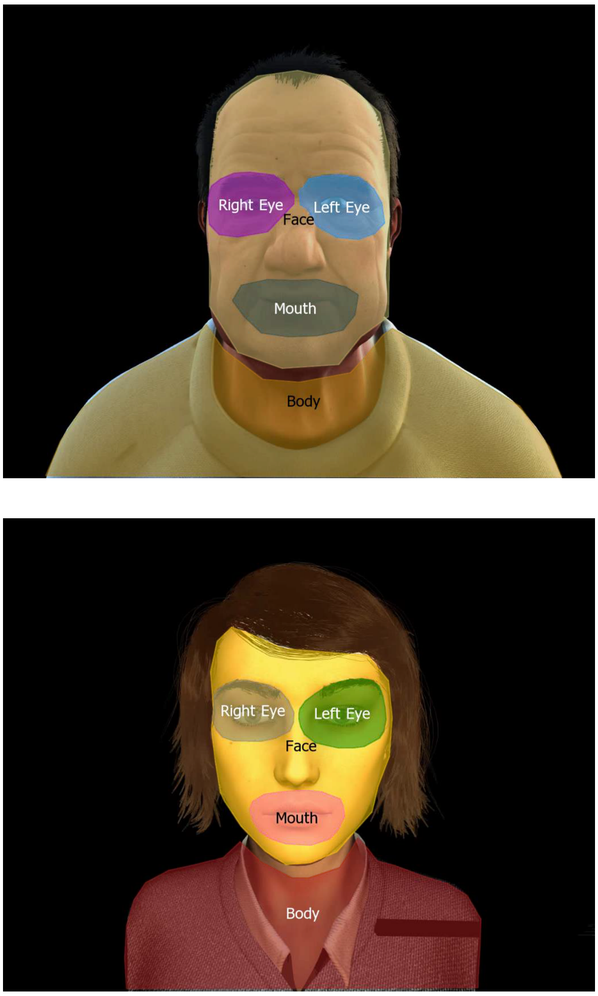

Fig. 6 Examples of stimuli with defined AOIs

7.3.2 Hypotheses

The research consisted of a quasi-experimental between subjects design. We considered as independent variable *group affiliation (ASD vs. NTD)*, as dependent variable *gaze behavior (FNFD & EMD)*, and as presumed moderator variable *degree of alexithymia*. The latter was measured via TAS-26 [24] and BVAQ-AB [9]. The data collected was pre-processed using the eye tracking analysis program SMI BeGaze, Version 3.4.5.

Three hypothesis were formulated:

- H1: The group affiliation will influence gaze behavior. In detail, control participants (NTDs) will fixate social stimuli longer and more frequently than ASDs.
- H2: The higher a participant's alexithymia score, the less frequent and shorter social relevant stimuli will be fixated (i.e. lower FNFD and EMD difference).
- H3: The group affiliation influences the degree of alexithymia. In detail, ASDs will score higher on the alexithymia instruments than NTDs.

7.3.3 Measures

To test the first hypothesis, we conducted a one factorial ANOVA with the independent variable *group affiliation (ASD vs. NTD)* and the dependent variable *gaze behavior (FNFD & EMD)*.

For the second hypothesis, we computed Spearman correlations regarding *gaze behavior*, and mean scores of TAS-26 and BVAQ-AB.

As for the third hypothesis, we calculated another one factorial ANOVA with the independent variable *group affiliation (ASD vs. NTD)* and the variable *degree of alexithymia*.

7.4 Results

A one-way ANOVA on age and IQ revealed no significant group difference in terms of age ($F(1{,}33) < 1$), or as to the IQ ($F(1, 33) < 1$).

7.4.1 Hypothesis 1

We tested whether the group affiliation (ASD vs. NTD) would influence a person's gaze behavior. Regarding the FNFD, a Mann-Whitney U test was used to check a significant effect of group affiliation on this gaze behavior. In this respect, no significant main effect was found ($U = 139.00, z = -0.27, p = .803$). This implies that there were no larger differences between ASDs ($M = 2074.87, SD = 58.82$) and NTDs ($M =$

2042.88, $SD = 154.99$) (Figure 7). When observing the data on a descriptive level, one can notice that NTDs actually showed lower FNFD values than ASDs, which is contrary to hypothesis 1.

