International Journal of Social Robotics (2024) 16:2261–2276
https://doi.org/10.1007/s12369-024-01185-3

Challenges in Observing the Emotions of Children with Autism Interacting with a Social Robot

Duygun Erol Barkana[1] · Katrin D. Bartl-Pokorny[2,3] · Hatice Kose[4] · Agnieszka Landowska[5] · Manuel Milling[2] · Ben Robins[6] · Björn W. Schuller[2] · Pinar Uluer[7] · Michal R. Wrobel[5] · Tatjana Zorcec[8]

Accepted: 14 October 2024 / Published online: 8 November 2024

Abstract

This paper concerns the methodology of multi-modal data acquisition in observing emotions experienced by children with autism while they interact with a social robot. As robot-enhanced therapy gains more and more attention and proved to be effective in autism, such observations might influence the future development and use of such technologies. The paper is based on an observational study of child-robot interaction, during which multiple modalities were captured and then analyzed to retrieve information on a child's emotional state. Over 30 children on the autism spectrum from Macedonia, Turkey, Poland, and the United Kingdom took part in our study and interacted with the social robot Kaspar. We captured facial expressions/body posture, voice/vocalizations, physiological signals, and eyegaze-related data. The main contribution of the paper is reporting challenges and lessons learned with regard to interaction, its environment, and observation channels typically used for emotion estimation. The main challenge is the limited availability of channels, especially eyegaze-related (29%) and voice-related (6%) data are not available throughout the entire session. The challenges are of a diverse nature—we distinguished task-based, child-based, and environment-based ones. Choosing the tasks (scenario) and adapting environment, such as room, equipment, accompanying person, is crucial but even with those works done, the child-related challenge is the most important one. Therapists have pointed out to a good potential of those technologies, however, the main challenge to keep a child engaged and focused, remains. The technology must follow a child's interest, movement, and mood. The main observations are the necessity to train personalized models of emotions as children with autism differ in level of skills and expressions, and emotion recognition technology adaptation in real time (e. g., switching modalities) to capture variability in emotional outcomes.

Keywords Autism · Robot-enhanced therapy · Social robots · Automatic emotion recognition

✉ Michal R. Wrobel
michal.wrobel@pg.edu.pl

Duygun Erol Barkana
duygunerol@yeditepe.edu.tr

Katrin D. Bartl-Pokorny
katrin.bartl-pokorny@medunigraz.at

Hatice Kose
hatice.kose@itu.edu.tr

Agnieszka Landowska
nailie@pg.edu.pl

Manuel Milling
manuel.milling@uni-a.de

Ben Robins
b.robins@herts.ac.uk

Björn W. Schuller
bjoern.schuller@uni-a.de

Pinar Uluer
puluer@gsu.edu.tr

Tatjana Zorcec
tzorcec@gmail.com

[1] Faculty of Engineering, Department of Electrical and Electronics Engineering, Yeditepe University, 26 Agustos Yerlesimi, Kayışdağ Caddesi, 34755 Istanbul, Turkey

[2] Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Eichleitnerstraåe 30, 86159 Augsburg, Germany

[3] Division of Phoniatrics, Medical University of Graz, Auenbruggerplatz 26, 8036 Graz, Austria

[4] Faculty of Computer and Informatics, Department of AI and Data Engineering, Istanbul Technical University, Ayazaga Campus, Maslak, Sariyer, 34469 İstanbul, Turkey

1 Introduction

Recent studies in socially assistive robotics (SAR) domain enable the use of robots in many application areas such as healthcare, education, entertainment, and edutainment as well as various applications for vulnerable populations such as people with hearing disabilities, elderly, or children with autism spectrum disorder (ASD) [20, 26, 30].

Children with ASD are known to have limited social and emotional skills in their everyday interactions with others [12]. The field of social robotics is showing promising results in using robots to assist children with ASD to develop their social and emotional skills, to help them overcome social barriers, and to get them more involved in their interactions [12, 18].

A systematic literature review on robot-based interventions targeting emotion-related skills for children with ASD revealed growing interest in robot-based interventions, especially with regard to emotional intelligence skills [9]. However, among the papers reviewed, few studies have investigated automatic emotion recognition of children with ASD during interaction with a social robot. The main motivation of the presented study was to investigate the feasibility of the available practices and work towards a novel approach to create an affective loop in child-robot interaction that would enhance the intervention regarding emotional intelligence building in children with autism.

The study was undertaken within the EMBOA project entitled "Affective loop in Socially Assistive Robotics as an intervention tool for children with autism" that was a research project that aimed to combine affective computing technologies with social robot intervention for children with ASD. It was an international research project combining three research areas: autism therapy, social robots, and automatic emotion recognition, to develop a practical evaluation of the application of emotion recognition technologies in robot-assisted intervention for children with autism (https://emboa.eu/).

Incorporating automatic emotion recognition techniques into social robot therapy for children with ASD opens up a wide range of new possibilities. In the future, the social robot might be able to automatically recognise the child's emotions and adapt its responses according to the child's current needs. For instance, let us consider an interaction scenario between a child and a social robot companion: the robot detects that the child is bored and adapts its intervention routine accordingly, asking the child to sing a song together or start a funny game. This would allow for a more natural, engaging, and joyful interaction from the child's perspective. It would also reduce the amount of manual control needed by the therapist or operator during a robot-assisted intervention session.

The research question of our study presented in this paper might be given as follows: "how to effectively observe emotional states of a child with autism interacting with a robot?" To answer this question we have first referred to existing literature on the subject. We found out, that most of the studies that referred to robotics in autism therapy had no automatic emotion recognition involved, even when they were addressing emotional and social skills [9]. Moreover, the other studies that concerned emotion recognition in autism frequently were based on typically developing participants, with the classifiers trained on datasets that did not include people with autism. We found that very few studies combined robots and emotion recognition together and that the interaction with a robot differs from interactions with other technological solutions (such as mobile apps). Therefore, the goal of the study was to explore interaction between a robot and a child with autism, with regard to automatic emotion recognition.

This paper presents insights, findings, and comments from the observational sessions of children with ASD interacting with social robots. Although we discuss the results of an analysis of the availability of channels used by emotion recognition systems, we do not systematically assess their capabilities and feasibility during these interactions. The remainder of the paper is organised as follows: Sect. 2 discusses related studies on emotion recognition in children with autism; Sect. 3 presents the methodology of the observational study conducted, including the experimental setting, participants and interaction scenarios; the challenges identified in conducting the observational study are then presented in Sect. 4 and Sect. 5; finally, Sect. 6 summarises the findings and Sect. 7 presents conclusions and future works.

2 Related Works

Children with ASD suffer from socio-communicative deficits and often have significant problems identifying and expressing emotions [6]. The literature suggests that children with ASD may benefit from therapy with social robots, however, robot-assisted interventions face a number of challenges [3, 12, 21].

5 Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, ul. Narutowicza 11/13, 80-233 Gdansk, Poland

6 School of Physics, Engineering and Computer Sciences, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, United Kingdom

7 Dept. of Computer Engineering, Galatasaray University, Faculty of Engineering and Technology, Çırağan Caddesi No:36, İstanbul 34349, Turkey

8 University Children's Hospital, Medical Faculty, University of Skopje, 50 Divizija no. 6, Skopje 1000, North Macedonia

Liu et al. conducted a study with a robot-based basketball game in which the robot recognized the individual level of satisfaction and engagement of children with ASD in relation to the game configuration and selected appropriate behaviors based on this. In observational studies of both low-functioning and high-functioning children, they reported that children's low expressivity was a major challenge [25]. Aziz et al. came to similar conclusions in a study of high-functioning children with the NAO humanoid robot [7]. In another study with the NAO robot, conducted on a sample of 36 children, it was shown that automatic engagement recognition needs to be personalized for each child, especially in the case of different cultural backgrounds [32]. Kouroupa and colleagues conducted a meta-analysis of 12 randomised controlled trials (RCTs) revealing that robotic interventions significantly enhanced social functioning of children with autism. However, no improvement in emotional or motor functioning was observed [21]. A randomised controlled trial conducted by Holeva et al. showed that there were minimal statistically significant differences in developmental improvement, as indicated by neuropsychological testing and parental report, between the robot-assisted intervention group and a control group that received only human intervention [17]. In a systematic review, Sani-Bozkurt and Bozkus-Genc found that while children with autism generally respond positively to robots as social partners, the effectiveness of robots in developing joint attention skills in children with autism is still unclear [34].

Several studies have outlined the perceptual problems of children with ASD. In their study, Pop et al. investigated the use of a Probo robot to improve the ability of children with ASD to identify emotions. In their conclusions, they emphasized that they cannot determine whether the participants understood the emotions shown by the robot or whether they merely reproduced them [31]. Similar conclusions were reached by English et al., based on a study conducted with NAO and Mini Darwin robots [15].

