Bodily Expression of Social Initiation Behaviors in ASC and non-ASC children: Mixed Reality vs. LEGO Game Play

Batuhan Sayis
Department of Information and
Communication Technologies
Universitat Pompeu Fabra
Barcelona, Spain
batuhan.sayis@upf.edu

Narcis Pares
Department of Information and
Communication Technologies
Universitat Pompeu Fabra
Barcelona, Spain
narcis.pares@upf.edu

Hatice Gunes
Department of Computer Science and
Technology
University of Cambridge
Cambridge, United Kingdom
hatice.gunes@cl.cam.ac.uk

ABSTRACT

This study is part of a larger project that showed the potential of our mixed reality (MR) system in fostering social initiation behaviors in children with Autism Spectrum Condition (ASC). We compared it to a typical social intervention strategy based on construction tools, where both mediated a face-to-face dyadic play session between an ASC child and a non-ASC child. In this study, our first goal is to show that an MR platform can be utilized to alter the nonverbal body behavior between ASC and non-ASC during social interaction as much as a traditional therapy setting (LEGO). A second goal is to show how these body cues differ between ASC and non-ASC children during social initiation in these two platforms. We present our first analysis of the body cues generated under two conditions in a repeated-measures design. Body cue measurements were obtained through skeleton information and characterized in the form of spatio-temporal features from both subjects individually (e.g. distances between joints and velocities of joints), and interpersonally (e.g. proximity and visual focus of attention). We used machine learning techniques to analyze the visual data of eighteen trials of ASC and non-ASC dyads. Our experiments showed that: (i) there were differences between ASC and non-ASC bodily expressions, both at individual and interpersonal level, in LEGO and in the MR system during social initiation; (ii) the number of features indicating differences between ASC and non-ASC in terms of nonverbal behavior during initiation were higher in the MR system as compared to LEGO; and (iii) computational models evaluated with combination of these different features enabled the recognition of social initiation type (ASC or non-ASC) from body features in LEGO and in MR settings. We did not observe significant differences between the evaluated models in terms of performance for LEGO and MR environments. This might be interpreted as the MR system encouraging similar nonverbal behaviors in children, perhaps more similar than the LEGO environment, as the performance scores in the MR setting are lower as compared to the LEGO setting. These results demonstrate the potential benefits of full body interaction and MR settings for children with ASC.

1 INTRODUCTION

Sharing information, feelings and thoughts in social interaction is achieved through initiations and responses between individuals and groups [27]. A number of studies have established that high functioning children with Autism Spectrum Condition (ASC) are able to maintain social interactions when an initiation is directed towards them. However, the number of social initiation acts generated by themselves, originating from their own will, is low [58]. Moreover, although recent research has shown that children with ASC can identify emotions in body language just as accurately as their non-ASC peers [49], the accepted standard of non-verbal social behavior can be difficult to understand by children with ASC [12, 17].

Existing research in Information and Communication Technologies (ICT) and autism has emphasized the potential of technology as a key element in acquiring social behaviors [9]. Motivation for social interaction can be increased by practicing social behaviors in a Mixed Reality (MR) Environment which helps children create new behavioral patterns. In these MR environments, communication is not limited to verbal interaction and can incorporate embodied interaction which goes hand in hand with social perception and social understanding. In these environments, the body plays the central role including gestures and non-verbal communication which are necessary in interpersonal communication. The advantage of full-body, face-to-face interaction for Mixed Reality systems over other types of interaction and interfaces [19, 48, 63] (such as head-mounted displays), is that they can naturally include corporal, proxemic, proprioceptive, and kinesthetic indicators in the experiences of the participants and do it through non-encumbering,

non-invasive, and non-isolating experiences. Our belief is that to fully benefit from ICT environments, embodiment should be integrated into the design of these systems. In 2010 di Paolo [18] supported this potential of Embodied Cognition emphasizing the importance of embodiment for social understanding and social perception. Even earlier, in 2005, the project MEDIATE applied the design principles of full-body interaction and allowed the participating children to use gesture, touch and movement to engage in the interactive environment [47]. This work was one of the first studies to explore how placing ASC children in digital settings could help promote and enhance their performance in both communication and social realms. Full-Body interaction setups might provide a useful medium for play therapy and intervention, as recent evidence shows that an increase in body activity in multiplayer settings improves the social dynamics of the gaming experience [8, 41]. Moreover, large scale full-body interactive environments make physical exploration and face-to-face interactions possible between participants [45], where communication and interaction extend beyond verbal expressions to incorporate physical activity such as cooperative behaviors, joint attention and proxemics.

Evaluating the efficacy of interventions on social skills for ASC is a major challenge as it is difficult to tackle the complexity of naturalistic interpersonal behavior. Manual video coding methods still remain one of the most common approaches in evaluating behavior of children with ASC. Microanalytic video coding [29] has the advantage of applying current knowledge on social behaviors of children with ASC. However, since they are undertaken by human coders, they incorporate a certain degree of subjectivity. To compensate for this, multiple coders work together to achieve high inter-rater or inter-coder agreement. This is a time-consuming and resource intensive process which cannot be scaled to large amounts of video datasets. Therefore, machine learning techniques trained on high-quality coding can automate this process to make the process scalable. Moreover, they can uncover the rules of social interaction which are still largely unknown [3, 7, 33] and help better quantify typical and atypical non-verbal behaviors in social interactions. This can provide an objective computational approach that can tackle the complexity of interpersonal behavior. When it comes to automatically analyzing nonverbal behaviors in children, face and body inputs have been extensively studied [5]. For example, Sanghvi [56] used upper body silhouette features to train a set of classifiers for engagement prediction in children playing with a robot. Using a similar approach, researchers have been able to predict engagement based on the children's facial cues [2] and body movements [13]. Esteban [22] used gaze direction, facial expressions and body posture to classify stereotypical behaviors and "social" engagement of children with ASC. Rudovic et al. achieved personalized emotional perception and engagement using contextual and personal information of children with ASC during robot-assisted therapy [51]. The focus of these studies was on child-robot interaction using children's behavioral cues [43]. In the context of interaction between individuals, Piana [50] presented an emotion recognition system that used full-body movement features to help children with ASC. Specifically, they explored a collection of body features that can be extracted from video sequences to recognize emotions automatically. Their framework monitors ASC children while they interact with others or play a serious game and evaluates their ability to express and understand emotions. However, the dataset used to test the feature extraction algorithms was based on non-ASC individuals and was therefore not yet aligned with the final objective of helping ASC children.

