A Virtual Reality Based System for the Screening and Classification of Autism

Marta Robles*, Negar Namdarian*, Julia Otto*, Evelyn Wassiljew,
Nassir Navab *Fellow, IEEE*, Christine M. Falter-Wagner, Daniel Roth *Member, IEEE*

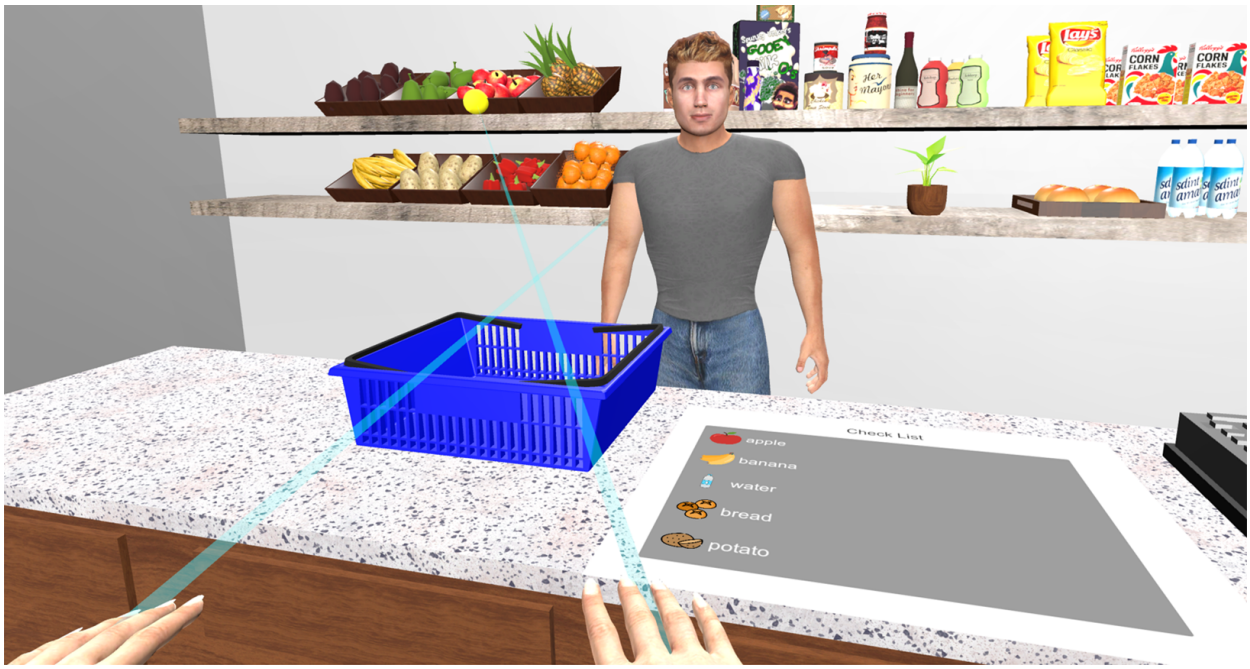

Fig. 1. **Example of the trial phase of the simulation.** A participant, embodied by an avatar, picks an item on the shopping list. After providing the item, the embodied agent representing the shop salesperson describes details and background information on the item, initiating a social interaction, resulting nonverbal responses from the participant. We record these nonverbal behaviors (e.g., head motion, eye gaze, gaze focus) and are able to classify autistic responses with high accuracy.

Abstract—Autism – also known as Autism Spectrum Disorders or Autism Spectrum Conditions – is a neurodevelopmental condition characterized by repetitive behaviours and differences in communication and social interaction. As a consequence, many autistic individuals may struggle in everyday life, which sometimes manifests in depression, unemployment, or addiction. One crucial problem in patient support and treatment is the long waiting time to diagnosis, which was approximated to seven months on average. Yet, the earlier an intervention can take place the better the patient can be supported, which was identified as a crucial factor.
We propose a system to support the screening of Autism Spectrum Disorders based on a virtual reality social interaction, namely a shopping experience, with an embodied agent. During this everyday interaction, behavioral responses are tracked and recorded. We analyze this behavior with machine learning approaches to classify participants from an autistic participant sample in comparison to a typically developed individuals control sample with high accuracy, demonstrating the feasibility of the approach. We believe that such tools can strongly impact the way mental disorders are assessed and may help to further find objective criteria and categorization.

Index Terms—Virtual reality, autism, machine learning, agents, embodiment, diagnosis

1 Introduction

Autism (ICD-11 6A02) is an entity used to define a set of persistent symptoms throughout the life cycle, characterised by (1) differences in communication and reciprocal social interaction and (2) the presence of repetitive behaviors and restricted interests [4, 52]. Concretely, the nonverbal communication skills of autistic individuals are particularly different compared to those of typically developed (TD) individuals. Nonverbal communication includes aspects such as initiating, maintaining or modulating gaze during a social interaction, modulating one's tone of voice when speaking or using gestures to accompany speech, among other features. The causes of this condition remain unclear (for a recent review, see [28]) but eye gaze patterns in autism have been pointed out as possible biomarkers of this condition [11, 16, 18]. More

- *Authors contributed equally to the project.
- *Marta Robles, Christine Falter-Wagner, and Evelyn Wassiljew are with the Department of Psychiatry and Psychotherapy, Medical Faculty, LMU Munich. Marta Robles is also with Department of Clinical and Health Psychology, Autonomus University of Barcelona (UAB).*
- *Nassir Navab, Negar Namdarian, and Julia Otto are with TU Munich.*
- *Daniel Roth is with Friedrich-Alexander-Universität Erlangen-Nürnberg, Department Artificial Intelligence in Biomedical Engineering, and is the corresponding author (d.roth@fau.de).*

Manuscript received 6 Sept. 2021; revised 3 Dec. 2021; accepted 7 Jan. 2022. Date of publication 16 Feb. 2022; date of current version 29 Mar. 2022. Digital Object Identifier no. 10.1109/TVCG.2022.3150489

concretely, eye-tracking has been used to investigate the gaze patterns of autistic individuals (for a systematic review and meta-analysis see [54]). The results of the mentioned review reveal that children with autism spectrum disorders (ASD have significantly reduced gaze fixation to the eye region of faces, when compared to TD individuals. Autism is one of the most prominent and widely discussed human conditions [31]. Despite of studies showing that diagnosis can be reliably established from the age of 2 years [43], many people with autism remain without a diagnosis, unrecognized, until adulthood [37]. Moreover, the average time to diagnose autism in adults has been estimated to 7 months [55], leading to long waiting lists and time. Further, the reliability of the common assessments for the diagnosis of autism, such as the Autism Diagnostic Observation Schedule (ADOS-2; [42], seems to be lower in adulthood [17, 44]. Not surprisingly, the diagnostic of autism in adulthood is one of the ten priority areas for autism research as published by Autistica [14]. Thus, there is a need for an objective measure to provide with a reliable and time economic diagnostics of ASD in adulthood.

1.1 Contribution

We present a system combining an agent-induced social virtual reality (VR) interaction with nonverbal behavior recording and pattern classification. We believe our approach could aid the diagnosis of autism and argue that it could be adapted in the future to also assist the screening of other social and communicative conditions or disorders. Our results are promising regarding the successful classification of autism based on a machine learning model trained and tested with the recorded data.

2 Background and Related Work

2.1 Virtual Environments and Autism Research

The use of VR technologies in autism research and therapy has grown in recent years, due to the strong level of experimental control. To date, VR has often been applied to interventions for children and adolescents on the autistic spectrum. For example, in the context of social communication, interaction and skill training [47, 57, 72], the training of emotion recognition, facial expression, as well as body gestures [22], phobia interventions [45, 46], the practice of fine motor skills [71] and driving exercises [66], (see [9] for a recent review). In populations with neurodegenerative diseases [53] or attention deficit hyperactivity disorder (ADHD) [2, 3], VR technologies have also been used as a tool to aid the evaluation processes to diagnose these conditions. Nowadays, the diagnosis is usually made in the clinic based on visible clinical signs and symptoms, and patients often have to wait for years for a correct diagnosis [53]. In the case of autism, this can be even more difficult, as the clinical heterogeneity of this condition is well known [10, 48]. This frequently leads to long evaluation processes including patients having to visit different experts, misdiagnosis and improper treatment [5]. Of course, these aspects have an impact on the patient's mental health [38].

In terms of ASD screening, few studies focused on the use of modern or novel technologies for ASD assessment. Koirala and colleagues [36] were the first to explore sensory abnormalities in ASD children with VR technology, whereas the automatic detection of ASD individuals revealed preliminary significant results in their study. VR in particular has shown an enormous contribution in clinical populations, in which eye-tracking on its own was only made possible to a limited extent [13]. Compared to regular VR devices, as well as eye-tracking technologies, immersive VR provides ecological validity in controlled environments by enabling a natural experience for the participants and therefore more reliable data collection [50].