Another set of challenges arises from the highly heterogeneous nature of autism disorder; what works for one child may not work for another [8]. Successful therapy therefore requires personalised and tailored interventions for each individual. Shi et al. developed and validated personalised models for robot perception of arousal and valence in children with ASD [36]. Alnajjar and colleagues proposed an adaptive robotic intervention system for ASD assessment and therapy. The results of the empirical study of six ASD patients in an autism rehabilitation centre showed that the adaptive approach significantly improved the attention levels of most patients in long-term therapy [5].

Silva et al. mentioned the varying developmental levels of children with ASD as a major challenge in observational studies of robot interactions. The study used robots built on a Lego Mindstorms NXT platform, where some activities were below children's abilities, leading to low motivation to complete tasks [37]. In contrast, in a study by Yun et al. using the iRobiQ and CARO robots, some children were unable to complete tasks without the help of therapists, due to the excessive complexity [44]. Furthermore, Landowska and Robins noticed that autistic children may show refusal or other atypical behaviors that do not match the planned interaction, environment, or equipment [22].

A more in-depth analysis of the state of the art in robotic interventions for children with ASD can be found in a comprehensive systematic review conducted as part of the EMBOA project [9].

The list of devices being used to capture person-centered data for emotion and other affect-related recognition experiments is extensive. Arguably, the most common practices for emotion recognition are based on (a combination of) audio-visual and physiological data [45] with multimodal approaches generally outperforming unimodal approaches [1]. This holds as well for the special case of children on the autism spectrum [33]. A core advantage of these modalities is that the necessary hardware, including eye trackers [46], show low levels of invasiveness, for instance compared to electromyography (EMG) or electroencephalogram (EEG) devices, as used in [13]. Nevertheless, the capturing devices for physiological signals, as well as lapel microphones need to be attached to participants, which can lower the acceptance rates, as further discussed in our study.

The issue of recording emotional symptoms when children with ASD interact with social robots has been addressed in a number of papers. For example, in the case of emotion recognition based on facial expressions, these include the children's movement during the observation, which makes it impossible to find a face, or the subject's inappropriate distance from the camera. These issues are described more extensively in a systematic literature review on automatic emotion recognition in children with ASD [23].

3 Methodology

This paper is based on an observational study that was conducted to assess the feasibility of using available automatic emotion recognition technologies, in terms of the availability of input channels, during the interaction of children with ASD with a social robot. The observational sessions were carried out in therapeutic centers in Poland, North Macedonia, Turkey, and the UK. All sessions in all centers were conducted using the same procedures and, where possible, the same or similar equipment. The preparation of the study involved addressing the ethical aspects of the research, defining criteria for the inclusion of participants, developing interaction scenarios, identifying appropriate observation

channels, designing the study setup, and finally developing the procedure.

3.1 Ethics Statement

The parents and therapists in the cooperating centers were informed about the study. The main assumption was that a child would be under no circumstances forced to interact with a robot and might quit at any point in time. The parents of the children provided written informed consent, in which they declared agreement for their child to participate and (optionally) being recorded. They also gave consent to process their children's data within the GDPR. The Ethical Board of the Gdansk University of Technology, Poland, approved the study protocol and raised no objections.

3.2 Participants

To ensure the credibility of the studies conducted, criteria for the inclusion of participants were defined. Three criteria have been specified:

- children between the ages of 2 and 12 years,
- formally diagnosed with ASD,
- without any other known neurological or psychological diagnosis.

In addition, it was decided that both treated and untreated children could participate, moreover, a level of functioning was also not specified, although we were recruiting children that had at least some imitation skills.

As outlined, the observational studies were conducted in four collaborating countries (the institutional abbreviations for the partners are reported between parentheses): Macedonia (MAAP), Turkey (ITU-YU), United Kingdom (UH), and Poland (GUT). Data on the number of children, their ages, and the number of repeated sessions are shown in Table 1.

3.3 Kaspar Robot

Human communication involves many subtleties (e. g., in facial expressions, in gestures, in body language, in speech etc), making it especially difficult for children with ASD to process into a coherent and meaningful whole. A common characteristic of children at the middle- to lower-end of the autistic spectrum is the difficulty to cope with social interaction, experiencing this as unpredictable, overwhelming, and frightening, causing anxiety and often withdrawal. A socially assistive robot designed to address some of these difficulties could be used as an effective tool to assist these children [12].

Kaspar (Fig. 1) is a social humanoid robot developed by researchers at the University of Hertfordshire, UK, that was specifically designed to help children with ASD develop social interaction and communication skills [43]. It is a child-size robot and has been purposefully designed with realistic but simplified human-like features offering a more predictable form of communication, making social interaction simpler, non-judgmental, and more comfortable for the child. Kaspar has a child-like appearance, in a sitting position and is approximately 56 cm tall. The robot has 22 Degrees of Freedom (DOF) and is equipped with sensors, cameras, and vocal communication that allow it to respond to external stimuli. Kaspar is capable of a range of movements, gestures, and facial expressions, (e. g., eye movements, blinking, nodding, shaking its head, waving its arms, open mouth and smile, portraying 'happy' or 'sad' expressions).

Kaspar can be controlled using a remote control keypad that is an integral part of the robot setup. Each of the interaction scenarios has an overlay that is placed on the keypad with relevant symbols/emojis or pictures on each key. The keypad can be used by the adult not only to operate the robot, but also as part of the interaction with the child, motivating the child to take initiative or respond, and sometimes giving the keypad to the child to control, build their confidence, allow them to take initiative and/or manage a collaborative game with the robot or another person. In addition to the keypad, the robot responds autonomously (within the context of the game scenario) when one of its touch sensors is activated.

Kaspar has been used in studies at schools, families' homes, and clinical centers with about 300 children (long-term studies where each child interacts with the robot over several weeks or months) [12]. Kaspar can engage children with ASD in a variety of therapeutic/educational games, e. g., turn-taking, joint attention and collaborative games, cause and effect games etc, that encourage the children to interact with the robot as well as using the robot as a mediator in interaction with other people (peers and adult care givers).

3.4 Interaction Scenarios

Kaspar is a robot designed for social skills training. Each scenario of Kaspar's interaction addresses some social skill components such as joint attention, involvement in interaction, turn-taking, vocalization imitation, etc. It follows a learning through play paradigm. For the study, a subset

Table 1 Study participants

Institutional Abbreviation	# of Children (Gender)	Age Range	# of Sessions (Min,Max)
MAAP	11 (9 M, 2 F)	2-6	(2,11)
ITU-YU	12 (11 M, 1 F)	6-10	(1,2)
UH	7 (6 M, 1 F)	10-12	(2,3)
GUT	3 (M)	6-6	(1,4)

Fig. 1 Kaspar, courtesy of The Robotics Research group, University of Hertfordshire, UK

of interaction scenarios with the Kaspar robot was chosen or implemented with the aim to evaluate the feasibility of available automatic emotion recognition technologies in robot-assisted intervention settings. Each scenario with Kaspar is based on the principle of turn-taking, imitation, and role changing. These interaction scenarios require basic receptive language skills related to emotions, animals, and body parts, as well as movement and vocalisation imitation skills.

In a *Standard* (starting) scenario, Kaspar introduces itself, and plays a movement imitation game with the child. The scenario does not require high communication skills; once Kaspar introduces itself, the therapist prompts the robot to perform some upper body gestures using only its arms, such as raising up its right hand, pointing left with its left hand, etc. and asks the children if they can copy Kaspar's movements (see Fig. 2). If the child performs correctly, then, a positive feedback is triggered by the researcher/therapist operating the robot, if not, a neutral feedback is given and the robot asks the child to try again. There is also a song included in this scenario, and it is used for familiarization with the robot (icebreaker) or draw the attention of the child back to the robot when a child gets bored or distracted. The song is "If you're happy and you know it" or a local equivalent per country.

In the *Emotions* scenario, Kaspar performs emotional expressions such as: happy, sad, surprised, and scared, followed by a tired expression. Expressions might be launched interchangeably, and the robot shows them with the movement of hands and limited facial (mouth and eyelids) actions. Then, the robot asks a child to mimic the expression. As in all scenarios, there is a number of prompts, reinforcements, and a song as outlined included as well, to keep a child interested and engaged.

In the *Animals* scenario, animal names and sounds are the theme of play. Kaspar asks the child what does the dog, cow, cat, duck, pig, and monkey say and requests the child to imitate these animals both verbally and behaviorally. The robot also 'says' the animal sounds to engage and amuse the children. This is a basic turn-taking scenario, however, children sometimes just enjoy hearing and observing rather than performing sounds.

The *Body parts* scenario consists of a pointing game where Kaspar asks the children to show their head, nose, mouth, eyes, ears, toes, hands, etc. When it is the robot's turn, it points out and says the name of the corresponding body part as well. Apart from naming body parts, this scenario practices imitation, turn-taking as well as my-your body part differentiation.