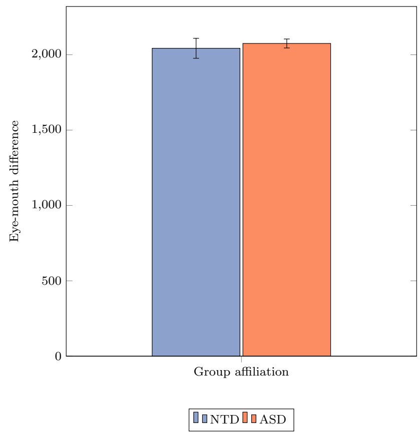

Fig. 7 Group difference regarding the FNFD of AOI-based total dwell times. Error bars represent 95% CI.

As for EMD, the ANOVA revealed no significant group effect either. This means that ASDs ($M = 407.98$, $SD = 766.26$) did not differ reliably from NTDs ($M = 261.93, SD = 817.16$), $F(1, 34) < 1$ (Figure 8). Once again our findings demonstrate a direction contrary to our hypothesis on a descriptive level indicating that NTDs had lower EMD scores than ASDs.

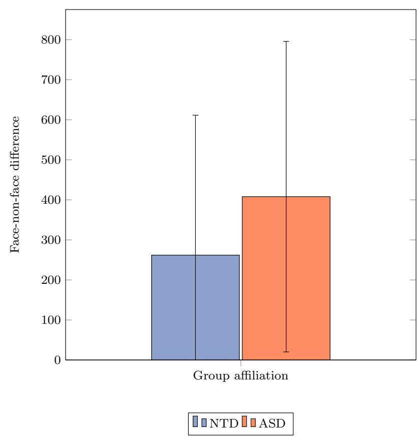

Fig. 8 Group difference regarding the EMD of AOI-based total dwell times. Error bars represent 95% CI.

7.4.2 Hypothesis 2

To test if the degree of alexithymia has a direct impact on the participant's gaze behavior we used Pearson correlations. We did the analysis on both the total sample and within groups.

Contrary to the hypothesis, none of the correlations between the alexithymia instruments and the FNFD value were significant. Neither for the total sample, nor within both groups regarded separately (Tables 3, 4 and 5).

One significant moderate relationship between the EOT (externally oriented thinking) scale of the TAS-26 and the EMD value ($r = .35, p = .046$) appeared in the total sample (Table 6). When observing this relationship within groups, it still remained significant for the NTDs ($r = .50, p = .024$), whereas there was no significant relationship within the ASD subsample (Tables 7 and 8). When considering NTDs only, we also found a moderate correlation between the verbalizing scale of the BVAQ-AB and the EMD value. However, this effect was only a statistical tendency ($r = .42, p = .067$). All of the other correlations between the alexithymia instruments and the EMD score were non-significant.

7.4.3 Hypothesis 3

This hypothesis states that the group membership influences the degree of alexithymia. To test this, two separate analysis of variance for the TAS-26 and BVAQ-AB were performed.

A Shapiro-Wilk test, concerning TAS-26 scores, revealed that in both groups the normality assumption was not violated, with the exception of the DIF (difficulty identifying feelings) scale: NTD ($W = .90, p = .046$), and the ASD ($W = .86, p = .033$). Therefore, a Mann-Whitney U test was applied for this scale, whereas ANOVAs where used for the remaining scales.

For the overall TAS-26 scale, a significant main effect was revealed, $F(1, 32) = 5.14, p = .030, g = 0.95$. It means that ASDs scored higher ($M = 43.79, SD = 9.78$) than NTD participants ($M = 37.15, SD = 7.32$) (Figure 9).