Social activity that incorporates full-body activity is a fundamental constructor of non-verbal social interaction behaviors between individuals. The project by Coppola [15] introduced one of the most informative sets of features based on two interacting participants. Although this was not designed for ASC specifically, their analysis provides a robust automation of body cue recognition. Their results show that their approach has a potential in ASC research.

In the present study we investigate how bodily expression differs between ASC and non-ASC children during social initiation in MR and LEGO settings. To analyze the body cues, we utilize and adapt the methodology of Coppola [15] and extract a set of individual and interpersonal spatio-temporal features from both subjects, using the skeleton joints. We utilize statistical hypothesis testing to determine if there are differences in the values of body features for social initiation, generated by ASC and non-ASC participants, in both platforms. As a next step, we evaluate a set of classification models to study the joint impact of these body features. We also undertake detailed visual analysis of social initiations of children with ASC and non-ASC in order to better understand the context around social initiation behaviors. Our analyses and results indicate that an MR platform can be utilized to positively manipulate the nonverbal body behavior between ASC and non-ASC children during social initiation as much as a traditional therapy setting (LEGO).

2 MATERIALS AND METHODS

2.1 The MR system (Experimental Condition)

The system is a Mixed Reality full-body interaction environment designed specifically for children with ASC as a space to practice social initiation with a non-ASC peer [45]. It is an installation based on a virtual environment (VE) that encourages an ecologically valid exploration and discovery of hidden virtual items and surprises. This installation targets an experience shared by an ASC child and a non-ASC child during which they are scaffolded to socialize and collaborate.

2.1.1 Setup. The visual interface is configured as a circular projection on the floor, six meters in diameter, generated by images from two Full HD projectors. The children interact with the projected VE through a physical object acting as a "pointer" and "placeholder" (Figure 1). This object allows them to better focus on the activity and have a better sense of control over that which they are affecting in the VE. A multi-camera system tracks these objects on the playing field plane.

2.1.2 Content. The interactive experience provides an enticing range of visual and sound effects to spark the social initiation behaviors in the child with ASC (Figure 2). The VE, animations, and characters were designed with the help of children with ASC through Participatory Design Workshops. The actions to be undertaken by the children are adapted and balanced in difficulty for each child to generate a fair and adequate equilibrium (Figure 2).

2.1.3 Interaction Design. The MR system designed provides a context for the two players to adopt exploration and discovery attitudes by navigating physical and virtual space simultaneously in a face-to-face configuration. Discovery of interesting situations is meant to generate the sufficient internal drive in the ASC child to want to communicate with their peer. This ecologically valid configuration is akin to situations they can find in real life, such as meeting children in a public park or in the school playground. The full-body interaction permits them use not only verbal communication but also facial expressions and body language. The configuration also allows for joint attention and collaborative activity.

2.2 Non-ICT play activity: Lego Game (Control Condition)

We defined a LEGO activity by following Daniel Legoff's effective and well-known play-based therapy for children with ASC [39]. Their findings showed improved acquisition of social skills, particularly when peer interaction was initiated. We have adopted this strategy and created a LEGO play environment with the advise from the team of experts that we collaborated in the Hospital Sant Joan de Déu. In our case, the children were asked to find a set of "lost" LEGO figurines, then build a ship for the figures. This installation targets an experience shared by an ASC child and a non-ASC child (Figure 2).
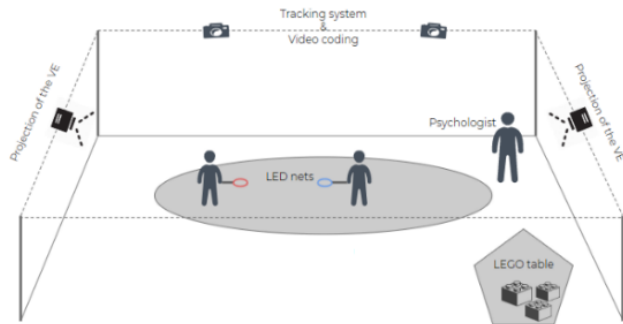

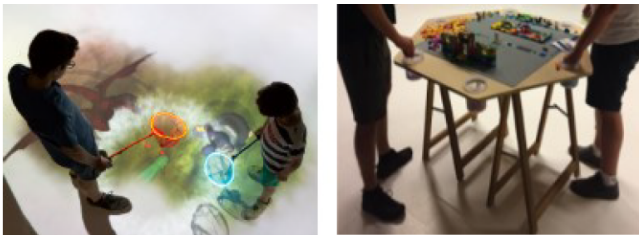**Figure 1: Diagram of the MR system during an experimental session**

Figure 2: Two players can collaboratively manipulate props in the scenario by bringing their nets together (left) / A control condition included LEGO social skills training elements where children were asked to collaborate in a shared building task (right)

2.3 Experimental Design and Procedure

All procedures performed were in accordance with the ethical standards of the ethical committee of the hospital we collaborated with and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Moreover, as a result of our collaboration with the Hospital Sant Joan de Déu, we contacted high-functioning ASC children on a local basis in the city of Barcelona. Eighteen ASC/non-ASC dyads (n = 3 female, n = 15 male), aged 8-12, formally diagnosed with ASC as determined by the Autism Diagnostic Observation Schedule (ADOS) module 3 reaching a value of 4 [40], participated in the study. The psychologist started the experimental session similar to a social skills therapy session weighing the emotional sensitivity of children with ASC. A visual support activity called "Jumby is Calm" was used to predict possible emotional reactions which can arise during the session. Children were then provided with the components used during the trial. Participants were fitted with a wearable designed by our lab for recording psycho-physiological measurements and showed an introductory video of how to use the handheld pointers to interact in the MR System. The children with ASC spent 15 minutes in the MR environment playing with their non-ASC peer, and 15 minutes in the LEGO setup with the same partner. Children answered interview questions administrated by tablets before and after each condition. All children participated in both experimental conditions, following a randomized order for each pair to compensate for order effects, with a 5 min break and a relaxation exercise between conditions.

3 DATA ANALYSIS

3.1 Data

3.1.1 Definition. Aligned with the previous research [3, 39, 44, 46, 55], the social initiations were coded using an observational grid with the categories derived from Bauminger's Social Emotional Intervention study [3]. With the help of our collaborator Hospital and the project's psychologist we adapted Bauminger's scheme to fit the real-time specificities of the data set we were analyzing. This scheme allowed us to determine the social behavior "initiation". Initiation is defined as follows in this scheme: The child begins a new social sequence directed towards other participant, distinguished from a continuation of a previous sequence by a change in activity.