The *Vowels and syllables* scenario tries to pursue active speech, as well as turn-taking in verbal activities. Kaspar and a child play a turn-taking game of repeating vowels and basic syllables. For children with limited speech, it is an occasion to practice vocalisations, while for children with more advanced speech skills, we make it a memory game—a child repeats a sequence of vowels or syllables—making this scenario to practice attention and short-term memory.

Each session starts with Kaspar introducing itself, prompting a child's name, and inviting the children to play with it. Sessions end with Kaspar saying "Bye-bye, see you next time" or "Thank you". All scenarios include positive and neutral feedback phrases as reinforcements, as well as the "Auch, you're hurting me!" reaction when a child acts too harshly upon a robot. The latter reaction was used in some centers for teaching children to respect physical boundaries.

The verbal feedback and the songs are displayed in the native language of each country. Some songs were—as also outlined above —adjusted to local requirements as well.

3.5 Observation Channels and Experimental Setup

Automatic emotion recognition can be based on different observation channels and modalities derived from them. In the study, we decided to use multiple channels in parallel, in order to compare and combine them. The observation channels were selected based on their ability to capture symptoms of emotion in child-robot interaction and a final set of modalities included: facial expressions, posture, eye gaze, speech prosody, and physiological signals. The setup for the interaction and data collection was implemented in an allocated and reserved room (Fig. 3). As we aimed at the comparability of the results, all of the partners in the four countries performing the studies used the same set of equipment, as agreed in the project, for conformance of observations. The equipment we used in our observational studies is given as follows:

Fig. 2 A child mirroring Kaspar's arm movements

- Empatica E4 wristband for capturing physiological signals;
- Gazepoint GP3 Eye tracker to capture eye-gaze;
- 2 microphones: Zoom H4n Pro and AKG C 417 L lapel microphone with an adapter;
- 2 cameras (facial and capturing scene)—cameras were not standardized.

The criteria used for device selection were as follows: the lowest intrusiveness of measurement for a child; a possibility of long-term measurements; robustness to disturbance; a data export function; and quality to price ratio (bearing on mind potential future mass deployment).

The physiological signals were collected with a smart wristband. The children were equipped with the Empatica E4 placed on the wrist during the interaction session. The facial expressions of children were collected using two video cameras, one placed above Kaspar (as seen in Fig. 3 on the transparent screen ('cam 1')) to capture the facial expressions of the children, and the second one placed on the right side of the robot ('cam 2') to capture the whole test setup. In addition, the entire session was recorded with a supplementary video camera positioned to monitor Kaspar's movement. Furthermore, gaze movement, duration, and fixation data were captured by a Gazepoint Eye Tracker, positioned below Kaspar's seat. Finally, audio recordings of children were captured by a H4n Pro sound system, which was placed under the table, closer to the children's side, to prevent the cable clutter on the table.

As the main goal of this paper was to capture challenges and lessons learned with regard to automatic emotion recognition solutions, we were analyzing all the captured inputs

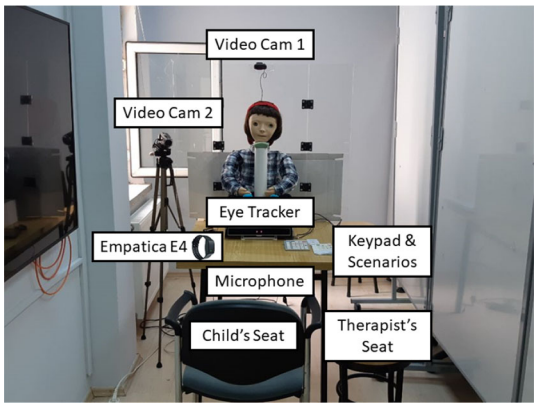

Fig. 3 Data collection setup for EMBOA user studies

with regard to their availability, the value they brought, as well as their limitations.

3.6 Procedure

Some general assumptions for observational sessions were as follows:

- all setups should be ready before a child enters the room;
- multiple sessions might be held with a child;
- familiarisation sessions are encouraged;
- there is a pre-prepared common list of scenarios to perform;
- the scenarios are to be translated into national languages and adjusted, if necessary (for example children sing different songs in different countries);
- we try to follow the specified scenarios, but it was allowed to follow a child—dropping or adding other interactions in between on the run;
- during the session, we write down the most important observations and the session might be annotated afterwards;
- we record, store, and share data (anonymized and coded) within the consortium.

The child is accompanied to the room by his/her therapist; if present, the parents are also allowed to monitor the interaction. When the child enters the room, (s)he is asked to sit in front of Kaspar. The child is then equipped with the E4 wristband. The placement of the camera, microphone, and eye-tracker are all checked, and the eye-tracker is calibrated. Therapists and researchers introduce themselves and Kaspar to the children. The interaction starts when Kaspar says, "Hello, I am Kaspar". Then, the flow of the entire interaction session is determined by the therapist or experimenter considering the child's profile or reactions. The interaction scenarios are not ordered, and they can be repeated multi-

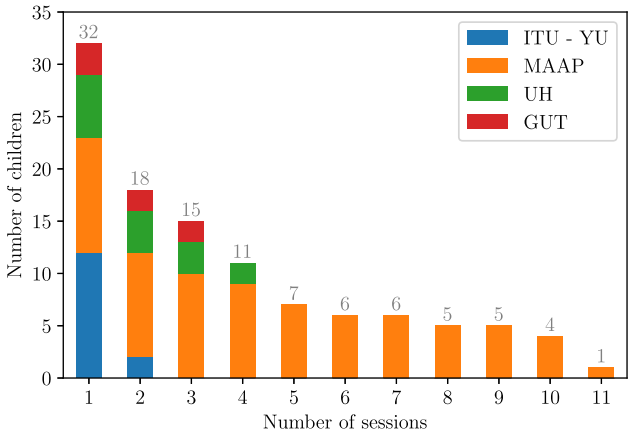

Fig. 4 The number of sessions and the number of children for each session

ple times, or some scenarios can be skipped if the child gets bored, or is not willing to cooperate. Most of the time, the first interaction is 'singing', and it helps children to get familiar with the robot. The interaction takes at least 10 to 15 min if the child is responsive and willing to continue. The duration of the sessions depended mainly on the developmental stage and initial level of social skills of a particular child. There were differences between countries, but these were due to age and ability of recruited participants rather than their cultural background.

Sometimes, the children notice the robot is controlled via the keypad, so they may want to use the keypad themselves to control the robot. If the therapist approves, then the child can also control the robot to initiate some actions. Ideally, the child performs the tasks during the interaction scenario by following the robot's instructions. However, some children need to be encouraged by the therapist, so the therapist repeats the task instruction to the child. The interaction ends with Kaspar saying, "Bye bye, see you next time".

Based on the profile of test participants (children with ASD), subjective evaluation questionnaires were not possible for feedback. However, to compensate for the lack of subjective evaluation of children and their impression of the solution, the comments of the therapists and experimenters have been used. The number of children and the number of sessions they participated in were displayed in Fig. 4.

4 Challenges Regarding Environment

4.1 Room and Accompanying People

Children with ASD are sensitive to external conditions they are in, including room, equipment, and people. Sometimes, they are reluctant to novel circumstances.

The robot was new to the participants, but the novelty of the robot, including the state of surprise, was the factor we wanted to observe, while the novelty of the room, people, or other equipment was a confounding factor. Therefore, we refer to the novelty of the environment as a challenge, while the novelty of the robot is considered a natural element of the observed interaction.

We aimed at finding a quiet, isolated room and minimize the number of people present. However, this was sometimes difficult to obtain. A minimal number of people involved a Kaspar operator. We also allowed a child caregiver in the room in order for a child to be more comfortable. Frequently, there were more people involved and the room we found was not so quiet, with external sounds coming as a noise to child vocalisations recording. With regards to recording the voice channel, also room echo and furniture movement or usage was generating additional noise. As we created a complex recording environment, apart from Kaspar's operator, also a person for starting/stopping/adjusting devices was present. Sometimes, additional therapists were present as well for the child's comfort or observation, simply. There were a few children who needed more time to start playing with Kaspar, as the environment was new. Moreover, some children refused to start playing and were allowed to walk out. The reasons include: the room arrangement, number of people present, or Kaspar itself. In some cases, familiarisation sessions helped children overcome their reservations.

4.2 Placement of Equipment

The research we report in this paper, includes a feasibility study before making robots equipped with a set of devices to capture emotional symptoms; therefore, we were experimenting with camera, microphone, and eye tracker placement. There were two cameras (facial and general) and two microphones (table and lapel) used. Some children refused to wear the lapel microphone on their clothes. Others agreed to wear the lapel microphone, but sometimes noise due to the movements of the clothes appeared that hampered the child voice activity detection. Background noise in general turned out to be problematic with regard to voice activity detection. Most child vocalisations were detected if the child's voice was clearly audible on the audio file, was not overlaid by the voices of others, and if there was no or only little background noise from furniture, other people, etc. [28]

The placement of the camera was problematic as well. The facial camera frequently did not capture child's face as they moved around, leaned forward or sideways (part of the face visible), were seated and standing interchangeably. The same challenge applies to the eye tracker, which generally requires a steady head position and calibration to work properly. The general camera captured the scene and was better for behavior

and posture analysis, but not for facial expressions, as the face was a small fracture of an image.