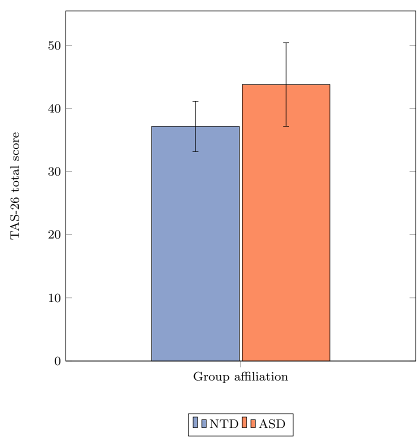

Fig. 9 Group differences on the total TAS-26 score.

Table 3 Correlations between the TAS-26 respectively the BVAQ-AB and the FNFD for the total sample

	Correlations between TAS-26 and FNFD ($n = 34$)					
	Overall score	DIF	DDF	EOT	DIF+DDF	
FNFD	$r = -.08$ $p = .648$	$r = .15$ $p = .394$	$r = -.15$ $p = .389$	$r = -.20$ $p = .251$	$r = .002$ $p = .993$	
	Correlations between BVAQ-AB and FNFD ($n = 34$)					
	Overall score	Emotionalizing	Verbalizing	Fantasizing	Identifying	Analyzing
FNFD	$r = -.13$ $p = .484$	$r = -.18$ $p = .309$	$r = -.02$ $p = .921$	$r = .05$ $p = .764$	$r = -.05$ $p = .777$	$r = -.24$ $p = .182$

Note. TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Scale, FNFD = face-non-face difference, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking.

Table 4 Correlations between the TAS-26 respectively the BVAQ-AB and the FNFD for ASDs

	Correlations between TAS-26 and FNFD ($n = 14$)					
	Overall score	DIF	DDF	EOT	DIF+DDF	
FNFD	$r = -.09$ $p = .749$	$r = -.21$ $p = .466$	$r = .20$ $p = .488$	$r = -.22$ $p = .458$	$r = -.02$ $p = .939$	
	Correlations between BVAQ-AB and FNFD ($n = 14$)					
	Overall score	Emotionalizing	Verbalizing	Fantasizing	Identifying	Analyzing
FNFD	$r = -.25$ $p = .399$	$r = -.17$ $p = .570$	$r = -.01$ $p = .964$	$r = -.19$ $p = .514$	$r = -.08$ $p = .780$	$r = -.26$ $p = .363$

Note. TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Scale, FNFD = face-non-face difference, ASD = autism spectrum disorder, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking.

Table 5 Correlations between the TAS-26 respectively the BVAQ-AB and the FNFD for NTDs

	Correlations between TAS-26 and FNFD ($n = 21$)					
	Overall score	DIF	DDF	EOT	DIF+DDF	
FNFD	$r = -.18$ $p = .458$	$r = .34$ $p = .141$	$r = -.37$ $p = .110$	$r = -.22$ $p = .355$	$r = -.07$ $p = .765$	
	Correlations between BVAQ-AB and FNFD ($n = 21$)					
	Overall score	Emotionalizing	Verbalizing	Fantasizing	Identifying	Analyzing
FNFD	$r = -.15$ $p = .533$	$r = -.22$ $p = .347$	$r = -.05$ $p = .851$	$r = .18$ $p = .453$	$r = -.13$ $p = .588$	$r = -.31$ $p = .186$

Note. TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Scale, FNFD = face-non-face difference, ASD = autism spectrum disorder, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking.

Another significant difference was found for the scales DIF (difficulty identifying feelings) and DDF (difficulty describing feelings) combined, $F(1, 32) = 5.29, p = .028, g = 1.02$, where participants with ASD scored significantly higher ($M = 28.21, SD = 10.14$) than their NTD peers ($M = 21.90, SD = 5.85$).

Additionally, we found a marginally significant difference between ASDs and NTDs concerning the DDF scale, $F(1, 32) = 3.02, p = .092, g = 0.59$, indicating once again higher values for ASDs ($M = 14.14, SD = 5.20$) than for NTDs ($M = 11.40, SD = 4.01$). See Table 9 for an overview.