2.2 Machine Learning-Based Autism Investigations

In medical context, machine learning (ML) has successfully been applied to objectively diagnose many different kinds of diseases and disorders, skin cancer [20] and heart diseases [56] being only two examples. In recent years, there has been a growing development of computer-aided investigations of ASD through ML on the basis of static images (e.g., [21, 32, 40, 41, 70]). Further, interpersonal synchronicity [23] has been investigated and classified with real-world motion data using motion energy analysis with a classification accuracy of 75.9%. Drimalla and colleagues [18] demonstated that a classification of facial behavior recorded from a video-based simulated dialogue study led to 73% accuracy in detecting ASD. Similarly, Yaneva and colleagues [69] could also detect autism automatically with around 74% accuracy. However, their approaches were based on a simulated interaction with a pre-recorded video [18] and in web page searches [69], which may not be fully capable to account for the full dynamics of social interactions.

In this regard, previous works presented potential methods and concepts to assess ASD using virtual characters [24, 61]. Further, specific study platforms for a potential behavior investigation have been developed [23, 59, 63]. In a recent study, Roth and colleagues [58] could automatically classify autistic individuals from a sample of ASD individuals and TD participants with up to 92.9% accuracy using a neural network trained from nonverbal cues from eye gaze and head movements recorded from avatar-mediated, dyadic social interactions in a desktop environment. While Georgescu and colleagues used motion analysis technologies to classify behaviors of real interactions [23], Roth and colleagues tracked the behavior of interactions that happened between two real people that were remotely tracked and represented to each other as avatars on desktop screens [58]. In contrast to these works, our goal was to implement a single user VR scenario that could allow to collect nonverbal data automatically, replicable, with high validity and experimental control.

3 Approach and Implementation

3.1 Virtual Environments and Scenarios

In order to create a virtual environment suitable for ASD assessment in VR, we identified different social settings that could be used for standardized social interactions. We used Autodesk 3ds Max[1], Blender[2], and the native tools in Unity 3D[3] to create our virtual environment. Some of the 3D models were acquired from Sketchfab[4].

Following a design discussion with clinical partners, we decided to implement a everyday life situation and standardized tasks in a virtual supermarket. For individuals with autism, shopping is a challenging daily living skill. When faced with an unfamiliar environment, such as at the supermarket, it was shown that diagnosed individuals show altered behaviors and affect [1]. Therefore, we anticipated different behavior from individuals with ASD while engaging in this simulation. Fig. 3 shows the final version of the shop used in our user study.

A social setting was considered essential for eliciting authentic nonverbal responses from participants during the simulation. Therefore, we created a virtual agent to act as the social partner in the role of the shop seller. As part of the simulation, the participants were instructed to purchase items shown on a shopping list (see Fig. 6). Purchasing the item required asking the seller to deliver it by ray-cast pointing and selection using the HTC Vive Controller. In the case of a correct selection, the shop seller agent would pick up the selected product and put it into the shopping basket. Following this action, the agent was designed to initialize a social interaction by narrating a short story or facts about the sold item accompanied by nonverbal behaviors. The narratives were co-designed with the clinical partners to match the right level between factual information and social engagement.

Taking into account the fact that different users may have varying levels of experience with VR, we created an introductory level within our simulation (see Fig. 2). Participants were presented with series of tutorials which covered the interaction with virtual objects, and how to adjust the volume of their headphones, etc. After finishing the tutorials, participants could see a start button on display, which could take them to the next phase of the simulation (see Fig. 2). Several studies have utilized virtual mirrors in order to increase the perception of embodiment toward a user's avatar [25, 60]. As a second part of the tutorial phase, the participants were therefore exposed to a virtual mirror to increase their awareness toward their presence and avatar within the simulation and to understand that their body behaviors are replicated and thus foster natural responses.

[1] Autodesk, 2020, San Rafael, USA. autodesk.com/products/3ds-max/

[2] Blender foundation, 2020. blender.org

[3] Unity Technologies, 2020, San Francisco, USA. unity.com

[4] Sketchfab, 2020, New York, USA. sketchfab.com

Fig. 2. **Tutorial Phase of the experiment.** Left: Tutorial environment. Center: Controller instructions. Right: Exposure to the virtual mirror.

Fig. 3. **Virtual shop environment.** The final design of virtual supermarket and surrounding scenario used in the study.

Fig. 4. **Virtual characters.**The virtual characters used for female participants (left), male participants (center) and the shop seller agent (right) in the study.

3.2 Avatar and Agent

Previous work reported that the perception of and interaction with virtual characters can be similar to a face-to-face interaction [15]. Therefore, it was argued that virtual characters, used as avatars (i.e., controlled by human behavior [6]) and agents (i.e., controlled by computer algorithms [6]) may act as a method to investigate social interactions in experimentally controlled fashion [24].

We used male and female virtual characters to represent the participants in the simulation accordingly. The avatar's height could be adjusted based on the participant's height. An HTC VIVE Pro head mounted display (HMD) with HTC VIVE controllers and HTC VIVE trackers (see Fig.6) in combination with SteamVR and the Unity Vive Input Utility allowed for the rendering as well as inverse kinematic tracking [62] in order to replicate the user behaviors to the avatar. Therefore, the participant's avatar's body movements corresponded to the participant's movements.

Similarly, we used a male virtual character as representation for the embodied agent (see Fig. 4). All characters were created using Autodesk Character Generator[5]. For the agent's motion and behavior, we integrated an animation state machine to drive the agent's action according to the current simulation status. We used both, Unity's internal animation system with customized keyframe animations (e.g., grabbing the products from the shelves, controlled arm rotations) as well as third party animation clips from the Unity Asset Store[6] and Mixamo[7] in order to construct all varieties of the agent's behavior. Each animation served as a state within the state machine, and various events could trigger the transition between these states. We took into account several factors such as velocity, and range of movement, to ensure smooth transitions.

To realize a more realistic gaze interaction, the virtual agent (seller) was capable of maintaining eye contact with the participant. Gaze shifting toward the participant involves eyes, head, and upper body movements, and eye, eyelid, blinking, as well as head animations and realistic lip-synced mouth movements were realized using a natural motion plugin (SALSA LipSync Suite Version 2.5.0.).

For the agents verbal discourse, a natural human voice was recorded for the verbal interaction. In order to have a voice that conveyed realistic emotions, the performer adapted his speech in accordance with the narrative.

3.3 Scenario and Logic

The agent's actions are triggered when the user selects an item with HTC VIVE handheld controller. Fig. 5 shows the flowchart of the avatar's actions in relation to the task procedure and status. If the participant selects an item that is not on the shopping list, the agent will ask them to try again. In the case that participants select an item on the list, the agent would bring the products to them. Following this, the agent will narrate a short story or fact about the sold item for the purpose of initiating social interaction. The participant could not select any other item while the agent is narrating the story. Once the final product is delivered, the agent will request payment. For payment, the participant has to drag a pack of cash visible on the counter towards the cash register and drop it there. After successful payment, the game will end with the agent saying goodbye to the participant.

3.4 Data Acquisition and Logging

Our system is designed to collect and log data during the social setting of the shopping scenario, which is the period in which the seller agent narrates a story for the participant. Since our research focused mainly on gaze, head, and body motion comparison between individuals with ASD and TD controls, we collect the body movements and eye gaze data of the participants. The body movement data include the position and rotation of head and hands. The gaze data was collected logging the gaze focus point along each axis (i.e., the gaze focus point in the 3D world) as well as the dwell time the participants focused on dynamic AOIs, virtually attached to landmarks of the agents face and hands, see Fig. 7. Once the raycast hits one of the colliders, the AOI, which was looked at by the participant, is detected. To prevent repetition, once a collision with an inner AOI (such as the eyes) is logged, the larger

[5]https://charactergenerator.autodesk.com/
[6]www.assetstore.unity.com
[7]www.mixamo.com

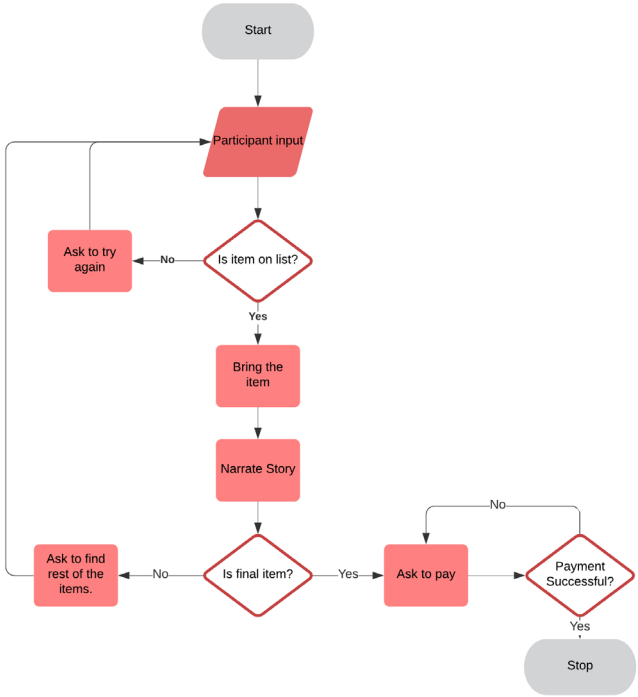

Fig. 5. **Scenario procedure.** The flowchart describes the simulation logic and user actions/agent reactions.