3.1.2 Data Annotation. In our study, the behaviors of social initiation have primarily been identified by video recording and coding for each condition of the trials. In order to code the overt behaviors of children with ASC, Boris [26] video coding software was used. All sessions were captured by two cameras installed at opposite sides of the playground. The intercoder reliability has been calculated for the given class, both through percentage agreement and Kappa (Cohen's Kappa). Kappa scores were between 0.60 to 0.69 and the percentage level of agreements were between 0.71 to 0.78 with three coders. Moreover, to include the complex dynamics of body gestures in our analysis, we defined clips of the social initiation moments as temporal windows of 10 seconds (5 seconds before and 5 seconds after the coded social initiation moment).

3.2 Preprocessing and Feature Extraction

3.2.1 Automatic body and head pose tracking. To obtain the full body skeleton joints, we used the fully automatic multi-person 2D body tracker OpenPose [11]; which provides the joint locations of the human body (e.g., hand, shoulder, etc.) in RGB sequences. Similar to the approach followed in [15], we focused on 15 joints for extracting features related to body pose. To track head pose we adopted the Openheadpose framework [10] which leverages on OpenPose. In order to characterize the labeled social interaction clips, we extracted body features using tracked skeleton information. These features were divided into two main categories: (i) individual features with n=51 for each individual totaling to 102 individual features; and (ii) interpersonal features with "proximity" (prox) n=245, and "visual focus of attention"(vfoa) n=6, in accordance with the labeled social initiation.

3.2.2 Spatio-Temporal Features from Individual Data. In order to obtain features from both subjects, we implement the feature extraction methodology of [24] and [25] which have been used successfully in human day-to-day activity detection. These features also encapsulate majority of the features that Kleinsmith [36] first introduced in 2005 where they have attempted to understand whether emotion categories and affective dimensions can be predicted from affective postures. Following [24] and [25], we obtained 51 spatio-temporal features, such as: euclidean distances between joints; angles formed between joints; and torso inclination. Each individuals individual features grouped under the category individual features with 102 features in total.

3.2.3 Spatio-Temporal Features from Interpersonal Data. Gaze is considered the primary focal point of social interactions [42]. Visual Focus of Attention (vfoa) is a version of gaze that reveals "who is looking and where" [57]. In order to detect the vfoa of each participant we used the head pose estimation features expressed as angles of pitch, roll and yaw for each individual (6 features in total). We also used Hall's [30] proxemics theory, to estimate proximity features. Since the MR system facilitates the collaborative activities on joint actions, children's proximity might be a factor leading to initiations. In this regard, similar to the strategy followed in [15], we define social features as the ones that describe the relationship between two skeletons based on physical proximity, i.e. inter-body distance during the interaction. This set of features encompasses different subsets of features as follows: (a) log-covariance of the joints distances between bodies (120 features); (b) the minimum distance between any joint of one person and the torso of the other (2 features); (c) torso to torso distance (1 feature); (d) the energy over the euclidean distances from all joints of skeleton one to the torso of skeleton two and vice-versa (2 features); and (e) features that were computed similarly to the first subset, however, in a temporal way (a temporal window, herein defined as 10 frames). From all subsets, we extracted 251 interpersonal features per frame given both skeletons are interacting, which is similar to other feature sets. We formed our feature vectors per frame given that each 10 second clip for social initiation moments consists of 300 frames (recording 30 fps). As an initial step, we resampled every social initiation clip and it was done with respect to the mean, computed over 300 frames for each social initiation clip.

3.3 Experiments and Results

3.3.1 Statistical Hypothesis Testing. In this section, we present the details of the statistical hypothesis testing to determine if there are differences in the values of each body cue features for social initiation generated by ASC and non-ASC participants in both platforms (MR system and LEGO). Previous work found that children with ASC exhibit a clear deficit in movement observable from birth [60] and evident throughout life [14, 17, 20, 23, 52, 53, 61, 62]. Nonetheless, it has also been shown that children with ASC can identify emotions in body language just as accurately as their non-ASC peers [49].

Social communication includes both verbal and nonverbal communication, therefore when fostering social initiation skills in children with ASC with ICT we believe we must strive to incorporate nonverbal communication as well as verbal. Nonverbal behavior expression in children with ASC might have a role in the complexity of the social context. In MR environments, communication is not limited to verbal interaction and incorporates embodied interaction which goes hand in hand with social perception and social understanding. However, no research has yet assessed the impact of an MR full-body interactive system based on individual and interpersonal body cues in fostering social initiation in children with ASC, and compare it with a traditional intervention approach. In this regard, our hypotheses are formulated as follows:

H1: *There are statistically significant differences between ASC and non-ASC nonverbal (body) behavior at individual and interpersonal level in LEGO during social initiation.*

H2: *There are statistically significant differences between ASC and non-ASC nonverbal (body) behavior at individual and interpersonal level in the MR platform during social initiation.*

To validate these hypotheses, the Mann-Whitney U test was run as the difference in the values of each body cue features (averaged over all the initiations taking place during each play session for each participant) for the ASC and non-ASC. We present and compare the results of the Mann-Whitney U tests for LEGO and the MR system in terms of individual and interpersonal features. All the significance values reported in following result sections were at the significance level p < .05. As a next step, we develop classification models as a way to study the joint impact of aforementioned nonverbal body cues.

Results for H1. Based on the Mann-Whitney U tests with the target variable ASC/non-ASC, we found that several individual and interpersonal feature types were significantly different between ASC and non-ASC during social initiation in LEGO (see Table 1). Among the six angles obtained from triangles formed by: shoulder, hand and elbow; hip, shoulder and knee; hip, knee and foot, all considering left and right sides for individuals; an angle obtained from shoulder, hand and elbow was significantly different between participants during social initiation. Moreover, the skeleton joint distance between foot to nose was significantly different between participants during social initiation.The spatio-temporal feature type, the velocities and energy of the upper joints of the skeleton were also significantly different between social initiations of the participants. In terms of spatiotemporal proximity features 1 out

Table 1: Types of features which differ between ASC & non-ASC during initiation in LEGO.

Feature Type	Feature Category
foot to nose	individual
angle right shoulder elbow hand	individual
upper joints velocity	individual
log-cov. of the joints distances (temporal)	interpersonal-prox

Table 2: Types of features which differ between ASC & non-ASC during initiation in MR

Feature Type	Feature Category
hand to hand	individual
angle right shoulder elbow hand	individual
nose to hip	individual
angle hip knee foot	individual
foot to nose	individual
vfoa roll	interpersonal-vfoa
vfoa yaw	interpersonal-vfoa
energy joints ind. to ind. torso	interpersonal-prox
ind. to ind. torso	interpersonal-prox
torso to torso	interpersonal-prox
log-cov. of the joints distances	interpersonal-prox
log-cov. of the joints distances (temporal)	interpersonal-prox

of 120 features was significantly different between ASC and non-ASC social initiation during LEGO. We have not observed any visual focus of attention related features being significantly different between participants.