On the contrary, the group difference in respect of the overall BVAQ-AB scale was not significant ($F(1, 32) < 1$), and neither were group differences in terms of the subscales.

However, we were able to reveal one marginally significant main effect as for the subscale DIF, $F(1, 32) = 3.14, p = .086$, meaning that NTDs ($M = 15.55, SD = 4.95$) exhibit less problems when it comes to the identifi-

Table 6 Correlations between the TAS-26 and BVAQ-AB respectively the EMD for the total sample

	Correlations between TAS-26 and EMD ($n = 34$)					
	Overall score	DIF	DDF	EOT	DIF+DDF	
EMD	$r = .31$	$r = .16$	$r = .15$	$r = .35$	$r = .18$	
	$p = .078$	$p = .379$	$p = .385$	$p = .046$	$p = .318$	
	Correlations between BVAQ-AB and EMD ($n = 14$)					
	Overall score	Emotionalizing	Verbalizing	Fantasizing	Identifying	Analyzing
FNFD	$r = .27$	$r = -.15$	$r = .24$	$r = .19$	$r = .11$	$r = .22$
	$p = .118$	$p = .398$	$p = .172$	$p = .277$	$p = .531$	$p = .214$

Note. TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Scale, EMD = eye-mouth difference, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking.

Table 7 Correlations between the TAS-26 and BVAQ-AB respectively the EMD for ASDs

	Correlations between TAS-26 and EMD ($n = 14$)					
	Overall score	DIF	DDF	EOT	DIF+DDF	
EMD	$r = .23$	$r = .27$	$r = .12$	$r = -.03$	$r = .23$	
	$p = .439$	$p = .344$	$p = .677$	$p = .930$	$p = .438$	
	Correlations between BVAQ-AB and EMD ($n = 14$)					
	Overall score	Emotionalizing	Verbalizing	Fantasizing	Identifying	Analyzing
FNFD	$r = .17$	$r = -.28$	$r = -.01$	$r = .20$	$r = .25$	$r = .09$
	$p = .559$	$p = .334$	$p = .975$	$p = .483$	$p = .385$	$p = .751$

Note. TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Scale, EMD = eye-mouth difference, ASD = autism spectrum disorder, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking.

Table 8 Correlations between the TAS-26 and BVAQ-AB respectively the EMD for NTDs

	Correlations between TAS-26 and EMD ($n = 21$)					
	Overall score	DIF	DDF	EOT	DIF+DDF	
EMD	$r = .31$	$r = -.06$	$r = .10$	$r = .50^{*}$	$r = .04$	
	$p = .191$	$p = .804$	$p = .666$	$p = .024$	$p = .871$	
	Correlations between BVAQ-AB and EMD ($n = 21$)					
	Overall score	Emotionalizing	Verbalizing	Fantasizing	Identifying	Analyzing
FNFD	$r = .35$	$r = -.07$	$r = .42^{+}$	$r = .18$	$r = -.13$	$r = -.31$
	$p = .133$	$p = .769$	$p = .067$	$p = .453$	$p = .588$	$p = .186$

Note. TAS-26 = Toronto Alexithymia Scale, BVAQ-AB = Bermond-Vorst Alexithymia Scale, EMD = eye-mouth difference, ASD = autism spectrum disorder, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking.
$+p < .10, *p < .05, **p < .01, ***p < .001$

cation of emotions than ASDs ($M = 19.79, SD = 8.95$) do (Figure 10, Table 10).

The analysis of the correlation between recognition of emotions and overall scales and subscales for the assessment of alexithymia (TAS-26 and BVAQ-AB) showed that there were no significant correlations.

Conversely, the correlation between the total scores of both instruments (TAS-26 and BVAQ-AB) is a strong one ($r = .69, p < .001$).