Some children were sensitive to wearable devices and did not want to put on the E4 on their wrists. They got uncomfortable and did not cooperate with the robot as long as the wristband was on their wrist. Sometimes, colored bandannas were used to cover the wristband as an accessory to minimize the children's discomfort. If a child refused to wear a wristband, the session continued without one. In total, 11% of sessions were recorded without a wristband.

4.3 Interaction Challenges

Children were mostly happy to see the robot, and eager to take part in the scenarios especially after they get used to the robot. Only two children with autism refused to approach the robot.

The song scenario was generally the most popular, even if the children who did not sing along with the robot, clapped synchronously during the song as a part of the interaction emerging between the child and the robot. In Poland, the therapists were also inviting typically developing children to interact and they observed that they are reluctant to play with a robot, in contrast to children with ASD. It is worth mentioning that the robot we used was designed for autism therapy, and its limited expressions, movement, and appearance was adjusted to fit the needs of those children.

It was practically impossible to keep the planned linear scenario order. For children to keep interacting, we had to adjust and mix, going back to favourite types of interaction. Rather than keeping to a preliminary plan, we followed a child in that matter, making it more interesting for a child, but less suitable for research.

A few children did not comply with playing with Kaspar. They refused to get closer to the robot (only 2) or sit in front of it, and some of them tried to hit the robot and harm the keypad and its accessories. There was a verbal and behavioral manifestation about their impression—the children who were reluctant to play with the robot got stressed, irritated, and angry. They started to display negative behaviors and wanted to leave the allocated room. The reasons behind it might be robot-related, but also environment (room) related, or based on the general reluctance to novelty of a particular child. Some of those children were highly sensitive to noise. Even though Kaspar's sound level was adjusted, they plugged their ears during the interaction, and they did not cooperate. We allowed a child to leave the room, with therapists motivating, but not forcing child to get into interaction.

Most children were interested in interaction and play, including being interested in additional equipment like cameras, etc. as well as Kaspar's operating keypad. Once the interaction happened, children were able to practice imitation, turn-taking, and verbal skills.

5 Challenges Regarding Emotion Processing Technologies

5.1 Facial Expressions Analysis

Facial expression analysis is one of the most widely used and efficient methods of emotion recognition. In order to identify problems with emotion recognition from video recordings of the faces of children on the autism spectrum during interaction with a social robot, an analysis of the collected recordings was conducted. The output files with recognized emotions were obtained by processing video files with the FaceReader software [11], version 9, released in 2021. We are aware that newer deep neural network-based emotion recognition systems based on facial expression analysis are now available; however, our focus in this study was on channel availability rather than emotion recognition accuracy.

In order to determine the capability of emotion recognition from video recordings, we evaluated availability of the channel and the possibility to analyse facial expressions. Each frame was marked as FIND_FAILED—could not find the face, FIT_FAILED—could not fit the face model, or DETECTED—emotion (facial expression) was detected.

The chart in Fig. 5, shows the results of the automated analysis for 52 session recordings, for different children from different centers. Green indicates the proportion of times when any emotional state, including neutral, was detected, yellow when no face was detected, and red when no emotional state was detected. In only three cases were emotions recognized for more than 80% of the duration of the recording, and for as many as 36 (69.2%), emotions were recognized in less than half of the recordings time.

All videos were then manually reviewed to identify problems that might have affected the level of emotion recognition. Among the most common issues that recurred in recordings with low emotion recognition rates were the position of the camera, i.e., too far away or at too great an angle to the child's face, and overexposure of the video. Further factors were related to the child's appearance (e.g., thick glasses, long fringe) or behavior—lowering head, leaning forward, looking around, or covering the face with a hand. Some of the behaviors were even parts of the scenarios we planned— e.g., touching the nose caused face occlusion.

5.2 Eye Tracker-Based Analysis

The most commonly used features of eye-tracking sensors for automatic emotion recognition are pupil diameter and fixation duration. Thus, in order to evaluate the availability of this channel, we decided to analyze two parameters from each available eye-tracker recording: (1) a percentage of the time when the eye gaze was detected; (2) a percentage of the time when the fixation point was detected. Without the first

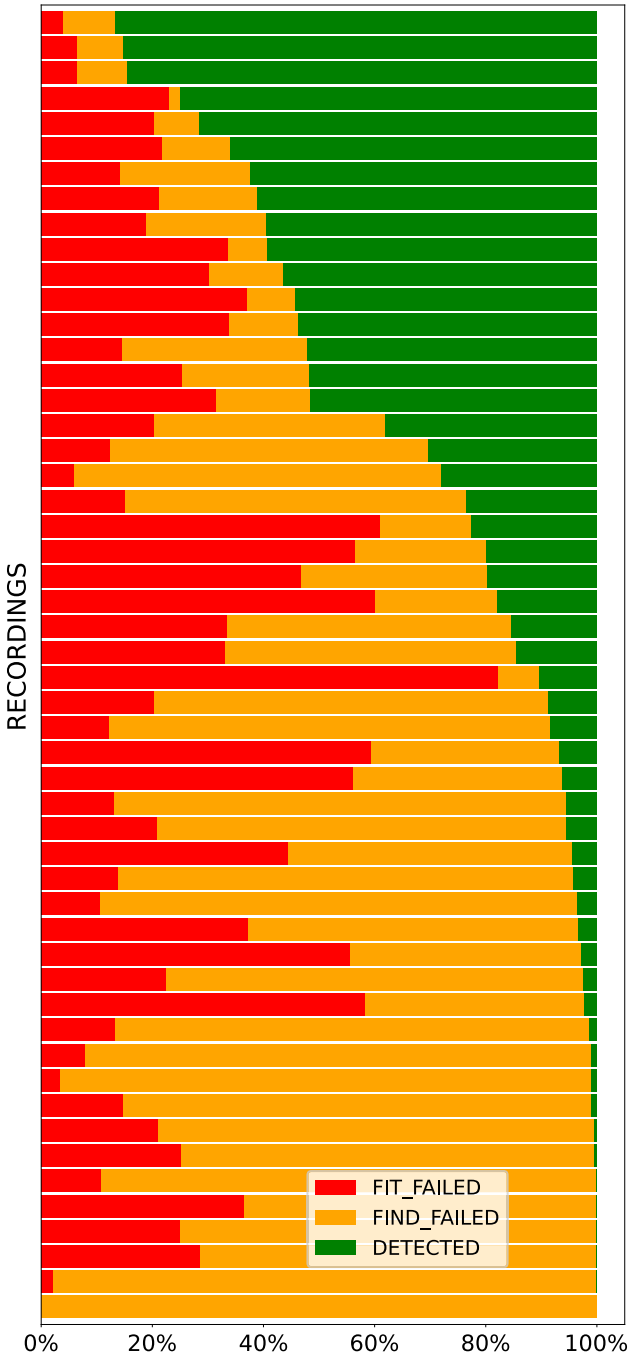

Fig. 5 Distribution of successful (green) or unsuccessful (other) moments where attempts at emotion recognition could be made

parameter, i. e., without an eye gaze detected, we are unable to obtain any eye-tracking feature, including a pupil diameter. The second parameter is the total duration of fixation in the recorded session. An overview of the results is presented in Table 2.

From our analysis, it can be concluded that the eye-tracking data has a limited availability on the designed environment with children with ASD. It was a disappointing finding, since current researches reported promising results for automatic emotion recognition from eye- tracking data [24, 40]. Therefore, each session recording was examined by an expert in order to find a possible explanation for the problem. The most common recurring issues were identified in three areas: (1) technical, (2) interaction characteristics, and (3) child behavior.

The technical problems are usually due to the eye tracker being placed incorrectly with regard to the child's face. It was pointing too high from one side when it should be pointing frontally. In some recordings, a child was sitting too far away from the eye tracker, or the room was too dark. One of the most frequent technical issues, also related to children's behavior, is the way in which the eye tracker finds a fixation point. This is achieved by moving the gaze, while younger children change their point of focus by moving their head rather than their gaze.

Many issues are a result of the characteristics of a session. The eye tracker works best when a person is sitting still. However, children are asked to move and play during the robot-based interventions. One of the tasks even required them to cover their face with their hands. This meant that the children's eyes were not visible for some of the time. In some recordings, especially with younger children, the therapist's hand is visible between the child and the eye tracker. This also reduces the time that eye gaze can be detected.

The most common issues related to a child's behavior are the movements of the child and closed eyes. Some of the situations are related to an intervention scenario, while others are specific to a particular child, e. g., younger children tend to move and cover faces more.