Results for H2. Based on the Mann-Whitney U tests with the target variable ASC/non-ASC, we found that several individual and interpersonal feature types were significantly different between ASC and non-ASC during social initiation in the MR system (see Table 2). Similarly to the LEGO condition we found that an angle obtained from shoulder, hand and elbow was significantly different between participants during social initiation. Additionally, the angle obtained from hip, knee and foot was significantly different between participants during social initiation. Again similar to the LEGO condition, foot to nose was significantly different between participants during social initiation. However, this type of individual difference also included nose to hip and distance between hand features in the MR system. In terms of spatiotemporal proximity features, 69 out of 120 features show significant differences between participants. We also observed differences between torso distance-based features. The minimum euclidean distance among all joints from individual one to the torso of individual two, and vice-versa, were significantly different between participants during social initiation. The obtained energy over the euclidean distances from all joints of skeleton one to the torso of skeleton two, and vice-versa were, also significantly different between participants during social initiation. Moreover, we have seen significant differences in head pose features expressed as yaw and roll between participants during social initiation. We found that among 102 individual features, four features for LEGO and five features for the MR system showed significant differences between participants during social initiation. On the other hand, we observed that among 251 (vfoa + prox) interpersonal features, 1 feature for LEGO and 81 features for the MR system showed significant differences between participants during social initiation. These results show that most of the features showing the nonverbal behavior differences between ASC and non-ASC were individual features in LEGO, while in the MR system these were interpersonal features. As a next step, we develop classification models as a way to study the joint impact of aforementioned nonverbal body features.

3.3.2 Automatic Classification. Automatic classification allows us to study the differences between ASC and non-ASC nonverbal behavior in detail for both platforms (MR system and LEGO). The problem is formulated as a 2-class classification task for both platforms. The initiations from ASC and non-ASC in the MR system

were named as ASCMR(n=117) and non-ASCMR(n=113). The initiations from ASC and non-ASC in LEGO were named as ASCLEGO (n= 121) and non-ASCLEGO (n=140).

We used the same features that were derived from the skeletons based on individual and interpersonal body movements (proxy & vfoa based social features). As a first step, all features were considered as a single feature set, as a combination of all the individual and interpersonal features, for each of the 2 scenarios (LEGO and MR).

We used Weka Data Mining Software for automatic classification [31]. We used the accuracy, precision, recall and F1-score metric to evaluate the models. We compared the findings for classifiers widely used in automated human motion analysis, namely Random Forest (RF) and Support Vector Machine (SVM) [1, 6, 28]. The hyperparameters of each model are tuned based on two grids of parameters. The grid of parameters used for SVM model refers to; 1) kernel (radial basis function -rbf, the polynomial kernel), 2) penalty parameter C of the error term (5 values ranging from 0.1 to 1000), and 3) gamma Kernel coefficient for rbf (4 values ranging from 0.001 to 1). The best set of parameters leading to the best performance of the SVM model was obtained across a 10-fold cross validation. Moreover, the number of trees for our RF model was set to 500, performing 10-Fold CV with each combination of values. Table 3 and Table 4 depict the overall scores obtained for the feature set which consists of all the individual and interpersonal features, for each participant in a subject-independent manner for LEGO and the MR system, respectively.

With automatic classification we were able to show that social initiation (ASC or non-ASC) recognition scores based on body features, as a combination of individual and interpersonal feature sets, for both LEGO and MR conditions are above chance level. Model scores were significantly different from the baseline performance (ZeroR). Moreover, we observed that classification scores were not statistically different between LEGO and the MR system.

3.3.3 Video Analysis of the nonverbal behaviors between ASC and non-ASC. Following the analysis from the previous 2 sections, we also undertook a detailed visual analysis of social initiations of children with ASC and non-ASC in order to better understand the

Table 3: Classification Performance of Different Algorithms on LEGO data

Method	Class	Precision	Recall	F-Measure	CV Acc
ZeroR	ASC	-	-	-	53.64%
ZeroR	non-ASC	53.60%	100.00%	69.80%	
RF	ASC	64.30%	52.10%	57.50%	64.37%
RF	non-ASC	64.40%	75.00%	69.30%	
SMO	ASC	**64.90%**	**59.50%**	**62.10%**	**66.28%**
SMO	non-ASC	**67.30%**	**72.10%**	**69.70%**	

Table 4: Classification Performance of Different Algorithms on MR data

Method	Class	Precision	Recall	F-Measure	CV Acc
ZeroR	ASC	50.90%	100.00%	67.40%	50.87%
ZeroR	non-ASC	-	-	-	
RF	ASC	60.20%	60.70%	60.40%	59.57%
RF	non-ASC	58.90%	58.40%	58.70%	
SMO	ASC	**60.90%**	**59.80%**	**60.30%**	**60.00%**
SMO	non-ASC	**59.10%**	**60.20%**	**59.60%**	

context around social initiation behaviors. The head and hands are seen responsible for the most movement and have received the highest attention in developing adequate coding strategies [32]. Therefore, we decided to focus on observing the movement of head and hands in LEGO which reflects the features that are significantly different between ASC and non-ASC in LEGO. MR environment involves a large-scale setting which lets players engage in collaborative activities by moving closer to their partner. Moreover, in body movement coding research the primary proxemic variable has been the distance between interactants, and it has garnered the most investigative attention [32]. Therefore, we decided to focus on observing the distance between participants in each social initiation moment from ASC and non-ASC in the MR environment which might cover most of the proximity based features discovered as significantly different in the previous sections. From 18 trials we randomly selected 4 trials (appx corresponding to %25 of all the trials) and from these 4 trials, a trained video coder, watched each social initiation video clip from ASC and non-ASC for LEGO and the MR system.