Fig. 6. **Embodiment method.** HTC VIVE Pro Eye and HTC VIVE trackers that allow for the embodiment of the user in the simulation by using inverse kinematics and body pose solving.

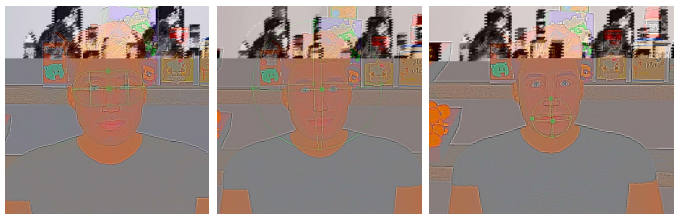

Fig. 7. **Dynamic areas of interest.** Dynamic head, eyes, and mouth AOIs assigned to the agent in order to acquire the focus dwell times.

area (such as the whole face) is ignored. During the virtual reality scenario a total of 27 features, including hand, head, and gaze behavior, were recorded using the HTC VIVE Pro Eye VR System and the Tobii XR SDK, respectively. We exported the acquired data into a CSV file format for analysis and further processing with Python and using Scikit-learn and tensorflow as machine learning libraries.

4 Evaluation

Based on previous research findings regarding the gaze pattern as a valid predictor for ASD [54] and based on our previous work from [58], it was suggested that *H1: there is a difference in mean fixation times on the eyes area, mouth area and the background during social interaction with a virtual character between ASD and TD individuals.* Our main research question was, in consequence, *RQ1: Can we classify ASD on the basis of the expressed nonverbal behavior (gaze, voice, head motion) acquired through an patient-agent system?*

4.1 Design and Task

For the data acquisition, we employed a between-group design and tested individuals with ASD vs. TD control participants. We used the virtual environment to simulate a social situation in which the nonverbal data from participants can be acquired and recorded to provide reliable data set for machine learning algorithms. As part of the simulation, the participants were instructed to purchase items shown on a list in the virtual environment representing a supermarket by pointing to and selecting the product in the shop using HTC VIVE handheld controllers via the controller's trigger button. This was available with both controllers accounting for different handedness of the participants. In the case that participants made the correct selection, the agent would bring the products to them and narrate about the sold item. An example would read as follows:

"Oh, these bananas are great. Did you know that bananas are rich in minerals such as magnesium, potassium and folic acid? Also, bananas are rich in vitamins B and C. And I tell you something: These bananas come from Ecuador. Ecuador has the perfect climate for its cultivation and exports every year more than six million bananas to all countries of the world."

During this time, the integrated eye-tracking system would gather data of the participant's eye gaze, and the transformation data from the head mounted display and controllers was recorded. The participants had to buy five items, which were presented in the same order for all participants. The narratives about the items were neutral facts and had a duration of about 90 seconds on average. Each data file collected during the scenario would then contain approximately 1850 measurement rows containing position, rotation and gaze values as well as the corresponding area of interest at the given moment during the simulation.

Taking into account the fact that different users might have varying levels of experience with VR, we created an introductory level within our simulation. To accomplish this, we designed a virtual space similar to an entrance to a market, to provide participants with the opportunity to become familiar with virtual technologies and understand how controllers work. An additional goal of the introductory level was to increase participants' sense of embodiment. During the simulation, participants were able to control an avatar representation of themselves. A study by Slater and Steed [64] showed that participants who interacted with virtual objects via a virtual body had a higher sense of presence in comparison to those using a traditional interface (like pressing a button) as a means of interaction. Virtual embodiment can lead to psychological effects such as increased social presence in users that control the avatar [60, 65]. In an attempt to evoke virtual embodiment, we considered the virtual mirror metaphor in our design. In this metaphor, users can see a simulated mirror reflection of their avatar. Several studies have used and tested virtual mirrors. A study by González-Franco and colleagues [25] concluded that seeing the avatar reflection of oneself in a virtual mirror, while the movements are synchronous with the user, would result in a higher subjective sense of embodiment. Assuming that a greater perception embodiment would also result in more natural behavioral responses, we implemented a mirror in the introductory part of the virtual simulation.

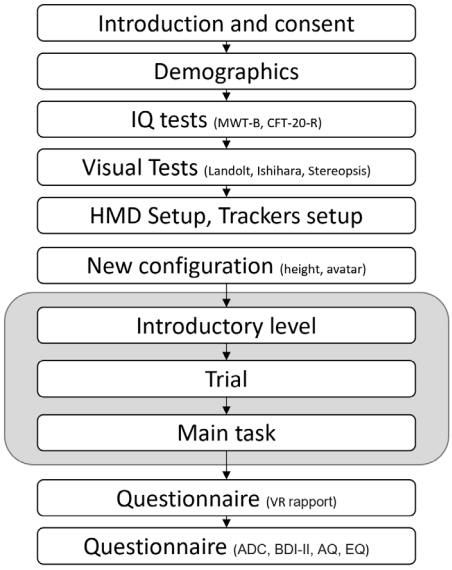

Fig. 8. **E**xperimental procedure.

4.2 Procedure

The experimental session can be roughly divided into three phases. First, the participants answered a batch of psychological and optical questionnaires and tests, to confirm the inclusion criteria of the participants (including a Landolt C-test, Ishihara Color test, and Steropsis test). Then, an eye calibration was performed and the VR simulation was conducted. We informed participants about potential cybersickness and that asked them to immediately notify investigators. The VR task lasted about 20-30 minutes. After the participants completed the task, they completed further questionnaires related to the VR task as well as psychological questionnaires. The session lasted for approximately 90 to 120 minutes. The participants were each compensated financially for taking part in the study. All participants gave written informed consent before study participation. The study was approved by the ethics committee of the Ludwig-Maximilians University Munich Hospital, in agreement with the Declaration of Helsinki [68]. The full procedure is depicted in Fig. 8.

4.3 Measures

The German versions of the following questionnaires where administrated: a demographic questionnaire, to collect biographical data and previous VR experience; the Multiple Choice Vocabulary Intelligence Test (verbal intelligence; MWT-B; [39]) and the Basic Intelligence Scale Scale 2 - Revision (non-verbal intelligence; CFT-20-R; [67]) to corroborate inclusion criteria of the participants' IQ; the Landolt C test, the Stereo Optical test [51] and the Ishihara colour-blindness test [12] for ocular and colour deficiencies. In phase three, after the VR task, participants were asked to answer a rapport questionnaire to assess the quality of the interaction with the avatar [35]; the autism-spectrum-quotient (measuring extent of autistic traits; AQ, [8]) and the empathy quotient (to assess empathy; EQ, [7]). In addition, we measured motor difficulties with the Adult Dyspraxia Checklist (ADC) [33] and depressive symptoms with the Beck Depression Inventary (BDI-II) [27]. We will not go into detail of this reporting due to the fact that the underlying research questions are not in the focus of the present study.

4.4 Participants

A total of 28 participants took part in the study. Twenty TD participants were recruited via social networks and acquaintances. Eight individuals with a clinically confirmed diagnosis of ASD were recruited through the specialised autism outpatient clinic of the University Hospital of Munich. We excluded participants who stated or reported being very tired since their gaze paths are likely to be altered, as well as participants who were distracted or did not follow the task instructions, leading to 6 TD control and 2 ASD participants data sets being excluded from the dataset. In addition, one male TD participant had to be excluded as technical issues led to data distortion. Therefore, the final sample was composed by 13 TD (9 female, 4 male, age $M = 23.31$, $SD = 2.39$) and 6 ASD (3 female, 3 male, age $M = 28.83$, $SD = 8.98$) participants ($N = 19$). Descriptive statistics can be found in Table 1.

	ASD	Matched	Random
Age	28.8 (8.9)	23.16 (2.0)	23.5 (2.5)
Verbal IQ	110.0 (5.0)	108.0 (14.2)	104.16 (10.6)
Non-verbal IQ	110.66 (16.45)	114.1 (14.5)	114.1 (13.7)

Table 1. Descriptive statistics: M (SD) of participants data of the matched control and the random control data sets compared to the ASD set.

					95% CI	
	t	df	p	$Cohen's\ d$	$Lower$	$Upper$
Gender	0.00	10.00	1.00	0.00	-1.132	1.132
Age	1.508	10.00	0.163	0.871	-0.341	2.044
Verbal IQ	0.324	10.00	0.753	0.187	-0.952	1.317
Non-verbal IQ	-0.390	10.00	0.705	-0.225	-1.356	0.916

Table 2. T-test results (matched set).

5 Results

5.1 Analysis Strategy

To analyse collected data for a balanced group comparison that is better applicable to machine learning classification approaches, six TD participants (3 female, 3 male, age $M = 23.16$, $SD = 2.0$) were case-wise matched to ASD participants based on age, IQ, and gender, see Table 1. T-test results showed that participants could be matched on the basis of gender and IQ scores, see Table 2). For the age, Levene's test was significant ($p < .05$) suggesting a violation of the equal variance assumption.