Apart from eye gaze detection from the eye tracker, we have also tried the video-based approach, such as OpenFace, and GazeTracker solution developed by one of the partners. The gaze recognition rates were comparable, while the challenges related to the children's face turning and coverage, and interaction scenarios remained.

5.3 Automated Voice and Vocalization Analysis

The analysis of speech or, in general, vocalizations is another common way to gain insights into the affective state of an individual. The main advantages are generally the non-invasive nature of audio recordings, a (relatively high) robustness to the device placement [29], and a particular manifestation of arousal in the acoustic signature of the voice[39, 42].

A main disadvantage, however, is similar to the other modalities— the availability of the modality. Vocalizations are not always present, especially in unscripted intervention sessions with autistic children. Additionally, automatic analysis models may confuse the vocalizations of different persons when analyzing the affective state of a single indi-

Table 2 Overview of results on child eye gaze activity detection

Activity	Unit	Eye gaze detection	Eye gaze fixation
Session duration;			
mean ± std	[s]	592.00 ± 218.00	592.00 ± 218.00
Time with detected child eye activity per session; mean ± std	[s]	171.90 ± 119.39	3.54 ± 4.82
Proportion of time with detected child eye activity per session; mean ± std	[%]	28.75 ± 17.21	0.57 ± 0.69

vidual. As a result, the availability of the voice of a child was available as little as 6 % of the session time. For this reason, it is common practice to use voice activity detection (VAD) systems prior to the application of speech emotion recognition (SER) systems [4]. In [27], a specific voice activity detection system and a cascaded speech emotion recognition system, both based on deep learning architectures, were trained based on vocalizations of children with ASD in a robot-assisted intervention setting, which serve as a basis for evaluation here. Generally, performance drop-offs are to be expected when applying a deep learning model to out-of-domain (OOD) data. Our study setting shows overall quite some similarities with the data on which the SER pipeline [27] was trained, as both are concerning intervention sessions with autistic children. Major challenges in these settings can arise from differences in recording devices, the acoustic response behavior of the observation rooms, the age distribution of participants, or most notably, the language and cultural background of the participants. These aspects hinder the robustness of VAD and SER systems for children with ASD, which in itself is a challenging task given the peculiarities in the affect expression of children with ASD. Nevertheless, we used the deep learning-based tools provided by [27] to analyze the effects of varying circumstances in an application scenario, the results of which are reported in [28].

In that study, we were able to analyze our recordings with two different microphones in the same situation—except for a few sessions with technical problems or a child not wanting to wear a lapel microphone. Different outputs of the analysis tools could thus be directly linked to different recording settings, which manifested itself in different types of microphones and their placement; one microphone was placed centrally in the room, while the other was attached to the child or researcher. Concerning VAD, we found that the different recording settings only had a clear effect on the amount of voice activity detection events in the Polish study with almost twice as many detections recorded with table microphone, while the other study arms showed a similar amount of detection events between microphones (only up to 27% more detections with the table microphone). Beyond, an analysis with the SER system showed that the different microphones (lapel or table) had a relatively small impact on estimated valence and arousal prediction, as the mean absolute deviations between predictions based on the two microphones stays below 0.1. However, larger deviations occurred if we applied different SER systems, i. e., one SER system with VAD-based pre-filtering and one without. The highest mean difference between the two system predictions occurred with 0.208 for arousal. In the provided results, we could thus observe that the choice of SER systems had a seemingly higher impact on outcomes than the choice of microphones. Nevertheless, we were not able to make estimations about the accuracy of the applied analysis tool, as for data collected within the EMBOA project, no ground truth labels were provided [28].

5.4 Physiological Signals Analysis

Another modality for the investigation of the affective state of children during their interaction with the robot is to monitor their physiological signals. The physiological signals may provide insights into the emotional changes that other modalities, such as facial or vocal expressions, would not manifest.

In this study, the blood volume pulse (BVP), electrodermal activity (EDA), and skin temperature (ST) data of children were collected by the Empatica E4 smart wristband while they were interacting with Kaspar. In the first round of interaction studies, 32 children volunteered to interact with the robot in one or more than one sessions, separated periodically. The collected data were analyzed, and the data with low quality of the signal were excluded from the E4 data set—details on the evaluation of the signal quality follow.

To detect the affective state of children, first, a set of heart rate variability (HRV) features were extracted from the BVP signal. The 6 HRV features used for this assessment are the standard deviation of NN intervals where NN or R-R intervals are described as the period between two consecutive heartbeats (SDNN), the total number of NN (R-R) intervals divided by the fraction of NN50 intervals where NN50 is described as the number of times two consecutive NN or R-R intervals diverge by more than 50 milliseconds (pNN50), the root mean square of the successive differences (RMSSD), and the heart rate (HR) mean as time-domain features; whereas the low (LF) and high (HF) frequency power and their ratio (LF/HF) serve as frequency-domain features. The features and the corresponding reference intervals pre-

Table 3 Reference intervals for the HRV features extracted from the previous studies in the literature

Ref Study	Age Range	SDNN (ms)	pNN50 (%)	RMSSD (ms)	HR Mean (bpm)	HF (ms2)	LF (ms2)	ASD	Gender
[10]	18–33	70.12 ±28.03	19.78 ±9.46	71.44 ±53.99	None	910.10 ±915.31	1020.00 ±1016.20	No	M
[10]	18–33	70.90 ±29.13	23.17 ±10.27	77.431 ±35.82	None	731.79 ±901.30	1261.69 ±1237.95	No	F
[35]	6–8	25.00–116.00	2.20–71.90	22.00–149.00	69.20–101.80	221.00–3124.00	188.00–7690.00	No	All
[38]	7.40 ±1.10	133.00 ±32.00	24.00 ±11.00	75.00 ±50.00	None	None	None	No	F
[19]	4–17	51.00–236.00	6.00–48.00	25.00–92.00	99.62 99.62	406.30–2200.60	173.30–1612.10	No	All
[16]	7–12	154.10 ±40.20	None	44.00 ±35.40	None	1889.50 ±1116.00	2308.40 ±1958.40	No	All
[41]	10.70 ±0.90	None	None	None	72.00 ±9.00	2243.00 ±3230.00	3127.00 ±3911.00	Yes	All

viously reported in the literature are displayed in Table 3, as well as the demographic profile of study participants.

A scoring method was used to measure the quality of the collected signal based on the reference studies given in Table 3. New reference intervals were extracted to assess the quality of the collected signals: For each HRV feature, minimum and maximum values were taken to be the broadest range considering every row. For example, the SDNN minimum value was specified as 25 ms, and the maximum was 154.10 ms. If the feature values extracted from the collected signal were between the predetermined intervals, then the signal was labeled as 'good quality' for the analysis and scored 1. If not, it was labeled as 'not okay' and not scored. Since the selected set of HRV features consists of 6 signals, the total quality score ranged from 0 to 6. The scored recordings and their quality are shown in Fig. 6.

The recordings with a total quality score equal to or higher than 4 were selected for the physiological signal analysis based on the HRV features: While 68 % of the recordings were labeled as 'good quality', the remaining recordings were discarded.

In order to monitor the affective state of children, we conducted a preliminary study to detect their stress level based on the LF and HF power extracted from the BVP signal. LF power has been shown to increase, and HF power to decrease during stressed conditions. When the LF and HF features were analyzed, the results revealed that in most cases, they output conflicting results. Even though in specific cases, the LF and HF were accurate in monitoring the fluctuating stress level of children, higher at the beginning of the interaction session and lower after the familiarization with the robot, the findings showed that multi-modal data is crucial to achieve better accuracy in the monitoring and the prediction of the affective state; for further information on the stress detection study, see [14].

In another study, we combined the LF and HF features with the peak count and signal amplitude extracted from the EDA signal to explore if there is any improvement in the stress detection task. Similar to the previous study, although in specific cases the combination of the signals provided accurate results indicating the accurate level of stress for the children, the results of the physiological signal analysis were not accurate in detecting the stress of the children; for further information, refer to [2]. We used an automatic annotation tool in both studies to identify children's facial expressions from the video recordings. The physiological signal analysis results were validated with the emotion labels extracted from the session recordings, highlighting the contribution of the multi-modal data collection.