LEGO. As the upper joints velocity was significantly higher in non-ASC initiation moments compared to ASC initiation moments (Figure 3), we focused on hand and head movements. In the first 5 min of the sessions, searching the LEGO pieces from boxes was common for both participants which required a lot of hand movements and reaching out gestures to all boxes (Figure 4). However, there are less pieces on the LEGO table in the non-ASC initiation moments compared to the ASC initiation moments. This might signal that during non-ASC initiation moments, the participants are still in the searching process (Figure 4) and as a result they might have had more body movements. Moreover, there were more initiations (although it is not significantly different) in the first 5

mins of the session by the non-ASC kids compared to ASC kids (Figure 3).
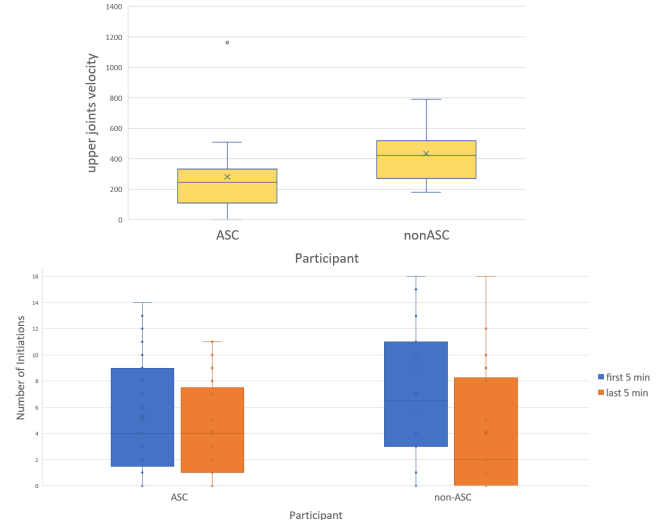

Figure 3: Top: upper joints velocity differences between ASC and non-ASC kids in social initiation moments during LEGO (left). Bottom: number of initiations generated by ASC and non-ASC in the first and last 5 mins of the LEGO sessions (right).

MR. As the torso to torso distance was significantly higher in non-ASC initiation moments compared to ASC initiation moments (see Figure 5), we focused on observing the torso to torso distances between ASC and non-ASC in those clips. We observed that during the initiation moments, when exploration is taking place in the playing plane, ASC kids had a tendency to follow the non-ASC kids more often and they were keeping a close proximity (see Figure 6). As it can be seen in the Figure 6 top image, the ASC (red net) kid initiates while he is following his non-ASC peer (blue net) during exploration. On the bottom image it can be observed that non-ASC discovers the virtual door under the virtual fog and shares this experience by creating a social initiation (pointing out to the door, calling the ASC kid) with the ASC kid (red net) who is at that stage outside of the playing field.

4 DISCUSSION

In our statistical hypothesis testing, we saw that the angle obtained from shoulder, hand and elbow, and the skeleton joint distance between foot to nose, were significantly different between ASC and non-ASC during social initiation in LEGO. In MR environment, the angle obtained from hip, knee and foot; and nose to hip and distance between hands, were significantly different between participants during social initiation. The aforementioned features are useful when discriminating between postures related to standing and sitting or torso inclination. We observed that more of these features play a role in differentiating between ASC and non-ASC during initiation in MR setting. This might be related to the diversity of
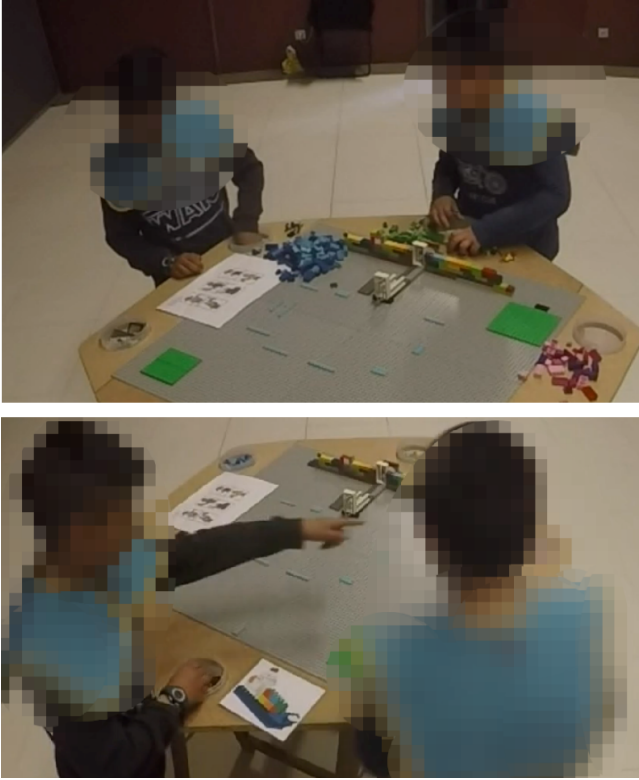
Figure 4: Top: non-ASC kid (on the left) and ASC kid (on the right) playing LEGO with social initiation by ASC kid. Bottom: non-ASC kid (on the left) and ASC kid (on the right) playing LEGO with social initiation by non-ASC kid.
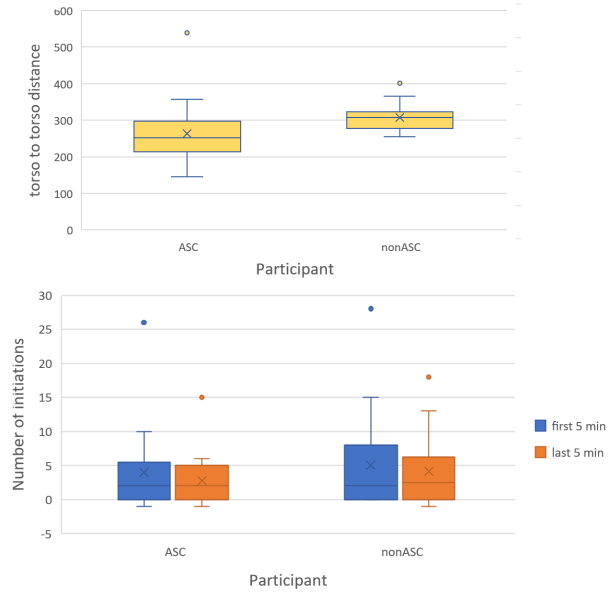
Figure 5: Top: torso to torso differences between ASC and non-ASC kids in social initiation moments during the MR system (left). Bottom: number of initiations generated by ASC and non-ASC in the first and last 5 mins of the MR system sessions (right).

full-body activities that the participants are motivated to undertake in MR setting.

An representative example of initiation in LEGO setting would involve changing the standing position which is likely to occur when picking up the LEGO piece dropped onto the floor. We also observed that the velocity and the energy of the upper joints were statistically different between ASC and non-ASC during social initiation in LEGO. This feature plays a role in the formation of key movement poses, e.g., in repeated movements, joints alternately accelerate and decelerate which leads to changes in the energy model representation. This information also helps the characterization of drastic changes in skeleton direction and velocity. Observing this type of feature in LEGO, but not in MR, might be related to the amount of repeated movements generated by the children with ASC in LEGO as compared to MR. Although generation of repetitive behavior varies from person to person, the reasons motivating such behavior might be the same, e.g., to deal with stress and anxiety and to block out uncertainty [34].