Additionally, a separate analysis based on matching the 6 ASD participants with 6 randomly chosen TD participants was conducted (see Table 1). Table 3 shows that the gender and IQ scores in random selection matched between the two groups, but the age did not. In the following we report both, the ASD vs. matched TD control comparison as well as ASD vs. random TD control comparison.

5.2 Descriptive Analysis and Comparisons

As expected, both groups showed differences in the AQ and EQ tests: the autistic individuals and the TD individuals in the matched data set comparison did defer in autistic traits (AQ: $t(10) = 2.09$, $p = .064$, $Cohen's\ d = 1.204$, 95% CI from 0.066 to 2.425) but not to a significant level, however did significantly differ in empathy skills (EQ: $t(10) = -2.48$, $p = .033$, $d = -1.429$, 95% CI from -2.692 to -0.113). In the random data set comparison, the groups deferred significantly in both constructs (AQ: $t(10) = 2.70$, $p = .022$, $d = 1.557$, 95% CI from 0.212 to 2.846; EQ: $t(10) = -2.43$ $p = .035$, $d = -1.404$, 95% CI from -2.661 to -0.093). Analysis of nonverbal behavior data collected during the monologues of the VR simulation was compared in between the two groups for both the matched and the random data set. As expected, the differences in gaze behavior were significant regarding the average dwell time on the eyes (matched group

					95% CI	
	t	df	p	$Cohen's\ d$	$Lower$	$Upper$
Gender	0.542	10.00	0.599	0.313	-0.834	1.445
Age	1.398	10.00	0.192	0.807	-0.395	1.973
Verbal IQ	1.210	10.00	0.254	0.699	-0.489	1.854
Non-verbal IQ	-0.400	10.00	0.698	-0.231	-1.361	0.911

Table 3. T-test results (random set).

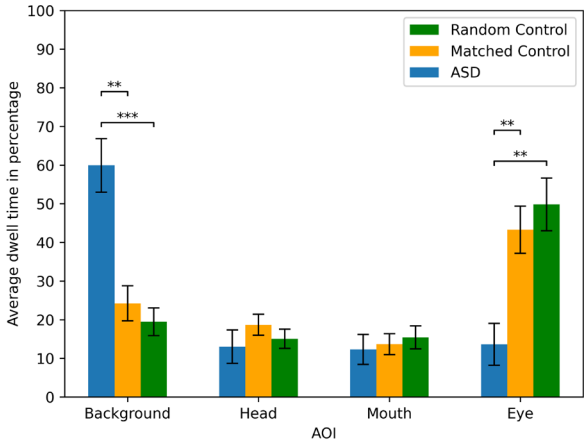

Fig. 9. **AOI dwell times.** The average dwell time on AOIs of the matched and the random set of participants. Note: Graphs denote $M \pm SE$. Asterisks denote significant differences resulting from Student independent t-tests. * indicate $p-value < .05$, ** indicate $p-value < .01$, *** indicate $p-value < .001$

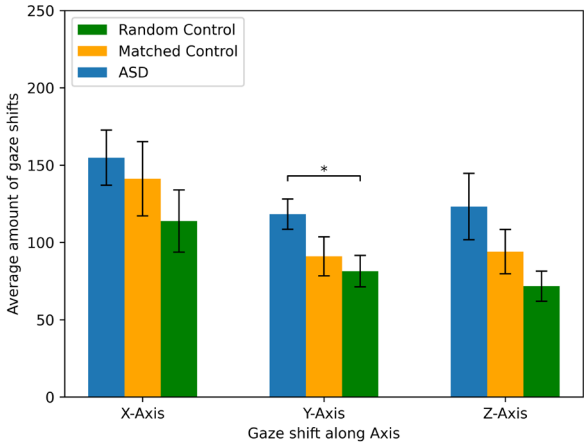

Fig. 10. **Gaze shifts.** The averages of the absolutes of the sum of gaze shifts of the focus point in 3D space *relative* to the head orientation (i.e., local translation) of the matched and the random set of participants. Note: Graphs denote $M \pm SE$. Values represent vector values not axis rotations. Z equals to the view direction, Y is the up axis. Asterisks denote significant differences resulting from Student independent t-tests. * indicate $p-value < .05$, ** indicate $p-value < .01$, *** indicate $p-value < .001$

comparison: $t(10) = -3.64$, $p = .005$, $d = -2.103$, 95% CI from -3.523 to -0.620; random group comparison: $t(10) = -4.17$, $p = .002$, $d = -2.405$, 95% CI from -3.909 to -0.837) and the background (matched group comparison: $t(10) = 4.32$, $p = .002$, $d = 2.494$, 95% CI from 0.900 to 4.025; random group comparison: $t(10) = 5.19$, $p < .001$, $d = 2.997$, 95% CI from 1.246 to 4.685), see Fig. 9. Results suggest a gaze shift towards the background for the ASD participants, whereas their focus on the eye region is reduced. However, we did not find the expected longer focus on the mouth region ($p > .05$). The differences in gaze shifts of the focus point in head-relative 3D space were not significant regarding X-axis (matched group comparison: $t(10) = 0.46$, $p = .657$, $d = 0.265$, 95% CI from -0.879 to 1.395; random group comparison: $t(10) = 1.53$, $p = .157$, $d = 0.883$, 95% CI from -0.331 to 2.057), Y-axis for the matched group comparison ($t(10) = 1.71$, $p = .118$, $d = 0.987$, 95% CI from -0.244 to 2.175) but significant for the random group comparison ($t(10) = 2.612$, $p = .026$, $d = 1.508$, 95% CI from 0.174 to 2.786), and again, non-significant for Z-axis (matched group comparison: $t(10) = 1.13$, $p = .286$, $d = 0.650$, 95% CI from -0.531 to 1.801; random group comparison: $t(10) = 2.18$, $p = .054$, $d = 1.259$, 95% CI from -0.022 to 2.490), see Fig. 10. Finally, the differences in head rotation were not significant regarding X-axis for the matched group comparison ($t(10) = 1.08$, $p = .305$, $d = 0.625$, 95% CI from -0.553 to 1.774), but significant for the random group comparison ($t(10) = 2.32$, $p = .043$, $d = 1.342$, 95% CI from 0.044 to 2.587) and non-significant for Y-axis (matched group comparison: $t(10) = 0.76$, $p = .468$, $d = 0.436$, 95% CI from -0.722 to 1.573; random group comparison: $t(10) = 1.84$, $p = .096$, d = 1.062, 95% CI from -0.182 to 2.260, and Z-axis (matched group comparison: $t(10) = 0.17$, $p = .871$, $d = 0.096$, 95% CI from -1.038 to 1.226; random group comparison: $t(10) = 0.69$, $p = .509$, $d = 0.396$, 95% CI from -0.758 to 1.531), see Fig. 11.

5.3 Preprocessing

Data files collected during the experiment were preprocessed. Foremost, we removed any invalid data due to tracking (system) errors. Invalid data for example arises when the tracker can not detect eye movement. The average amount of invalid data detected during the testing phase of the simulation and also the user study is less than 10% of the collected data frames per simulation (ASD: 6.7%, matched controls: 8.87%, random controls: 6.76%). We removed this data from the dataset. To validate approaches of previous works [23, 58], we transformed gaze vectors to present the gaze shift in local coordinate space and calculated gaze averages and further calculated averages for all features.

A similar approach was followed for the time series (i.e., individual monologue) based data set that were analyzed using a LSTM classification approach. We separated each monologue during the simulation and calculated the averages. An overview of the preprocessing is provided in Fig. 12. Most machine learning algorithms have difficulty handling largely varying scales of input features. Therefore, we scaled the data for all features in a limited range by min-max scaling, as it transforms all values to the range $[0, 1]$, which is the expected input for most neural network algorithms [26].

5.4 Classification

We used similar parameters for the logistic regression, support vector machine, and neural network than previous work. We further implemented an LSTM based on the data of the individual monologues. We chose a sigmoid activation function and binary cross entropy as loss function, and a stochastic gradient descent (SGD) as optimizer. The training set consisted of 80% of the available data while the test set evaluated for validation contains the remaining 20%. Both sets contained an equal amount of TD control and ASD sample data. The machine learning pipeline is depicted in Fig. 13.

5.4.1 Validation of Previous Findings

In order to quantify results of previous work and compare results of this thesis, previously implemented algorithms of a similar setting are tested on the new data [58]. In named study, autism is classified through application of three different types of machine learning algorithms, including logistic regression, a support vector machine and a neural network.

Each algorithm was evaluated applying 5-fold cross validation.

The logistic regression model was trained on all extracted features. Results from training the model on data collected during the user study, reveal an average accuracy of 80% ($SD = 0.4$), sensitivity of 80% ($SD = 0.4$) and specificity of 80% ($SD = 0.4$), with $C = 0.5$ and a maximum of 5000 iterations for the matched data set as well as the random data set.