8 Discussion

8.1 R-DECT

The results of the experiment validated the R-DECT (no floor effects or ceiling effects appeared), proving that it can be used to assess the general ability of categorizing facial expressions of emotions. Moreover, the test was able to detect the impairments in facial emotion

Table 9 Group differences on the total score and subscales of the TAS-26

	ASD ($n = 14$)		NTD ($n = 21$)		Statistics			
TAS-26 Scale	M	SD	M	SD	F	p		g
Overall scale	43.79	6.03	37.15	7.32	5.14*	.030		0.95
DDF	14.14	5.20	11.40	4.01	3.02+	.092		0.59
EOT	15.57	3.20	15.25	3.99	0.06	.804		0.08
DIF + DDF	28.21	6.27	21.90	5.85	5.29*	.028		1.02
					U	z	p	g
DIF	14.07	6.03	10.50	3.10	94.00	-1.61	.106	0.78

Note. TAS-26 = Toronto Alexithymia Scale, ASD = autism spectrum disorder, NTD = neuro-typically developed, DIF = difficulty identifying feelings, DDF = difficulty describing feelings to others, EOT = externally oriented thinking, g = Hedge's g
$+p < .10, *p < .05, **p < .01, ***p < .001$

Table 10 Group differences on the total score and subscales of the BVAQ-AB

	ASD ($n = 14$)		NTD ($n = 21$)		Statistics		
BVAQ-AB Scale	M	SD	M	SD	F	p	g
Overall scale	113.86	19.31	108.80	12.48	0.79	.380	0.32
Emotionalizing	24.27	4.20	24.30	3.23	0.00	.978	0.01
Verbalizing	24.50	7.59	23.06	5.41	0.42	.521	0.22
Fantasizing	22.14	8.71	24.49	6.57	0.81	.376	0.31
Identifying	19.79	8.95	15.55	4.95	3.14+	.086	0.61
Analyzing	23.64	6.27	22.10	4.67	0.68	.416	0.28

Note. BVAQ-AB = Bermond-Vorst Alexithymia Scale, ASD = autism spectrum disorder, NTD = neuro-typically developed, g = Hedge's g
$+p < .10, *p < .05, **p < .01, ***p < .001$

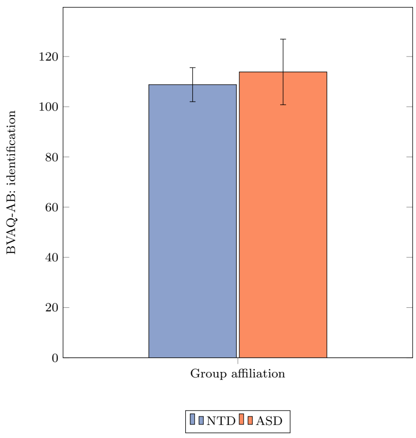

Fig. 10 Group differences on the BVAQ-AB subscale "identifying feelings".

categorization in the ASD group. However, the statistical power of the study was too low to assess which emotion(s) contributed to this general difference.

The R-DECT also served to validate the generated facial animations, shedding light on the improvements that needed to be done to convey the corresponding emotional meaning. The intensity of the displayed emotional facial expressions demonstrated to play a considerable role in categorization. To improve the test, some adjustments of animations had to be made especially in the condition where intensity of the expression was intended to be "medium" or "strong". That seems to be the case especially for the animation expressing the emotion *fear*.

8.2 NPR-DECT

The preliminary results of the NPR-DECT, considering only a sample of neurotypically developed (NTD) participants, gave some insight on the categorization and perception of the stylization methods used to abstract the characters' facial expressions.

As for recognizability and likeability, we could observe that, apart from the realistic-looking (original) style, the "image abstraction" stylization in all levels led to better rates in both characters. Regarding the other styles, we could not draw any definite conclusion given that no clear recognition or likeability pattern arose.

We also found a significant difference of accuracy between both virtual characters, being Nikita the one that facilitated the categorization of emotions across all NPR styles.