We also observed differences in torso distance-based features in MR. These features are important in identifying the most active person (i.e., the person entering into the other person's individual space). We also observed that majority of the features representing nonverbal behavior differences between ASC and non-ASC in

LEGO setting were the individual features, while in the MR setting these were the interpersonal features. The large scale setting of the MR system might have led to higher numbers of locomotive actions, as players would display collaborative actions by moving within proximity to their partner, sometimes coming from across the play area. In such situations, players can accomplish joint actions without the need for verbal communication. As the LEGO setting provided a context where players were already within proximity to each other, sharing a focal point of joint attention, features related to differences in proximity and vfoa were not as commonly observed.

Using computational models, the combination of the aforementioned features allowed us to recognize the type of social initiation moment (ASC or non-ASC) from body cues in both environments, validating our hypotheses. Past research indicated that different body cues work together [16] and may contribute more information when treated together. We did not observe significant differences between the evaluated models in terms of performance for LEGO vs. MR environment. This might be interpreted as the MR system encouraging similar nonverbal behaviors in children, as similarly as the LEGO environment, perhaps even more similar as the performance scores in MR setting are lower as compared to LEGO setting. This also implies that if the children's nonverbal behaviors are encouraged to be similar, they also get to be in a similar position to initiate social interaction.

When we consider the manual video analysis for LEGO settings, the most common nonverbal behavior for both participants during
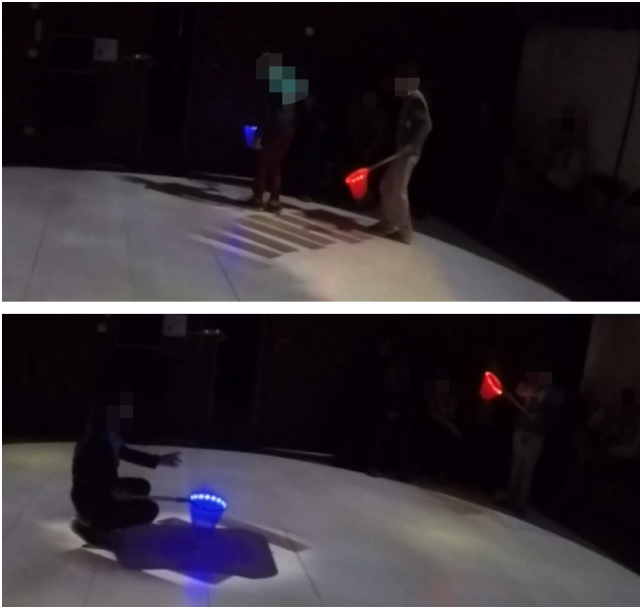
Figure 6: Top: non-ASC kid (with blue net) and ASC kid (with red net) playing in the MR environment and it is the moment of social initiation from ASC kid, Bottom: non-ASC kid (on the left) and ASC kid (on the right) playing in MR and it is the moment of social initiation from non-ASC kid.

initiation was related to searching LEGO pieces – i.e., hand movements and reaching out gestures. However, the number of LEGO pieces on the table for these initiation moments differed between ASC and non-ASC. This might signal that during non-ASC initiation moments, the participants are still in the process of searching (Figure 4) and therefore they might have had more body movements enabling easier detection of nonverbal behavior differences. Moreover, searching activity for LEGO pieces might be related to imitation as both participants are doing similar activity. Past work on ASC demonstrated that motor imitation related impairments were common during social communication [35, 54, 59]. On the other hand, we observed that during the initiation moments, when exploration is taking place in the play area, ASC kids had a tendency to follow the non-ASC kids and they would keep a close proximity. Data from several studies suggests that the ASC child usually has less chance to take a leading role in activities with mixed dyads of friendship. Therefore, settings that mediate communication between individuals should encourage balanced interactions to ensure that children with ASC have an equal opportunity in leading the interactions and the game [4]. As we were able to observe ASC initiation taking place at these close proximity moments, it can be said that the MR system might be providing balanced opportunity for ASC children to take the lead. We also note that it was more likely to observe the non-ASC kids exploring the play area individually during non-ASC initiation. Evidence from previous comparison studies suggest that non-ASC children interact less with ASC children [21, 37, 38]. This might explain why non-ASC kids are away from their ASC peers in non-ASC social initiation

moments. In MR setting, observing non-ASC children initiating is therefore promising in terms of fostering interaction between mixed dyads.

5 CONCLUSION AND FUTURE WORK

This study is a first step in analyzing the body behaviour differences between ASC and non-ASC children during social initiation in a Mixed Reality (MR) environment and a typical LEGO-based intervention on social skills, where both environments mediate dyadic face-to-face play. To analyze the body cue differences between ASC and non-ASC, we extracted a set of individual and interpersonal spatio-temporal features derived from the skeletons of both subjects. This allowed us to observe the specific individual and interpersonal features which differ between ASC and non-ASC children during social initiation in MR and LEGO settings. In general, most of the features showing the nonverbal behavior differences between ASC and non-ASC were individual features in LEGO setting and interpersonal features in MR setting. Observing this difference for the MR system is a promising result. MR systems exploit full-body interaction, and incorporate navigation of physical and virtual space, take advantage of body gestures, allow relationships in space with the other user, and, in general, include non-verbal communication compared to other types of intervention.

We also investigated the joint effect of the various sets of features. As a result, we were able to automatically recognize the type of social initiation moments (ASC or non-ASC) from body cues in LEGO and in MR. Computational models evaluated for social initiation classification did not show significant differences between LEGO and MR settings. This is a promising outcome indicating that an MR platform can be utilized to positively manipulate the nonverbal body behaviors between ASC and non-ASC children during social initiation as well as a traditional therapy setting (LEGO). However, further work is needed to discover hidden relationships between different body cues and their differences between ASC and non-ASC children.

Our manual video analysis confirmed that the body features utilised for automatic analysis were perceivable and meaningful to a human coder. However, these body features should be analyzed together with other data sources, such as system logs of game events, to have a better understanding of the differences between ASC and non-ASC social initiations. This will lead to a more insightful understanding of how a face-to-face, full-body interactive MR system can become a mediator to foster socialization between the ASC and the non-ASC children, and pave the way for designing novel tools that can be utilised by the ASC therapists and caregivers.

REFERENCES

[1] Insaf Ajili, Malik Mallem, and Jean-Yves Didier. 2018. Relevant LMA Features for Human Motion Recognition.
[2] Salvatore M Anzalone, Sofiane Boucenna, Serena Ivaldi, and Mohamed Chetouani. 2015. Evaluating the engagement with social robots. *International Journal of*

Social Robotics 7, 4 (2015), 465–478.