Training the support vector machine on data from the user study achieves an average accuracy of 63.3% ($SD = 0.306$), sensitivity of 80% ($SD = 0.4$) and specificity of 60% ($SD = 0.49$) for the matched

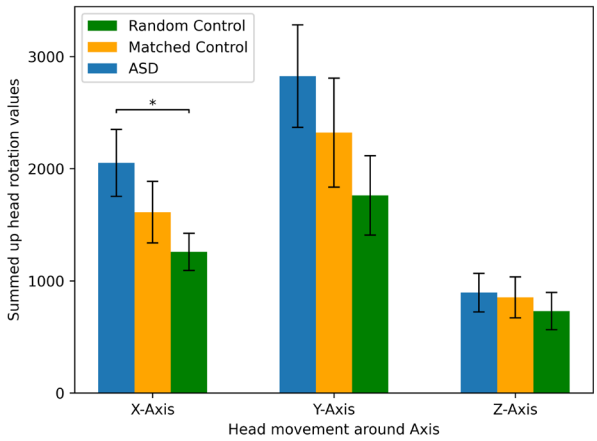

Fig. 11. **Head motion.** Sum of the total head rotation averages per group for each axis. Z equals to the view direction, Y is the up axis (i.e., captures the horizontal head rotation). Note: Graphs denote $M \pm SE$. Asterisks denote significant differences resulting from Student independent t-tests. * indicate $p-value < .05$, ** indicate $p-value < .01$, *** indicate $p-value < .001$

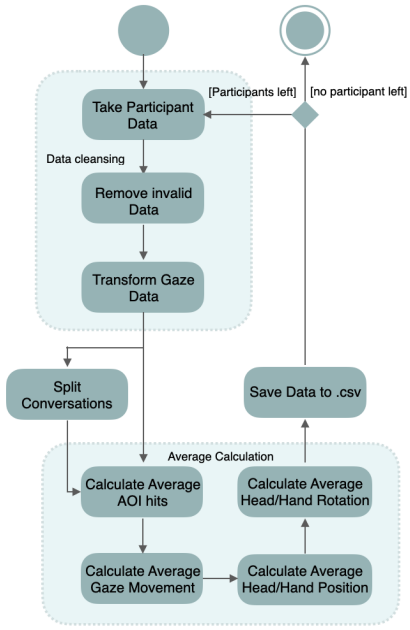

Fig. 12. **Preprocessing.** Overview of the data preprocessing including exclusion of invalid data and average calculation.

data set and an accuracy of 80% ($SD = 0.4$), sensitivity of 80% ($SD = 0.4$) and specificity of 80% ($SD = 0.4$) for the random data set. Parameters for the matched set were $\gamma = 0.001$, $C = 9.9$ and $\gamma = 0.009$, $C = 9.9$ for the random set.

An neural network consisting of one hidden layer achieved an accuracy of 76.7% ($SD = 0.29$), sensitivity of 80% ($SD = 0.40$) and specificity of 70% ($SD = 0.40$) for the matched data set with $hidden\_layer\_size = 5$. An accuracy of 93.30% ($SD = 0.13$), sensitivity of 80% ($SD = 0.40$) and specificity of 1.0 ($SD = 0.00$) for the random data set is achieved with $hidden\_layer\_size = 6$. On the other hand, an average accuracy of 86.70% ($SD = 0.16$), sensitivity of 70% ($SD = 0.40$) and specificity of 100% ($SD = 0.00$) is achieved with $hidden\_layer\_size = (6,21)$ for user study data matched set is achieved by an ANN consisting of two hidden layers and an accuracy of 93.30% ($SD = 0.13$), sensitivity of 80% ($SD = 0.40$), specificity of 100% ($SD = 0.00$) and $hidden\_layer\_size = (7,23)$ is achieved for the random set. Accuracy results are compared in Table 4

Testing previous algorithms on a data set not consisting of averages of the full conversation but by evaluating each conversation of the simulation separately reveals a more defined accuracy (see Table 5). The new data consists of a total of 60 sets instead of 12 as the simulation consists of five monologues.

5.4.2 Classification Using a LSTM

An LSTM network consisting of a single hidden layer was implemented. The best parameters for learning rate (0.1) and epochs (250) were chosen by test and evaluation. The LSTM achieved an accuracy of 100%, sensistivity of 83.0% and a specificity of 99.1% and 98.9% on all features, equally for the matched and random data set.

Feature Analysis In order to reduce computing time and increase accuracy, it is beneficial to look at different smaller feature combinations separately instead of evaluating the LSTM on all features at once. Some features might be stronger indicators for autism than others and some might not show a significant difference between the two groups. Those could in turn be left out of the calculation, saving computing time in the process. Accuracy is calculated for all possible two and three feature combinations of the 27 features of the collected data. The accuracy the algorithm can achieve is calculated for each single feature as well. In Table 6 and Table 7 some examples of feature combinations of two and three for the matched data set can be seen. Evaluation on the random data set shows similar results, see Table 8 and Table 9. Combinations of features that include one of the AOIs in general perform better than combinations of features only consisting of position or rotation features. However, there is no clear combination winner in the two evaluations taking into account two features. Evaluation of one single feature achieved the highest accuracy for Background and Eye (see Table 10) for the matched data set and for only Background for the random data set.

Multilayer LSTM A multilayer neural network is expected to improve accuracy compared to a single layer network for large amounts of data and may reduce over-fitting. As the amount of data in this study is limited, the outcome may not pose a significant difference. Including a second layer in the model resulted in an accuracy of 100%, a sensitivity of 82.9% and a specificity of 98.7% for classifying autism correctly on the matched and an accuracy of 100%, a sensitivity of 81.4% and a specificity of 93.1% for the random data set, with $learning\_rate = 0.1$ and $number\_of\_epochs = 250$ for both sets and thus even under-performed the single layer approach regarding the

	ASD vs. Matched TD	ASD vs. Random TD
Logistic Regr.	0.80 (0.40)	0.80 (0.40)
SVM	0.80 (0.31)	0.80 (0.40)
MLP 1 Layer	0.77 (0.29)	0.93 (0.13)
MLP 2 Layer	0.87 (0.17)	0.93 (0.13)

Table 4. Accuracy M (SD) for each approach based on the evaluation of the averages of the full data set.

	ASD vs. Matched TD	ASD vs. Random TD
Logistic Regr.	0.93 (0.10)	0.93 (0.10)
SVM	0.93 (0.10)	0.95 (0.10)
MLP 1 Layer	0.97 (0.04)	0.97 (0.04)
MLP 2 Layer	0.98 (0.03)	0.98 (0.03)

Table 5. Accuracy M (SD) for each approach based on the evaluation of the averages of each monologue.

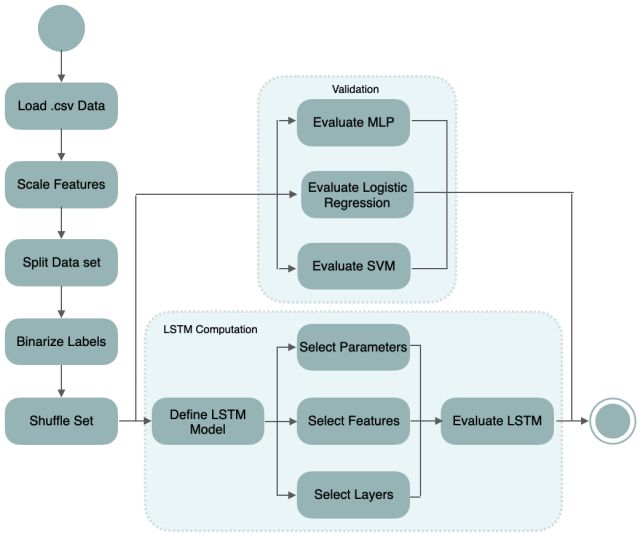

Fig. 13. **Classification procedure.** Representation of the machine learning pipeline depicting data preparation and evaluation of applied algorithms.

Features	Accuracy
(Background, Eye)	1.0
(Head, Gaze Vector Y)	1.0
(Mouth, Gaze Vector Y)	0.93
(Head Position Z, Left Hand Position Z)	0.87

Table 6. Accuracy of classification evaluating on possible two feature combinations for matched data set.

key performance indicators. However, we expect this approach to be more beneficial with larger data sets and the performance achieved can be considered sufficient for an assistive tool.

6 Discussion

In this paper, we present a VR-based system that could act as an assistive screening tool in objectively supporting the diagnosis of ASD and report on its evaluation. The simulation exposes the user to a social situation with an embodied virtual agent. We record resulting nonverbal behavior parameters of the user and use those to perform a classification judgement. With different approaches, we could reach accuracies up to 100% with our (limited) samples. While the sample size limits the generalizability of the approach due to the lack of a large general population representation, we interpret these results as very promising. Our procedure is not invasive, and we assume that, when put into practice, a screening test with this procedure would not take more than approximately 45 minutes till a result could be obtained.