8.3 Gaze Behavior

The results of our experiment regarding the first hypothesis did not replicate anomalies on the gaze variables FNFD (face-non-face difference) and EMD (eye-mouth difference). In fact, the conducted analysis revealed no differences between both the ASD and NTD groups.

Some arguments that might justify this result are: differences in the stimuli in our experiments, compared to the stimuli in Bird et al. [11] (we used animated characters and they employed footage with real actors); duration of the trials (theirs were considerably longer); and the presence of distracting content (in Bird et al.'s study, participants could choose not to look at the characters, while in our test the characters occupied most of the screen space). Another possible explanation might be a considerable number of ASD participants that took part in another study of our department about five years ago. On the one hand, this speaks in favor of an especially motivated sample. On the other hand, it also signifies that these patients received their diagnoses in early childhood and had the chance to work on their deficits. Finally, it might be that there is indeed no connection between high-functioning ASD and deviant gaze behavior, as concluded by Sawyer et al. [32].

Our findings concerning the second hypothesis indicated that alexithymia has only little to no influence on the gaze behavior (FNFD or EMD). One exception on the overall results was the subscale EOT (externally oriented thinking) from the TAS-26, where there was a moderate relationship between EOT and EMD. This correlation became stronger regarding only the NTD group, but it was not significant within the ASD group. A reason for this could be that EOT measures a construct different than the other two scales, DIF (difficulty identifying feelings) and DDF (difficulty describing feelings to others), or the overall scale.

Finally, our results partially support the assumption that ASDs show higher values of alexithymia. In detail, there was strong evidence for this hypothesis when considering the overall TAS-26 scale and the scales DIF and DDF combined. However, there were no significant group differences when utilizing the BVAQ-AB. One could argue that the TAS-26 scale has been a widely used instrument for the assessment of alexithymia, including adolescent samples from the age of 14 and above. On the contrary, the BVAQ-AB is a relatively new instrument that has not been validated yet for adolescent populations. Hence it might be that the BVAQ-AB is only validly applicable on adults. One has to notice that the BVAQ-AB comprises 40 items as compared to only 26 TAS-26 items. Considering that the participants had to complete four questionnaires prior to the experiment, a longer questionnaire like the BVAQ-AB might have resulted in minimized motivation rates.

Further limitations that might have contributed to the non-corroboration of the hypotheses are: the small overall sample ($n = 35$); ADHD comorbidity in the study; the use of the eye-tracking variable FNFD (face-non-face difference), which does not entirely fit into our research because the faces of the characters occupy most of the screen, as well as the time given to look at the stimuli (less than 2 seconds); or possible influence of training and psychotherapy on gaze behavior along the years.

9 Conclusions

In this paper we have compiled the results of previous experiments carried out within the project SARA. In particular, the focus was emotion categorization of facial expressions, and the relation between alexithymia, ASD and eye gaze.

As for the emotional categorization, we validated the use of animated facial expressions to work in autism research. They provide the flexibility and parameterization required in an experiment, which could be personalized according to the participants', or experimenters', needs.

We also introduced a novel element: stylization of facial expressions. The validation of the different abstracted representations was assessed by NTD subjects. However, the most important conclusion we could draw was that styles resembling the original one (e.g. image abstraction) were preferred. Nevertheless, we could not assess the differences between styles and what elements should be improved.

Regarding gaze behavior and its link to alexithymia and ASD, our results did not reproduce the ones from Bird et al. [11]. This might indicate that there is indeed a relationship between gaze behavior and ASD, which will also be explored in our future research.

In its current status SARA is more a research tool than an intervention tool. Therefore, the interaction with the tool has not been exploited to its maximum.

The former results serve as a guideline for future research in the area, in particular the outcome from the NPR-DECT study. More research needs to be done to improve, or come up with real-time NPR algorithms that allow the abstraction of certain facial regions, as well as a uniform scale for abstraction. Gaze behavior will also be assessed when categorizing abstracted facial expressions, which will shed more light on the effect of stylization on ASD.