[3] Nirit Bauminger. 2002. The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of autism and developmental disorders* 32, 4 (2002), 283–298.

[4] Nirit Bauminger, Marjorie Solomon, Anat Aviezer, Kelly Heung, John Brown, and Sally J Rogers. 2008. Friendship in high-functioning children with autism spectrum disorder: Mixed and non-mixed dyads. *Journal of Autism and Developmental Disorders* 38, 7 (aug 2008), 1211–1229. https://doi.org/10.1007/s10803-007-0501-2

[5] Tony Belpaeme, Paul Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuitl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová, Georgios Athanasopoulos, and Valentin Enescu. 2013. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction* 1, 2 (2013), 33–53.

[6] Daniel Bernhardt and Peter Robinson. 2007. Detecting affect from non-stylised body motions. In *International conference on affective computing and intelligent interaction*. Springer, 59–70.

[7] Nadia Bianchi-Berthouze. 2013. Understanding the role of body movement in player engagement. *Human–Computer Interaction* 28, 1 (2013), 40–75.

[8] Nadia Bianchi-Berthouze, Whan Woong Kim, and Darshak Patel. 2007. Does body movement engage you more in digital game play? and why?. In *International conference on affective computing and intelligent interaction*. Springer, 102–113.

[9] Jennifer Brown and Donna Murray. 2001. Strategies for enhancing play skills for children with autism spectrum disorder. *Education and Training in Mental Retardation and Developmental Disabilities* (2001), 312–317.

[10] Yuanzhouhan Cao and Olivier Can. [n.d.]. Leveraging Convolutional Pose Machines for Fast and Accurate Head Pose Estimation. ([n. d.]).

[11] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008* (2018).

[12] Meia Chita-Tegmark. 2016. Attention allocation in ASD: A review and meta-analysis of eye-tracking studies. *Review Journal of Autism and Developmental Disorders* 3, 3 (2016), 209–223.

[13] Mark B Colton, Daniel J Ricks, Michael A Goodrich, Behzad Dariush, Kikuo Fujimura, and Martin Fujiki. 2009. Toward therapist-in-the-loop assistive robotics for children with autism and specific language impairment. *autism* 24 (2009), 25.

[14] Jennifer L Cook, Sarah-Jayne Blakemore, and Clare Press. 2013. Atypical basic movement kinematics in autism spectrum conditions. *Brain* 136, 9 (2013), 2816–2824.

[15] Claudio Coppola, Diego R Faria, Urbano Nunes, and Nicola Bellotto. 2016. Social activity recognition based on probabilistic merging of skeleton features with proximity priors from rgb-d data. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5055–5061.

[16] Marco Costa, Wies Dinsbach, Antony S R Manstead, and Pio Enrico Ricci Bitti. 2001. Social presence, embarrassment, and nonverbal behavior. *Journal of Nonverbal Behavior* 25, 4 (2001), 225–240.

[17] Hanne De Jaegher, Hanne De Jaegher, Martha R Leary, and Private Practice. 2013. Embodiment and sense-making in autism. *Frontiers in integrative neuroscience* 7, March (2013), 15. https://doi.org/10.3389/fnint.2013.00015

[18] Ezequiel Di Paolo, Marieke Rohde, and Hanneke De Jaegher. 2010. Horizons for the enactive mind: Values, social interaction, and play. In *Enaction: Towards a new paradigm for cognitive science*.

[19] Nyaz Didehbani, Tandra Allen, Michelle Kandalaft, Daniel Krawczyk, and Sandra Chapman. 2016. Virtual reality social cognition training for children with high functioning autism. *Computers in Human Behavior* 62 (2016), 703–711.

[20] Rebecca Downey and Mary Jane K Rapport. 2012. Motor Activity in Children With Autism : A Review of. (2012). https://doi.org/10.1097/PEP.0b013e31823db95f

[21] Nabil Hassan El-Ghoroury and Raymond G Romanczyk. 1999. Play interactions of family members towards children with autism. *Journal of autism and developmental disorders* 29, 3 (1999), 249–258.

[22] Pablo G Esteban, Paul Baxter, Tony Belpaeme, Erik Billing, Haibin Cai, Hoang-Long Cao, Mark Coeckelbergh, Cristina Costescu, Daniel David, and Albert De Beir. 2017. How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn, Journal of Behavioral Robotics* 8, 1 (2017), 18–38.

[23] Maddalena Fabbri-Destro, Luigi Cattaneo, Sonia Boria, and Giacomo Rizzolatti. 2009. Planning actions in autism. *Experimental brain research* 192, 3 (2009), 521–525.

[24] Diego R. Faria, Cristiano Premebida, and Urbano Nunes. 2014. A probabilistic approach for human everyday activities recognition using body motion from RGB-D images. *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions* (2014), 732–737. https://doi.org/10.1109/ROMAN.2014.6926340

[25] Diego R Faria, Mario Vieira, Cristiano Premebida, and Urbano Nunes. 2015. Probabilistic human daily activity recognition towards robot-assisted living. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 582–587.

[26] Olivier Friard and Marco Gamba. 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in*

Ecology and Evolution 7, 11 (2016), 1325–1330.

[27] Erving Goffman. 1955. On face-work: An analysis of ritual elements in social interaction. *Psychiatry* 18, 3 (1955), 213–231.

[28] Hatice Gunes and Massimo Piccardi. 2008. Automatic temporal segment detection and affect recognition from face and body display. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39, 1 (2008), 64–84.

[29] Kim Kopenhaver Haidet, Judith Tate, Dana Divirgilio-Thomas, Ann Kolanowski, and Mary Beth Happ. 2009. Methods to improve reliability of video-recorded behavioral data. *Research in nursing & health* 32, 4 (2009), 465–474.

[30] Edward T Hall. 1963. A System for the notation of proxemic behavior 1. *American anthropologist* 65, 5 (1963), 1003–1026.

[31] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11, 1 (2009), 10–18.

[32] Jinni A Harrigan. 2013. Methodology: Coding and studying nonverbal behavior. (2013).

[33] Margaret Hauck, Deborah Fein, Lynn Waterhouse, and Carl Feinstein. 1995. Social initiations by autistic children to adults and other children. *Journal of autism and developmental disorders* 25, 6 (1995), 579–595.

[34] Patricia Howlin. 2004. *Autism and Asperger syndrome: Preparing for adulthood*. Routledge.