Our subjacent research goal was to investigate whether and to what degree VR technology can support the ASD detection process by automatically distinguishing behavior and gaze characteristics between TD individuals and individuals with autism through VR and nonverbal pattern analysis/pattern classification techniques. In this regard, we are the first to show the successful implementation of a VR driven

Features	Accuracy
(Background, Mouth, Head Position X)	1.0
(Background, Eye, Gaze Vector Y)	1.0
(Left Hand, Mouth, Eye)	1.0
(Head Position Z, Gaze Vector Z, Right Hand Position X)	0.87

Table 7. Accuracy of classification evaluating on possible three feature combinations for matched data set.

Features	Accuracy
(Background, Head)	1.0
(Eye, Gaze Vector Y)	1.0
(Gaze Vector Y, Right Hand Rotation X)	0.93
(Head Position Z, Left Hand Rotation X)	0.87

Table 8. Accuracy of classification evaluating on possible two feature combinations for random data set.

Features	Accuracy
(Background, Left Hand, Mouth)	1.0
(Background, Head, Gaze Vector Y)	1.0
(Background, Head, Eye)	1.0
(Head Position Y, Left Hand Rotation Y, Left Hand Rotation Z)	0.87

Table 9. Accuracy of classification evaluating on possible three feature combinations for random data set.

screening tool and thus argue that found supporting results indicating a positive answer to our *RQ 1: Can we classify ASD on the basis of the expressed nonverbal behavior (gaze, voice, headmotion) acquired through an patient-agent system?*, although not without limitations.

Our results show strong differences in the mean fixation times in the *eye region* of the virtual agent and the *background region* of the grocery shop can be observed between participants with autism and TD individuals. Contrary to previous studies [49, 58] that showed differences in focus times of the mouth area, we did not observe a significant difference in this measure. One interpretation may be that in relation, the overall focus of ASD participants was mainly the background area such that the head and head area generally did not receive much attention at all, because sufficient context cues where available to avoid this area completely. We can thus only partially support *H1: there is a difference in mean fixation times on the eyes area, mouth area and the background during social interaction with a virtual character between ASD and TD individuals*. However, we implemented and confirmed other indicators that can contribute to the screening, such as the head position and the gaze vector in 3D space, that confirms the results found in a previous study [58]. To this end, non-verbal gaze behavior has been shown to be a notable factor in the recognition of autistic features and has been investigated for many years [19, 34]. Previous works investigated a desktop-based virtual environment prototype that classifies autism and thus revealed significant results in distinguishing between the two groups of adult subjects with respect to their gaze pattern, with high categorisation accuracies (up to 92.9%) [58]. Yet, previous works have been using either a a) a still picture based assessment or b) a dyadic interaction assessment, which may a) not account for the full and subtle dynamics of social interaction or b) require two participants or one participant and a therapist to be part of the procedure. With the present system and study we could substitute one participant posed by an avatar with an embodied virtual agent and minimise the setting requirements to a one person configuration, maintaining and confirming the previous works' performance and increase the level of accuracy and other key performance indicators. In comparison to previous work, we could also use a large percentage of the data collected without invalid data points, since the tracker is fixed to the head and integrated in the HMD, accounting for changes in head orientation that would lead to errors with regular desktop trackers.

A recent review suggest that receiving an autism diagnosis has a significant emotional impact on adults and that accessibility and processes are inconsistent [30]. Moreover, earlier diagnosis could prevent secondary mental health problems in this population [29]. We believe that the present study could assist this processes and improve non-objective and time consuming standard assessments. This study also contributes to the field of diagnosis research evidences, one of the ten priority areas for autism research [14]. We believe that a tool, such as ours, could not only be extended to include a broader population, but also to distinguish and identify other social and communicative disorders, such as Borderline or Schizophrenia, that manifest in differences and

Feature	Accuracy Matched TD controls	Accuracy Random TD controls
Background	1.00	1.00
Eye	1.00	0.73
Head	0.53	0.33
Mouth	0.33	0.33
Left Hand	0.33	0.33
Right Hand	0.33	0.33
Head Position X	0.33	0.33
Head Position Y	0.40	0.40
Head Position Z	0.86	0.33
Head Rotation X	0.33	0.33
Head Rotation Y	0.33	0.33
Head Rotation Z	0.33	0.33
Gaze Vector X	0.33	0.60
Gaze Vector Y	0.93	0.73
Gaze Vector Z	0.80	0.33
Left Hand Position X	0.46	0.46
Left Hand Position Y	0.33	0.33
Left Hand Position Z	0.93	0.66
Left Hand Rotation X	0.33	0.93
Left Hand Rotation Y	0.80	0.46
Left Hand Rotation Z	0.33	0.59
Right Hand Position X	0.33	0.33
Right Hand Position Y	0.60	0.33
Right Hand Position Z	0.33	0.33
Right Hand Rotation X	0.33	0.33
Right Hand Rotation Y	0.33	0.33
Right Hand Rotation Z	0.33	0.33

Table 10. Accuracy of classification evaluating on single feature for the matched TD control and the random TD control data set.

altered patterns of social and nonverbal behavior in everyday life.

In particular, we believe that including a tool such as the proposed one in a screening procedure could substantially i) reduce the waiting time by speeding up the initial procedure and pathway decision process, ii) reduce the patient and medical system costs, and iii) provide additional certainty and assurance to the therapist and is superior to other subjective questionnaire assessments.

6.1 Limitations

Of course, our study and results cannot be blindly generalized and are not without limitations. We recognise that the sample size of the present study is an obvious limitation, that keeps from generalizing the findings to be applicable to diverse screening populations. In addition, the limited sample size poses the risk of introducing overfitting in the neural network. Therefore, the results presented should be interpreted with caution and future research should be conducted to significantly increase the sample size to support and corroborate the results shown. Future recruitment may also consider the diversity of the population included, including patients with disorders that are related in the behavior manifestations. In addition, our TD sample was not screened and therefore we cannot exclude that participant with mental disorders are present in the sample (above or below the average percentage in the general population). However, all TD participants stated that there is no presence of any disorders.

Further, in the present study we only included adult participants with no intellectual impairments. Thereby, future research should both include a wider age range and individuals with different cognitive styles. In this regard however, it is necessary to change and redesign the 3D environment and simulation accordingly. Yet, our scenario and simulation principle offers dynamics to simplify the task or environment to a degree understandable for people with intellectual deficits and children in developing ages. Finally, our agent is not yet capable of initiating and maintaining bidirectional, i.e., interactive, social communication. Future work may include such interaction either by a screen based dialogue and answer selection or by simplified yes and no answers recognized with speech recognition. However, from the current data, we do see that differences can be shown even with the simplified interaction type we implemented.

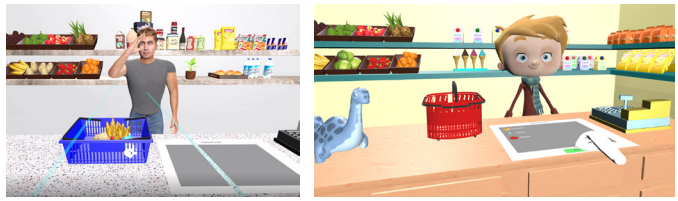

Fig. 14. **Future work.** Left: current environment. Right: Prototype of a potential adaptation of the scenario to younger samples.

6.2 Future Work

In future works, we aim to mitigate our limitations, to increase the sample size and include other disorders that manifest in behavioral differences. In that regard, our goal is to look for deep ML methods to be applied, such as by including the whole time series of the data collections. That would potentially enable another dimension, i.e. frame dependent measures, and may allow distinguishing between subtle differences in disorder types, or classify the severity within one neurodevelopmental disorder. Furthermore, we aim at adjusting our simulation characteristics for other populations, such as children, since it is especially important to have a diagnosis as soon as possible in order to establish interventions. Figure 14 presents a first impression on how style and simulation could be adapted for younger aged populations. The impact of the level of fidelty and naturalness of the communication behavior of the virtual agent leaves room for future endeavors. While our interaction was simplistic, it may be the target of future research to investigate more natural interactions, potentially allowing for a more realistic bidirectional communication and interaction. A more natural communication using a wizard of oz paradigm or a more intelligent agent may even improve the present results and system's performance. To this regard, future work may also consider to vary the number of agents and their proxemics, which was not subject of the current investigation. Rather, the interpersonal distance was chosen by an estimate of what would feel natural and physically sound. Finally, future research should also aim to clinically validate the presented screening tool, pursuing to classify the severity of the disorder.

7 Conclusion

Our proposed VR system for autism classifcation and the presented evaluation results showed that the system is capable of a nonverbal behavior pattern classification between autistic and typically developed individuals with a high accuracy, sensitivity, and specificity. Confirming previous studies, focus behavior as well as gaze movement were strong features. Our system could not only assist diagnostic procedures of autism but be extended and used for the assessment of other communicative and social disorders. We are convinced that our system could be successfully deployed as an assistive tool in the screening and diagnosis procedures to reduce waiting times and costs, as well as to provide an objective method of assessment.