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. This may also help to build computer-based interventions (CBI) in the future.

Acknowledgements The SARA project (officially *Impact of Non-Photorealistic Rendering for the understanding of emotional facial expressions by children and adolescents with high-functioning Autism Spectrum Disorders* is funded by DFG - German Research Foundation (AR 892/1-1, DE 620/18-1, RA 764/4-1). Thanks to Kai Goetz for the graphics.

References

1. The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research, 10th ed. edn. Geneva, Switzerland: World Health Organization (1992)
2. Facts About ASD (2014). Retrieved April 06, 2015, from Center for Disease Control and Prevention: http://www.cdc.gov/ncbddd/autism/facts.html
3. Alcorn, A., Pain, H., Rajendran, G., Smith, T., Lemon, O., Porayska-Pomsta, K., Foster, M.E., Avramides, K., Frauenberger, C., Bernardini, S.: Social communication between virtual characters and children with autism. In: AIED, pp. 7–14 (2011)
4. Alves, S., Marques, A., Queirós, C., Orvalho, V.: LIFEisGAME prototype: A serious game about emotions for children with autism spectrum disorders. PsychNology Journal **11**(3), 191–211 (2013)
5. American Psychiatric Association: Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.). Arlington, VA: American Psychiatric Publishing (2013)
6. Arellano, D., Helzle, V., Schaller, U., Rauh, R.: Animated faces, abstractions and autism. In: Intelligent Virtual Agents 2014, *LNCS*, vol. 8637, pp. 22–25 (2014)
7. Arellano, D., Schaller, U.M., Rauh, R., Helzle, V., Spicker, M., Deussen, O.: Intelligent Virtual Agents: 15th International Conference, IVA 2015, Delft, The Netherlands, August 26-28, 2015, Proceedings, chap. On the Trail of Facial Processing in Autism Spectrum Disorders, pp. 432–441. Springer International Publishing, Cham (2015)
8. Baio, J.: Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, united states, 2010. In: Surveillance Summaries, vol. 63, pp. 1–21 (2014)
9. Bermond, B., Vorst, H.C., Vingerhoets, A.J., Gerritsen, W.: The amsterdam alexithymia scale: Its psychometric values and correlations with other personality traits. Psychotherapy and Psychosomatics **68**(5), 241–251 (1999)
10. Bird, G., Cook, R.: Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. Translational psychiatry (3) (2013)
11. Bird, G., Press, C., Richardson, D.: The role of alexithymia in reduced eye-fixation in autism spectrum conditions. Journal of autism and developmental disorders **41**(11), 1556–1564 (2011)
12. Chita-Tegmark, M.: Attention allocation in asd: a review and meta-analysis of eye-tracking studies **3**, 209–223 (2016)
13. DeCarlo, D., Santella, A.: Stylization and abstraction of photographs. ACM Trans. Graph. **21**(3), 769–776 (2002)
14. Ekman, P., Friesen, W., Hager, J.: The Facial Action Coding System. Weidenfeld & Nicolson, London (2002)
15. Elsabbagh, M., Divan, G., Koh, Y., Kim, Y., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C., Wang, C., Yasamy, M., Fombonne, E.: Global prevalence of autism and other pervasive developmental disorders **5**(3), 160–179 (2012)
16. Gooch, B., Gooch, A.: Non-Photorealistic Rendering. A.K. Peters, Natick, MA (2001)
17. Gooch, B., Reinhard, E., Gooch, A.: Human facial illustrations: Creation and psychophysical evaluation. ACM Trans. Graph. **23**(1), 27–44 (2004)
18. Grawemeyer, B., Johnson, H., Brosnan, M., Ashwin, E., Benton, L.: Developing an embodied pedagogical agent with and for young people with autism spectrum disorder. In: ITS 2012, *LNCS*, vol. 7315, pp. 262–267. Springer, Heidelberg (2012)
19. Helzle, V., Biehn, C., Schlömer, T., Linner, F.: Adaptable setup for performance driven facial animation. In: ACM SIGGRAPH 2004 Sketches, p. 54 (2004)
20. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J. (eds.): Eye tracking: a comprehensive guide to methods and measures. Oxford, UK: Oxford University Press (2011)
21. Kirchner, J.C., Hatri, A., Heekeren, H.R., Dziobek, I.: Autistic symptomatology, face processing abilities, and eye fixation patterns. Journal of autism and developmental disorders **41**(2), 158–167 (2011)
22. 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 **59**(9), 809–816 (2002)
23. Krautheim, B.: The influence of alexithymia on gaze behavior patterns in male adolescents with autism spectrum disorder. Master's thesis, Albert-Ludwigs-Universität Freiburg. Institut für Psychologie (2015)
24. Kupfer, J., Brosig, B., Brhler, E.: TAS-26: Toronto-Alexithymie-Skala-26 (deutsche Version). Manual. Göttingen: Hogrefe (2001)
25. Law Smith, M.J., Montagne, B., Perrett, D.I., Gill, M., Gallagher, L.: Detecting subtle facial emotion recognition deficits in high-functioning autism using dynamic stimuli of varying intensities. Neuropsychologia **48**(9), 2777–2781 (2010)
26. Lozier, L.M., Vanmeter, J.W., Marsh, A.: Impairments in facial affect recognition associated with autism spectrum disorders: A meta-analysis. Development and Psychopathology **26**(4 Pt 1), 933–945 (2014)
27. Milne, M., Luerssen, M., Lewis, T., Leibbrandt, R., Powers, D.: 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 pp. 23–48 (2011)
28. Peirce, J.W.: Generating stimuli for neuroscience using psychopy. Frontiers in Neuroinformatics **2**(10) (2008)