6 Summary of Results and Discussion

The research question of the study presented in this paper was to indicate how to observe the emotional states of a child with autism interacting with a robot. In order to answer that question we performed an international observational study, that allowed us to name some challenges and lessons learned, as summarized below:

- challenges that exist in observing the emotions of a child with autism are of various nature—technical (those are the simplest ones to address), procedure or task-based, and related to the specificity of the participant group;
- observational studies with children with autism are challenging—there are multiple factors to take into

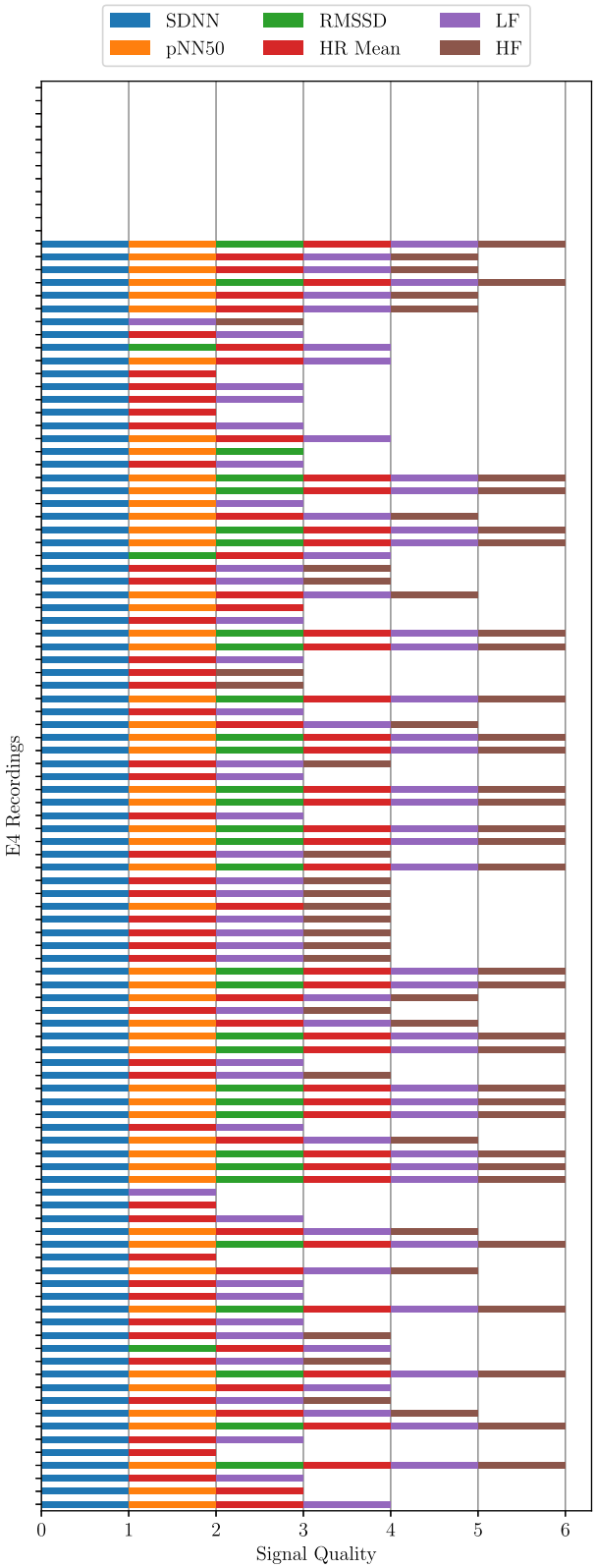

Fig. 6 The quality of E4 recordings based on HRV features

account, as such of the respective children potentially being demanding study participants—one has to take into account that despite careful planning and efforts, some children would not involve in interaction at all;

- the more complex environment we create and the more sophisticated equipment we use, the more technical problems might arise, including device placements, etc.;
- all observational channels we used had limited availability—none of the analyzed modalities (facial expressions, eye gaze, vocalizations, physiological signals) was available the entire time, and for some, their availability is really low;
- availability of the facial expression channel differed from session to session—in three sessions only any emotions (including neutral state) were recognized for more than 80 % of the duration of the session, for as many as 36 out of 52 sessions (69 %) the facial expression modality was available for less than half of the session time;
- availability of the sound channel was even lesser—during only 6 % of the sessions' time, vocalizations of a child were available—this is a result of not every task in the scenario being based on speech, as well as common difficulties with productive language in children with autism;
- availability of eye tracker data was also limited—although a child's eye gaze was detected during 29 % of the session time, and the eye gaze fixation was available for less than 1 % of the session time—we presume, that this outcome is a result of eye trackers requiring precise calibration and lack of major head movements to work properly and those two factors were not possible with the given participants;
- physiological signals, apart from sessions when a child refused to wear a wristband, were the most available—68 % of the recordings had quality good enough for further analysis, and this result is promising;
- there are a number of challenges with regard to procedure and interaction itself, as children with autism are demanding partners and all the planned scenarios sometimes were mixed or discarded to follow a child; although for research, it would be more beneficial to follow a predefined standardized procedure, it is worth remembering that the most important thing is to keep a child interested and engaged in interaction with a robot, as this is the main indicator influencing possible benefits for training skills of a child.
- the question on how much of the challenges are due to the nature of ASD rather than sensor or data limitation remains open— it is heavily dependent on the study—the equipment used, the robot itself, implemented scenarios, and participants. The proportion between the challenges' categories might be changed by improvement of the procedures and sensors, which are easier to address than challenges related to the specificity of ASD.

We are aware that our study, although carefully planned, is not free from some limitations. The main validity threats of the study and our mitigation methods are listed below, and further explained:

- individual differences bias,
- disposition-of-the-day bias,
- Hawthorne effect/context effect,
- instrumentation effect,
- maturation/history effect.

Individual differences bias is a threat to the external validity (generalizability) of the study. This threat is especially present in studies involving children on the autism spectrum, as they might significantly differ in disorder severity, intellectual abilities, skills, behaviors, and level of other limitations. In order to address this risk, we planned to invite 20 children, but we recruited even more (33). Most studies with children with autism and robots invite up to 10 children, as we found out in a literature study [9], so it seems that our study is among the largest. Moreover, we added inclusion criteria, such as age and formal diagnosis and additionally, we have asked some preliminary questions on the severity of autism, and the level of basic skills.

Disposition-of-the-day bias is immanent in all human-related studies and in our opinion, it applies in particular to children with autism due to the specificity of the disorder, which might influence the obtained results. We have foreseen this issue and addressed this in the research methodology of our study—we planned multiple sessions per child. Moreover, we added a before-session question (to caregivers or therapists) on whether anything special happened before the session that might influence the child's behavior.

There are two effects that refer to the situational context of an observation—one of those is the Hawthorne effect (people behave differently when knowing they are observed—also known as observer's paradox) and the context effect (the environment and circumstances of observations influence subjects of observation). As participants of the study were mainly children at the kindergarten or early school level, we think that the Hawthorne effect was minimal. However, the context effect was severe—children with autism are in general not open to new rooms, situations, and people, and all those circumstances were present in our study. We were unable to mitigate this risk, the only thing we could do and did was to set up the sessions in therapeutic centers (environments they knew) rather than inviting children to labs at the partners' premises. We also encouraged Kaspar familiarisation sessions before the actual measurements.

The instrumentation effect (using different instruments in groups/locations) had a limited influence on our study. All partners that recorded sessions were equipped with the same microphone sets, eye gaze trackers, and wristbands recording physiological signals. The effect applies to cameras—we used the cameras available to partners, not specifying their characteristics.

Another validity threat is the maturation/history effect (making the test/measurement/task for the second time influencing the result) and that effect was present in our study, as most of the children had multiple sessions. We even encouraged familiarisation sessions, for children to feel more comfortable. We noticed differences between the sessions in child behavior and emotional states, and as our goal was to analyze emotion recognition-related channels, this effect worked for the benefit of our study.

7 Conclusion & Future Work

The observational study reported here explored us potential difficulties and challenges that can occur when aiming to apply state-of-the-art automatic emotion recognition techniques in robot-supported intervention sessions with children with autism. We are aware that it will not be possible or reasonable to solve all observed issues. The solution to certain problems related to automatic emotion recognition, such as talking less to the child in order to increase the voice activity detection performance on the child's speech or avoiding eye contact with the child in order to recognize the face all the time from a certain angle, would likely reduce the engagement and enjoyment of robot-supported intervention sessions for the children and/or the effectiveness of the intervention. Still, we identified a number of issues that could be adapted to maximize the child's comfort and/or to likely improve (multimodal) automatic emotion recognition in this specific setting without risking reducing the intervention's success. The detailed recommendations on using particular channels would best be summarised in the form of detailed guidelines which – due to the size – could not be presented in this paper. We are aware that we did not analyze emotions versus interactions and scenarios, and we plan to do this in future works.

The general purpose of this study, however, was accomplished – we studied a number of observation channels frequently used in automatic emotion recognition and summarised the findings. This study was a preliminary feasibility study for future works on how to extend social robots so that they can perceive affect of children and act upon it. Our findings showed that voice and eye gaze had limited applicability here. Also, analysis of facial expressions has some challenges impossible to be fully addressed. The most available of the analysed channels – physiological signals, is more difficult to obtain in the context of human-robot interaction. Perhaps, multimodal observation is an option, however, the more sensors a robot has, the more expensive the production as well as the more complex run-time processing of emotional symptoms is. Having said this, we still find it promising for a robot

to perceive human emotions and be able to respond to them, adjusting activities and forming a human-technology affective loop. We believe that this research lays the groundwork for robots to automatically recognize emotions and interact with children with ASD in the future without additional external intervention.