Acknowledgments

The authors wish to thank Prof. Dr. Zhuanghua Shi from LMU Munich for his support during the data collection. We also thank Harry Unwin for creating the audio recordings for the agent.

References

[1] A. Adjorlu, E. R. Hoeg, L. Mangano, and S. Serafin. Daily Living Skills Training in Virtual Reality to Help Children with Autism Spectrum Disorder in a Real Shopping Scenario. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 294–302. IEEE, Nantes, France, Oct. 2017. doi: 10.1109/ISMAR-Adjunct.2017.93

[2] D. Areces, J. Dockrell, T. García, P. González-Castro, and C. Rodríguez. Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool. *PloS one*, 13(8):e0201039, 2018. Publisher: Public Library of Science San Francisco, CA USA.

[3] D. Areces, C. Rodríguez, T. García, M. Cueli, and P. González-Castro. Efficacy of a continuous performance test based on virtual reality in the

diagnosis of ADHD and its clinical presentations. *Journal of Attention Disorders*, 22(11):1081–1091, 2018. Publisher: Sage Publications Sage CA: Los Angeles, CA.

[4] A. P. Association and A. P. Association, eds. *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association, Washington, D.C, 5th ed ed., 2013.

[5] S. K. Au-Yeung, L. Bradley, A. E. Robertson, R. Shaw, S. Baron-Cohen, and S. Cassidy. Experience of mental health diagnosis and perceived misdiagnosis in autistic, possibly autistic and non-autistic adults. *Autism*, 23(6):1508–1518, Aug. 2019. Publisher: SAGE Publications Ltd. doi: 10.1177/1362361318818167

[6] J. Bailenson and J. Blascovich. Avatars.

[7] S. Baron-Cohen and S. Wheelwright. The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. *Journal of Autism and Developmental Disorders*, 34(2):163–175, Apr. 2004. doi: 10.1023/B:JADD.0000022607.19833.00

[8] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley. The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, 31(1):5–17, Feb. 2001. doi: 10.1023/A:1005653411471

[9] C. Berenguer, I. Baixauli, S. Gómez, M. d. E. P. Andrés, and S. De Stasio. Exploring the Impact of Augmented Reality in Children and Adolescents with Autism Spectrum Disorder: A Systematic Review. *International Journal of Environmental Research and Public Health*, 17(17):6143, Jan. 2020. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/ijerph17176143

[10] H. Bruining, L. d. Sonneville, H. Swaab, M. d. Jonge, M. Kas, H. v. Engeland, and J. Vorstman. Dissecting the Clinical Heterogeneity of Autism Spectrum Disorders through Defined Genotypes. *PLOS ONE*, 5(5):e10887, May 2010. Publisher: Public Library of Science. doi: 10.1371/journal.pone.0010887

[11] R. Camero, V. Martínez, and C. Gallego. Gaze Following and Pupil Dilation as Early Diagnostic Markers of Autism in Toddlers. *Children*, 8(2):113, Feb. 2021. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute. doi: 10.3390/children8020113

[12] J. H. Clark. The Ishihara Test for Color Blindness. *American Journal of Physiological Optics*, 5:269–276, 1924.

[13] V. Clay, P. König, and S. U. König. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12(1), Apr. 2019. Number: 1. doi: 10.16910/jemr.12.1.3

[14] J. Cusack and R. Sterry. Your questions: shaping future autism research. Technical report, Autistica, 2016.

[15] A. de Borst and B. Gelder. Is it the real deal? Perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, 6, May 2015. doi: 10.3389/fpsyg.2015.00576

[16] M. Del Valle Rubido, J. T. McCracken, E. Hollander, F. Shic, J. Noeldeke, L. Boak, O. Khwaja, S. Sadikhov, P. Fontoura, and D. Umbricht. In search of biomarkers for autism spectrum disorder. *Autism research*, 11(11):1567–1579, 2018. Publisher: Wiley Online Library.

[17] A. der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). S3-Leitlinie Autismus-Spektrum-Störungen im Kindes-, Jugend- und Erwachsenenalter Teil 1: Diagnostik. Technical report, DGKJP, DGPPN, 2015.

[18] H. Drimalla, T. Scheffer, N. Landwehr, I. Baskow, S. Roepke, B. Behnia, and I. Dziobek. Towards the automatic detection of social biomarkers in autism spectrum disorder: introducing the simulated interaction task (SIT). *npj Digital Medicine*, 3(1):1–10, Feb. 2020. doi: 10.1038/s41746-020-0227-5

[19] M. Elsabbagh, A. Volein, G. Csibra, K. Holmboe, H. Garwood, L. Tucker, S. Krljes, S. Baron-Cohen, P. Bolton, T. Charman, G. Baird, and M. H. Johnson. Neural correlates of eye gaze processing in the infant broader autism phenotype. *Biological Psychiatry*, 65(1):31–38, Jan. 2009. doi: 10.1016/j.biopsych.2008.09.034

[20] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, Feb. 2017. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7639 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Diagnosis;Machine learning;Skin cancer Subject_term_id: diagnosis;machine-learning;skin-cancer. doi: 10.1038/nature21056

[21] D. Fabiano, S. Canavan, H. Agazzi, S. Hinduja, and D. Goldgof. Gaze-based classification of autism spectrum disorder. *Pattern Recognition Letters*, 135:204–212, 2020. Publisher: Elsevier.

[22] S. Fridenson-Hayo, S. Berggren, A. Lassalle, S. Tal, D. Pigat, N. Meir-Goren, H. O'Reilly, S. Ben-Zur, S. Bölte, S. Baron-Cohen, and O. Golan. 'Emotiplay': a serious game for learning about emotions in children with autism: results of a cross-cultural evaluation. *European Child & Adolescent Psychiatry*, 26(8):979–992, Aug. 2017. doi: 10.1007/s00787-017-0968-0

[23] A. L. Georgescu, J. C. Koehler, J. Weiske, K. Vogeley, N. Koutsouleris, and C. Falter-Wagner. Machine Learning to Study Social Interaction Difficulties in ASD. *Frontiers in Robotics and AI*, 6:132, 2019. doi: 10.3389/frobt.2019.00132

[24] A. L. Georgescu, B. Kuzmanovic, D. Roth, G. Bente, and K. Vogeley. The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Frontiers in Human Neuroscience*, 8:807, 2014. doi: 10.3389/fnhum.2014.00807

[25] M. González-Franco, D. Pérez-Marcos, B. Spanlang, and M. Slater. The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 111–114, Mar. 2010. ISSN: 2375-5334. doi: 10.1109/VR.2010.5444805

[26] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. "O'Reilly Media, Inc.", Sept. 2019.

[27] M. Hautzinger, F. Keller, and C. Kühner. *Beck Depressions-Inventar: BDI II.Revision*. Harcourt Test Services, Frankfurt am Main, revision ed., 2006.

[28] H. Hodges, C. Fealko, and N. Soares. Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational Pediatrics*, 9(Suppl 1):S55–S65, Feb. 2020. doi: 10.21037/tp.2019.09.09

[29] M. Hosozawa, A. Sacker, and N. Cable. Timing of diagnosis, depression and self-harm in adolescents with autism spectrum disorder. *Autism*, 25(1):70–78, Jan. 2021. Publisher: SAGE Publications Ltd. doi: 10.1177/1362361320945540

[30] Y. Huang, S. R. Arnold, K.-R. Foley, and J. N. Trollor. Diagnosis of autism in adulthood: A scoping review. *Autism*, 24(6):1311–1327, Aug. 2020. Publisher: SAGE Publications Ltd. doi: 10.1177/1362361320903128

[31] K. Hume, J. R. Steinbrenner, S. L. Odom, K. L. Morin, S. W. Nowell, B. Tomaszewski, S. Szendrey, N. S. McIntyre, S. Yücesoy-Özkan, and M. N. Savage. Evidence-Based Practices for Children, Youth, and Young Adults with Autism: Third Generation Review. *Journal of Autism and Developmental Disorders*, Jan. 2021. doi: 10.1007/s10803-020-04844-2

[32] M. Jiang and Q. Zhao. Learning visual attention to identify people with autism spectrum disorder. In *Proceedings of the ieee international conference on computer vision*, pp. 3267–3276, 2017.