29. Rauh, R., Schaller, U.: Categorical perception of emotional facial expressions in video clips with natural and artificial actors: A pilot study. Tech. Rep. ALU-KJPP-2009-001, University of Freiburg (2009)
30. Rice, L., Wall, C., Fogel, A., Shic, F.: Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with ASD. Journal of Autism and Developmental Disorders pp. 1–11 (2015)
31. Saemundsen, E., Páll Magnússon Ingibjörg Georgsdóttir, E.E.V.R.: Prevalence of autism spectrum disorders in an icelandic birth cohort. BMJ Open **3**(6) (2013)
32. Sawyer, A.C.P., Williamson, P., Young, R.L.: Can gaze avoidance explain why individuals with asperger's syndrome can't recognise emotions from facial expressions? Journal of autism and developmental disorders **42**(4), 606–618 (2012)
33. Senju, A., Johnson, M.H.: Atypical eye contact in autism: models, mechanisms and development. Neuroscience and biobehavioral reviews **33**(8), 1204–1214 (2009)
34. Serret, S., Hun, S., Iakimova, G., Lozada, J., Anastassova, M., Santos, A., Vesperini, S., Askenazy, F.: 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
35. Sifneos, P.: The prevalence of alexithymic characteristics in psychosomatic patients. Psychotherapy and Psychosomatics **22**, 255–262 (1973)
36. Spicker, M., Arellano, D., Schaller, U., Rauh, R., Helzle, V., Deussen, O.: Emotion recognition in autism spectrum disorder: Does stylization help? In: Proceedings of the ACM Symposium on Applied Perception, SAP '16, pp. 97–104. ACM (2016)
37. Tanaka, J.W., Wolf, J.M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M.D., Schultz, R.T.: 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)
38. Uljarevic, M., Hamilton, A.: Recognition of emotions in autism: a formal meta-analysis. Journal of autism and developmental disorders **43**(7), 1517–1526 (2013)
39. Weiß, R.H.: Grundintelligenztest Skala 2 - Revision - (CFT 20-R). Göttingen: Hogrefe (2006)
40. Whyte, E., Smyth, J., Scherf, S.: Designing serious game interventions for individuals with autism. Journal of Autism and Developmental Disorders **45**(12), 3820–3831 (2014)