Acknowledgements We want to thank all families, therapists, and researchers who participated in our study. This work was supported by the European Commission's Erasmus+ Project (EMBOA, Affective loop in socially assistive robotics as an intervention tool for children with autism) under Contract 2019-1-PL01-KA203-065096. The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Data Availibility Statement Data collected during the observation sessions is available upon request.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

1. Abdullah SMSA, Ameen SYA, Sadeeq MA et al (2021) Multimodal emotion recognition using deep learning. J Appl Sci Technol Trends 2(02):52–58
2. Aktaş SNB, Uluer P, Coşkun B, et al. (2022) Stress detection of children with asd using physiological signals. In: 2022 30th Signal Processing and Communications Applications Conference (SIU), pp 1–4, https://doi.org/10.1109/SIU55565.2022.9864668
3. Al-Nafjan A, Alhakbani N, Alabdulkareem A (2023) Measuring engagement in robot-assisted therapy for autistic children. Behav Sci 13(8):618. https://doi.org/10.3390/bs13080618
4. Alghifari MF, Gunawan TS, Qadri SAA et al (2019) On the use of voice activity detection in speech emotion recognition. Bull Electr Eng Inf 8(4):1324–1332
5. Alnajjar F, Cappuccio M, Renawi A et al (2021) Personalized robot interventions for autistic children: an automated methodology for attention assessment. Int J Soc Robot 13:67–82. https://doi.org/10.1007/s12369-020-00639-8
6. American Psychiatric Association D, Association AP et al (2013) Diagnostic and statistical manual of mental disorders: DSM-5, vol 5. American psychiatric association Washington, DC, USA
7. Aziz AA, Mokhsin M, Moganan FFM et al (2018) Humanoid-robot as teaching mediator: Research model in demonstrating the autistic children learning motivation based on the emotional responses. Adv Sci Lett 24(4):2296–2300. https://doi.org/10.1166/asl.2018.10939
8. Banire B, Al Thani D, Qaraqe M (2023) One size does not fit all: detecting attention in children with autism using machine learning. User Modeling and User-Adapted Interaction pp 1–33. https://doi.org/10.1007/s11257-023-09371-0
9. Bartl-Pokorny KD, Pykała M, Uluer P et al (2021) Robot-based intervention for children with autism spectrum disorder: a systematic literature review. IEEE Access 9:165433–165450. https://doi.org/10.1109/ACCESS.2021.3132785
10. Berkoff D, Cairns C, Sanchez L et al (2007) Heart rate variability in elite American track and field athletes. J Strength Cond Res/Natl Strength Cond Assoc 21:227–31. https://doi.org/10.1519/R-20135.1
11. Brodny G, Kołakowska A, Landowska A, et al. (2016) Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions. In: 2016 9th International Conference on Human System Interactions (HSI), IEEE, pp 397–404, https://doi.org/10.1109/HSI.2016.7529664
12. Cabibihan JJ, Javed H, Ang M et al (2013) Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. Int J Soc Robot 5:593–618. https://doi.org/10.1007/s12369-013-0202-2
13. Chen J, Ro T, Zhu Z (2022) Emotion recognition with audio, video, EEG, and EMG: a dataset and baseline approaches. IEEE Access 10:13229–13242. https://doi.org/10.1109/ACCESS.2022.3146729
14. Coşkun B, Uluer P, Toprak E, et al. (2022) Stress detection of children with autism using physiological signals in kaspar robot-based intervention studies. In: 2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob), pp 01–07, https://doi.org/10.1109/BioRob52689.2022.9925485
15. English BA, Coates A, Howard A (2017) Recognition of gestural behaviors expressed by humanoid robotic platforms for teaching affect recognition to children with autism-a healthy subjects pilot study. In: Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9, Springer, pp 567–576, https://doi.org/10.1007/978-3-319-70022-9_56
16. Fazli C (2019) Pediatric heart rate variability normative values related to average heart rate and age in a developing country. J Cardiovas Res 2. https://doi.org/10.33552/OJCR.2019.02.000547
17. Holeva V, Nikopoulou V, Lytridis C et al (2022) Effectiveness of a robot-assisted psychological intervention for children with autism spectrum disorder. J Autism Dev Disord. https://doi.org/10.1007/s10803-022-05796-5
18. Ismail LI, Verhoeven T, Dambre J et al (2019) Leveraging robotics research for children with autism: a review. Int J Soc Robot 11:389–410. https://doi.org/10.1007/s12369-018-0508-1
19. Karabulut M (2015) Salıklı Çocuklarda kalp hızı deişkenlii. Fırat Med J 20:152–55
20. Kose H, Akalin N, Uluer P (2014) Socially interactive robotic platforms as sign language tutors. Int J Humanoid Rob 11(01):1450003. https://doi.org/10.1142/S0219843614500030
21. Kouroupa A, Laws KR, Irvine K et al (2022) The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis. PLoS ONE 17(6):e0269800. https://doi.org/10.1371/journal.pone.0269800
22. Landowska A, Robins B (2020) Robot eye perspective in perceiving facial expressions in interaction with children with autism. In: Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-

2020), Springer, pp 1287–1297, https://doi.org/10.1007/978-3-030-44038-1_117
23. Landowska A, Karpus A, Zawadzka T et al (2022) Automatic emotion recognition in children with autism: a systematic literature review. Sensors 22(4):1649. https://doi.org/10.3390/s22041649
24. Lim JZ, Mountstephens J, Teo J (2020) Emotion recognition using eye-tracking: taxonomy, review and current challenges. Sensors 20(8):2384. https://doi.org/10.3390/s20082384
25. Liu C, Conn K, Sarkar N et al (2008) Online affect detection and robot behavior adaptation for intervention of children with autism. IEEE Trans Rob 24(4):883–896. https://doi.org/10.1109/tro.2008.2001362
26. Martinez-Martin E, Escalona F, Cazorla M (2020) Socially assistive robots for older adults and people with autism: an overview. Electronics. https://doi.org/10.3390/electronics9020367
27. Milling M, Baird A, Bartl-Pokorny KD et al (2022) Evaluating the impact of voice activity detection on speech emotion recognition for autistic children. Front Comput Sci. https://doi.org/10.3389/fcomp.2022.837269
28. Milling M, Bartl-Pokorny KD, Schuller BW (2022b) Investigating automatic speech emotion recognition for children with autism spectrum disorder in interactive intervention sessions with the social robot kaspar. medRxiv pp 2022–02. https://doi.org/10.1101/2022.02.24.22271443
29. Milling M, Pokorny FB, Bartl-Pokorny KD et al (2022) Is speech the new blood? recent progress in ai-based disease detection from audio in a nutshell. Front Digit Health 4:886615. https://doi.org/10.3389/fdgth.2022.886615
30. Pennisi P, Tonacci A, Tartarisco G et al (2016) Autism and social robotics: A systematic review. Autism Res 9(2):165–183. https://doi.org/10.1002/aur.1527
31. Pop CA, Simut R, Pintea S et al (2013) Can the social robot probo help children with autism to identify situation-based emotions? a series of single case experiments. Int J Humanoid Rob 10(03):1350025. https://doi.org/10.1142/s0219843613500254
32. Rudovic O, Lee J, Mascarell-Maricic L et al (2017) Measuring engagement in robot-assisted autism therapy: A cross-cultural study. Front Robot AI. https://doi.org/10.3389/frobt.2017.00036
33. Rudovic O, Lee J, Dai M et al (2018) Personalized machine learning for robot perception of affect and engagement in autism therapy. Sci Robot 3(19):eaao6760. https://doi.org/10.1126/scirobotics.aao6760
34. Sani-Bozkurt S, Bozkus-Genc G (2023) Social robots for joint attention development in autism spectrum disorder: A systematic review. Int J Disabil Dev Educ 70(5):625–643. https://doi.org/10.1080/1034912X.2021.1905153
35. Seppälä S, Laitinen T, Tarvainen M et al (2013) Normal values for heart rate variability parameters in children 6–8 years of age: The panic study. Clin Physiol Funct Imag. https://doi.org/10.1111/cpf.12096
36. Shi Z, Groechel TR, Jain S et al (2022) Toward personalized affect-aware socially assistive robot tutors for long-term interventions with children with autism. ACM Trans Human-Robot Interact (THRI) 11(4):1–28. https://doi.org/10.1145/3526111
37. Silva S, Soares F, Costa S, et al. (2012) Development of skills in children with ASD using a robotic platform. In: 2012 IEEE 2nd Portuguese Meeting in Bioengineering (ENBENG). IEEE, pp 1–4, https://doi.org/10.1109/enbeng.2012.6331347
38. Silvetti M, Drago F, Ragonese P (2002) Heart rate variability in healthy children and adolescents is partially related to age and gender. Int J Cardiol 81:169–74. https://doi.org/10.1016/S0167-5273(01)00537-X
39. Stappen L, Baird A, Christ L, et al. (2021) The MuSe 2021 multimodal sentiment analysis challenge: sentiment, emotion, physiological-emotion, and stress. In: Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge. p 5–14, https://doi.org/10.1145/3475957.3484450
40. Tarnowski P, Kołodziej M, Majkowski A et al (2020) Eye-tracking analysis for emotion recognition. Comput Intell Neurosci. https://doi.org/10.1155/2020/2909267
41. Tiinanen S, Mättä A, Silfverhuth M, et al. (2011) HRV and EEG based indicators of stress in children with Asperger syndrome in audio-visual stimulus test. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp 2021–2024, https://doi.org/10.1109/IEMBS.2011.6090371
42. Wagner J, Triantafyllopoulos A, Wierstorf H et al (2023) Dawn of the transformer era in speech emotion recognition: closing the valence gap. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2023.3263585
43. Wood LJ, Zaraki A, Robins B et al (2021) Developing kaspar: a humanoid robot for children with autism. Int J Soc Robot 13:491–508. https://doi.org/10.1007/s12369-019-00563-6
44. Yun SS, Choi J, Park SK (2016) Robotic behavioral intervention to facilitate eye contact and reading emotions of children with autism spectrum disorders. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp 694–699, https://doi.org/10.1109/ROMAN.2016.7745194
45. Zeng Z, Pantic M, Roisman GI, et al. (2007) A survey of affect recognition methods: audio, visual and spontaneous expressions. In: Proceedings of the 9th international conference on Multimodal interfaces, pp 126–133, https://doi.org/10.1145/1322192.1322216
46. Zheng WL, Dong BN, Lu BL (2014) Multimodal emotion recognition using eeg and eye tracking data. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp 5040–5043, https://doi.org/10.1109/EMBC.2014.6944757