[33] A. Kirby, L. Edwards, D. Sugden, and S. Rosenblum. The development and standardization of the Adult Developmental Co-ordination Disorders/Dyspraxia Checklist (ADC). *Research in Developmental Disabilities*, 31(1):131–139, 2010. doi: 10.1016/j.ridd.2009.08.010

[34] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. 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, Sept. 2002. doi: 10.1001/archpsyc.59.9.809

[35] J. C. Koehler, A. L. Georgescu, J. Weiske, M. Spangemacher, L. Burghof, P. Falkai, N. Koutsouleris, W. Tschacher, K. Vogeley, and C. M. Falter-Wagner. Brief Report: Specificity of Interpersonal Synchrony Deficits to Autism Spectrum Disorder and Its Potential for Digitally Assisted Diagnostics. *Journal of Autism and Developmental Disorders*, July 2021. doi: 10.1007/s10803-021-05194-3

[36] A. Koirala, Z. Yu, H. Schiltz, A. Van Hecke, B. Armstrong, and Z. Zheng. A Preliminary Exploration of Virtual Reality-Based Visual and Touch Sensory Processing Assessment for Adolescents With Autism Spectrum Disorder. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:619–628, 2021. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering. doi: 10.1109/TNSRE.2021.3064148

[37] M.-C. Lai and S. Baron-Cohen. Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, 2(11):1013–1027, Nov. 2015. Publisher: Elsevier. doi: 10.1016/S2215-0366(15)00277-1

[38] A. Leedham, A. R. Thompson, R. Smith, and M. Freeth. 'I was exhausted trying to figure it out': The experiences of females receiving an autism diagnosis in middle to late adulthood. *Autism*, 24(1):135–146, Jan. 2020. Publisher: SAGE Publications Ltd. doi: 10.1177/1362361319853442

[39] S. Lehrl, J. Merz, G. Burkhard, and S. Fischer. *Mehrfachwahl-Wortschatz-*

Intelligenztest: MWT-B. Perimed-Fachbuch-Verlag-Ges., 1989.

[40] W. Liu, M. Li, and L. Yi. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8):888–898, 2016. doi: 10.1002/aur.1615

[41] W. Liu, X. Yu, B. Raj, L. Yi, X. Zou, and M. Li. Efficient autism spectrum disorder prediction with eye movement: A machine learning framework. In *2015 International conference on affective computing and intelligent interaction (ACII)*, pp. 649–655. IEEE, 2015.

[42] C. Lord, R. Luyster, K. Gotham, and W. Guthrie. *Autism Diagnostic Observation Schedule. 2nd. (ADOS-2) Manual (Part II): Toddler Module.* Western Psychological Services, Torrence, CA, 2012.

[43] C. Lord, S. Risi, P. S. DiLavore, C. Shulman, A. Thurm, and A. Pickles. Autism from 2 to 9 years of age. *Archives of General Psychiatry*, 63(6):694–701, June 2006. doi: 10.1001/archpsyc.63.6.694

[44] B. B. Maddox, E. S. Brodkin, M. E. Calkins, K. Shea, K. Mullan, J. Hostager, D. S. Mandell, and J. S. Miller. The Accuracy of the ADOS-2 in Identifying Autism among Adults with Complex Psychiatric Conditions. *Journal of Autism and Developmental Disorders*, 47(9):2703–2709, Sept. 2017. doi: 10.1007/s10803-017-3188-z

[45] M. Maskey, J. Rodgers, V. Grahame, M. Glod, E. Honey, J. Kinnear, M. Labus, J. Milne, D. Minos, H. McConachie, and J. R. Parr. A Randomised Controlled Feasibility Trial of Immersive Virtual Reality Treatment with Cognitive Behaviour Therapy for Specific Phobias in Young People with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 49(5):1912–1927, May 2019. doi: 10.1007/s10803-018-3861-x

[46] M. Maskey, J. Rodgers, B. Ingham, M. Freeston, G. Evans, M. Labus, and J. R. Parr. Using Virtual Reality Environments to Augment Cognitive Behavioral Therapy for Fears and Phobias in Autistic Adults. *Autism in Adulthood*, 1(2):134–145, June 2019. doi: 10.1089/aut.2018.0019

[47] J. Moon and F. Ke. Exploring the treatment integrity of virtual reality-based social skills training for children with high-functioning autism. *Interactive Learning Environments*, 0(0):1–15, May 2019. doi: 10.1080/10494820.2019.1613665

[48] L. Mottron and D. Bzdok. Autism spectrum heterogeneity: fact or artifact? *Molecular Psychiatry*, 25(12):3178–3185, Dec. 2020. doi: 10.1038/s41380-020-0748-y

[49] D. Neumann, M. L. Spezio, J. Piven, and R. Adolphs. Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social Cognitive and Affective Neuroscience*, 1(3):194–202, Dec. 2006. doi: 10.1093/scan/nsl030

[50] N. Newbutt, C. Sung, H. J. Kuo, and M. J. Leahy. The potential of virtual reality technologies to support people with an autism condition: A case study of acceptance, presence and negative effects. *Annual Review of Cyber Therapy and Telemedicine (ARCTT)*, 14, Mar. 2016.

[51] S. Optical. Original Stereo Fly Stereotest.

[52] W. H. Organization. ICD-11, 2019.

[53] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, and H. Devos. Emulation of Physician Tasks in Eye-Tracked Virtual Reality for Remote Diagnosis of Neurodegenerative Disease. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1302–1311, Apr. 2017. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2017.2657018

[54] E. A. Papagiannopoulou, K. M. Chitty, D. F. Hermens, I. B. Hickie, and J. Lagopoulos. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social neuroscience*, 9(6):610–632, 2014. Publisher: Taylor & Francis.

[55] M. Penner, E. Anagnostou, L. Y. Andoni, and W. J. Ungar. Systematic review of clinical guidance documents for autism spectrum disorder diagnostic assessment in select regions. *Autism: The International Journal of Research and Practice*, 22(5):517–527, July 2018. doi: 10.1177/1362361316685879

[56] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv:1711.05225 [cs, stat]*, Dec. 2017. arXiv: 1711.05225.

[57] C. R. Ramachandiran, N. Jomhari, S. Thiyagaraja, and M. Maria. Virtual Reality Based Behavioural Learning for Autistic Children. *undefined*, 2015.

[58] D. Roth, M. Jording, T. Schmee, P. Kullmann, N. Navab, and K. Vogeley. Towards Computer Aided Diagnosis of Autism Spectrum Disorder Using Virtual Environments. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 115–122. IEEE, Utrecht, Netherlands, Dec. 2020. doi: 10.1109/AIVR50618.2020.00029

[59] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik. Effects of Hybrid and Synthetic Social Gaze in Avatar-Mediated Interactions. In *Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018*. IEEE, Munich, Germany, 2018. doi: 10.1109/ISMAR-Adjunct.2018.00044

[60] D. Roth and M. E. Latoschik. Construction of the Virtual Embodiment Questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, Dec. 2020. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2020.3023603

[61] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente. Hybrid Avatar-Agent Technology – A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com*, 14(2):107–114, Aug. 2015. Publisher: Oldenbourg Wissenschaftsverlag Section: i-com. doi: 10.1515/icom-2015-0030

[62] D. Roth, J.-L. Lugrin, J. Büser, G. Bente, A. Fuhrmann, and M. E. Latoschik. A simplified inverse kinematic approach for embodied VR applications. In *2016 IEEE Virtual Reality (VR)*, pp. 275–276, Mar. 2016. ISSN: 2375-5334. doi: 10.1109/VR.2016.7504760

[63] D. Roth, J.-P. Stauffert, and M. E. Latoschik. Avatar Embodiment, Behavior Replication, and Kinematics in Virtual Reality. In *VR Developer Gems*. A K Peters/CRC Press, 2019. Num Pages: 26.

[64] M. Slater and A. Steed. A Virtual Presence Counter. *Presence*, 9:413–434, Oct. 2000. doi: 10.1162/105474600566925

[65] H. J. Smith and M. Neff. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. Association for Computing Machinery, New York, NY, USA, Apr. 2018.

[66] J. Wade, L. Zhang, D. Bian, J. Fan, A. Swanson, A. Weitlauf, M. Sarkar, Z. Warren, and N. Sarkar. A Gaze-Contingent Adaptive Virtual Reality Driving Environment for Intervention in Individuals with Autism Spectrum Disorders. *ACM Transactions on Interactive Intelligent Systems*, 6(1):3:1–3:23, Mar. 2016. doi: 10.1145/2892636

[67] R. Weiß. *Grundintelligenztest Skala 2—Revision (CFT 20-R) [Culture Fair Intelligence Test 20-R—Scale 2].* Jan. 2006.

[68] World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20):2191–2194, Nov. 2013. doi: 10.1001/jama.2013.281053

[69] V. Yaneva, S. Eraslan, Y. Yesilada, and R. Mitkov. Detecting high-functioning autism in adults using eye tracking and machine learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(6):1254–1261, 2020. Publisher: IEEE.

[70] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov. Detecting Autism Based on Eye-Tracking Data from Web Searching Tasks. In *Proceedings of the 15th International Web for All Conference*, W4A '18, pp. 1–10. Association for Computing Machinery, New York, NY, USA, Apr. 2018. doi: 10.1145/3192714.3192819

[71] H. Zhao, Z. Zheng, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar. Design of a Haptic-Gripper Virtual Reality System (Hg) for Analyzing Fine Motor Behaviors in Children with Autism. *ACM Transactions on Accessible Computing*, 11(4):19:1–19:21, Nov. 2018. doi: 10.1145/3231938

[72] Z. Zheng, Q. Fu, H. Zhao, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar. Design of an Autonomous Social Orienting Training System (ASOTS) for Young Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(6):668–678, June 2017. doi: 10.1109/TNSRE.2016.2598727