[35] Brooke Ingersoll. 2008. The Social Role of Imitation in Autism: Implications for the Treatment of Imitation Deficits. *Infants & Young Children* 21, 2 (2008), 107–119. https://doi.org/10.1097/01.IYC.0000314482.24087.14

[36] Andrea Kleinsmith, P. Ravindra De Silva, and Nadia Bianchi-Berthouze. 2005. Grounding affective dimensions into posture features. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3784 LNCS (2005), 263–270. https://doi.org/10.1007/11573548_34

[37] Fiona Knott, Charlie Lewis, and Tim Williams. 1995. Sibling interaction of children with learning disabilities: A comparison of autism and Down's syndrome. *Journal of child psychology and psychiatry* 36, 6 (1995), 965–976.

[38] Fiona Knott, Charlie Lewis, and Tim Williams. 2007. Sibling interaction of children with autism: Development over 12 months. *Journal of Autism and Developmental Disorders* 37, 10 (2007), 1987–1995.

[39] Daniel B LeGoff. 2004. Use of LEGO© as a therapeutic medium for improving social competence. *Journal of autism and developmental disorders* 34, 5 (2004), 557–571.

[40] Catherine Lord, Michael Rutter, Susan Goode, Jacquelyn Heemsbergen, Heather Jordan, Lynn Mawhood, and Eric Schopler. 1989. Austism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of autism and developmental disorders* 19, 2 (1989), 185–212.

[41] Laura Malinverni, Joan Mora-Guiard, Vanesa Padillo, Lilia Valero, Amaia Hervás, and Narcis Pares. 2017. An inclusive design approach for developing video games for children with Autism Spectrum Disorder. *Computers in Human Behavior* 71 (2017), 535–549.

[42] Malia F Mason, Elizabeth P Tatkow, and C Neil Macrae. 2005. The look of love: Gaze shifts and person perception. *Psychological science* 16, 3 (2005), 236–239.

[43] Maja J Matarić. 2017. Socially assistive robotics: Human augmentation versus automation. *Science Robotics* 2, 4 (2017), eaam5410.

[44] Camilla M McMahon, Laurie A Vismara, and Marjorie Solomon. 2013. Measuring changes in social behavior during a social skills intervention for higher-functioning children and adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 43, 8 (2013), 1843–1856.

[45] Joan Mora-Guiard, Ciera Crowell, Narcis Pares, and Pamela Heaton. 2016. Lands of Fog: helping children with Autism in social interaction through a full-body interactive experience. *ACM SIGCHI Conference on Interaction Design and Children 2016* (2016), manuscript submitted for publication. https://doi.org/10.1145/2930674.2930695

[46] Gina Owens, Yael Granader, Ayla Humphrey, and Simon Baron-Cohen. 2008. LEGO® therapy and the social use of language programme: An evaluation of two social skills interventions for children with high functioning autism and Asperger syndrome. *Journal of autism and developmental disorders* 38, 10 (2008), 1944.

[47] Narcis Pares, Paul Masri, Gerard Van Wolferen, and Chris Creed. 2005. Achieving dialogue with children with severe autism in an adaptive multisensory interaction: the" MEDIATE" project. *IEEE Transactions on Visualization and Computer Graphics* 11, 6 (2005), 734–743.

[48] Sarah Parsons and Sue Cobb. 2011. State-of-the-art of virtual reality technologies for children on the autism spectrum. *European Journal of Special Needs Education* 26, 3 (2011), 355–366.

[49] Candida C Peterson, Virginia Slaughter, and Celia Brownell. 2015. Children with autism spectrum disorder are skilled at reading emotion body language. *Journal of experimental child psychology* 139 (2015), 35–50.

[50] Stefano Piana, Alessandra Staglianò, Antonio Camurri, and Francesca Odone. 2013. A set of full-body movement features for emotion recognition to help children affected by autism spectrum condition. In *IDGEI International Workshop*.

[51] Rosalind W. Picard, Björn Schuller, Jaeryoung Lee, Ognjen Rudovic, and Miles Dai. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics* 3, 19 (2018), eaao6760. https://doi.org/10.1126/scirobotics.aao6760

[52] Beth Provost, Brian R Lopez, and Sandra Heimerl. 2007. A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. *Journal of autism and developmental disorders* 37, 2 (2007), 321–328.

[53] Nicole J Rinehart, Bruce J Tonge, Robert Iansek, Jenny McGinley, Avril V Brereton, Peter G Enticott, and John L Bradshaw. 2006. Gait function in newly diagnosed children with autism: cerebellar and basal ganglia related motor disorder. *Developmental medicine and child neurology* 48, 10 (2006), 819–824.

[54] Sally J Rogers, Susan L Hepburn, Tracy Stackhouse, and Elizabeth Wehner. 2003. Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of child psychology and psychiatry* 44, 5 (2003), 763–781.

[55] Lisa Ruble, Heather Willis, and Valerie McLaughlin Crabtree. 2008. Social skills group therapy for autism spectrum disorders. *Clinical Case Studies* 7, 4 (2008), 287–300.

[56] Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W McOwan, and Ana Paiva. 2011. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 305–312.

[57] Samira Sheikhi and Jean-Marc Odobez. 2012. Recognizing the visual focus of attention for human robot interaction. In *International Workshop on Human Behavior Understanding*. Springer, 99–112.

[58] Marian Sigman, Ellen Ruskin, Shoshana Arbelle, Rosalie Corona, Cheryl Dissanayake, Michael Espinosa, Norman Kim, Alma López, Cynthia Zierhut, and Carolyn B Mervis. 1999. Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the society for research in child development* (1999), i–139.

[59] Wendy L Stone, Opal Y Ousley, and Cynthia D Littleford. 1997. Motor imitation in young children with autism: What's the object? *Journal of abnormal child psychology* 25, 6 (1997), 475–485.

[60] Philip Teitelbaum, Osnat Teitelbaum, Jennifer Nye, Joshua Fryman, and Ralph G Maurer. 1998. Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences* 95, 23 (1998), 13982–13987.

[61] E B Torres, J Nguyen, C Suresh, P Yanovich, and A Kolevzon. 2013. Noise from the periphery in autism spectrum disorders of idiopathic origins and of known etiology. In *Paper Presentation at the Annual Meeting of the Society for Neuroscience SFN*.

[62] Caroline Whyatt and Cathy Craig. 2013. Sensory-motor problems in Autism. *Frontiers in integrative neuroscience* 7 (2013), 51.

[63] Sze Ngar Vanessa Yuan and Horace Ho Shing Ip. 2018. Using virtual reality to train emotional and social skills in children with autism spectrum disorder. *London journal of primary care* 10, 4 (2018), 110–112.