Duygun Erol Barkana is currently a Professor with Yeditepe University, Istanbul, Türkiye and the Director of the Robotics Research Laboratory. She is working on robot-assisted rehabilitation systems, social assistive robots, human–robot interaction, and affective computing. She is a part of several national and international projects in these fields. Her research, supported by the European Union, and the Scientific and Research Council of Turkey, has culminated in many publications. She was honored with the Career Award by the Scientific and Technological Research Council of Turkey in 2009, the Outstanding Young Scientists Awards by the Turkish Academy of Sciences in 2018, and the Research Incentive Award by the Middle East Technical University Professor Mustafa Parlar Research and Education Foundation, in 2019.

Katrin D. Bartl-Pokorny received the M.Phil. degree in Applied Linguistics from the University of Graz, Austria, in 2010, and the Ph.D. degree in Medical Science from the Medical University of Graz, Austria, in 2019. She is currently a Senior Lecturer at the Division of Phoniatrics, Medical University of Graz, Austria, as well as a postdoctoral researcher at the Chair of Health Informatics, Technical University of Munich, Germany. She has (co)authored four book chapters and 51 journal articles. Her research interests include digital health and speech-language and socio-communicative abilities of typically developing children and children with developmental disorders.

Hatice Kose is currently a Professor with Istanbul Technical University, Turkey, coordinating the Cognitive Social Robotics Research Group and the Game and Interaction Technologies Laboratory (GAMELab). She is working on social assistive robots, human–robot interaction, affective computing, and gamification, mostly in the domain of education, therapy, and treatment of children with disabilities. She is a part of several national and international projects in this field, including the FP6 Project Robotcub, the RISE Project Welding of E-Textiles for Interactive Clothing (Etexweld), the LUDI COST Action Play For Children With Disabilities, and the Erasmus+ Project EMBOA Affective loop in Socially Assistive Robotics as an intervention tool for children with autism.

Agnieszka Landowska is currently an Associate Professor with the Gdansk University of Technology, Poland. She is also a Leader of the Emotions in the Human–Computer Interactions Group (EMORG) and Head of Software Engineering Department. She leads a project that develops mobile applications dedicated to autism therapy (Friendly Apps). She supervised research in project AFFITS (Methods and tools for affect-aware intelligent tutoring systems), AUTMON (Automated therapy monitoring for children with autism spectrum disorder), and EMBOA (Affective loop in Socially Assistive Robotics as an intervention tool for children with autism), and took part in UE COST LUDI (Play for children with disabilities) and SHELD-ON (Smart habitat for the elderly) projects. Her research is focused on making the technology more humane, including topics of human-computer and human–robot interaction, accessibility and adoption of technology, user experience, and affective computing.

Manuel Milling received his Bachelor of Science in Physics and in Computer Science from the University of Augsburg in 2014 and 2015, respectively and his Master of Science in Physics from the same university in 2018. He is currently a PhD candidate in Computer Science at the chair of Health Informatics, Technical University of Munich. His research interests include machine learning with a particular focus on the core understanding and applications of deep learning methodologies.

Ben Robins received the Ph.D. degree from the School of Computer Science, University of Hertfordshire, U.K., focusing on assistive technology for children with autism, bringing together his expertise and experience in these two disciplines. He is currently a Senior Research Fellow with the School of Computer Science, University of Hertfordshire. His qualifications and many years of work experience lie in two disciplines, computer science and dance movement therapy. His research, which started in 2002 in the AURORA Project and continued in the FP6, FP7, and H2020 European projects IROMEC, ROBOSKIN, and BabyRobot, respectively, investigates the potential use of robots as therapeutic or educational tools, encouraging basic communication, and social interaction skills in children with autism. His recent work, part of the Erasmus+ projects SMART and EMBOA, and UH's KASPAR Project, has further investigated robot-assisted therapy and continued the development of the KASPAR robot as a therapeutic and educational tool (http://kaspar.herts.ac.uk). This included running long term studies with KASPAR and children with autism and other learning difficulties in collaboration with schools and medical centers internationally.

Björn W. Schuller received the Diploma, Doctoral, and Habilitation degrees in machine intelligence and signal processing from the Technical University of Munich, Germany, where he is currently a Full Professor and Chair of Health Informatics. Prior to this, he was the Chair of Embedded Intelligence for Health Care and Wellbeing at the University of Augsburg, Germany. He is also a Full Professor of Artificial Intelligence and the Head of GLAM at Imperial College London, U.K., and the co-founding CEO and current CSO of audEERING, an audio intelligence company based near Munich and in Berlin, Germany. He has (co)authored more than 1,500 publications (more than 65 k citations and H-index = 115). He is a Golden Core Awardee of the IEEE Computer Society, and Fellow of the ACM, AAAS, BCS, DIRDI, IEEE, and ISCA, and President-Emeritus of the AAAC. His more than 50 awards include having been honored as one of 40 extraordinary scientists under the age of 40 by WEF in 2015. He is the Field Chief Editor of Frontiers in Digital Health, Editor in Chief of the AI Open Journal, and was the Editor-in-Chief of the IEEE Transactions on Affective Computing. He served as a coordinator/PI in more than 15 European projects, and is an ERC Starting and DFG Reinhart-Koselleck Grantee and a Consultant of companies, such as Barclays, GN, Huawei, Informetis, or Samsung.

Pinar Uluer received her MSc degree from Computer Engineering, Galatasaray University, and her PhD degree from Mechatronics Engineering, Istanbul Technical University, Turkey. She is currently an Assistant Professor at the Department of Computer Engineering of Galatasaray University (GSU) and she is an active member of the Cognitive and Social Robotics (CSR) Laboratory, Istanbul Technical University, since 2013. As a PhD candidate, she has taken part in several research projects funded by the Scientific and Technological Research Council of Turkey (TUBITAK) focusing on the use of robot-based assistive systems in the domain of education and rehabilitation for deaf or hard of hearing children. She has worked as a researcher in an Erasmus+ project (EMBOA) to monitor and investigate the affective states of children with autism in robot-assisted therapy. Her current research focuses on the affective behavior modeling for socially assistive robots with an emphasis on the child-robot interaction.

Michal R. Wrobel received a Ph.D. in Computer Science from the Gdansk University of Technology in 2011. Since 2006 he has been working at the Faculty of Electronics, Telecommunications and Informatics, Department of Software Engineering, Gdansk University of Technology, where he is currently an Assistant Professor. He is also a member of the Emotions in HCI research group, where he researches software usability, affective computing and software management methods. His research interests include a modern approach to software development management, with a particular focus on the role of human factors in software engineering. He was project leader of the EMBOA project aimed at improving social robot intervention with children with autism using affective computing technologies.

Tatjana Zorcec is currently a Professor with the Medical Faculty and an Employee of the University Children's Hospital, Skopje, Macedonia. She has expertise in diagnostics and treatment of children with autism and is appointed as a National Coordinator for autism. She has participated in many national and international scientific projects.